Expert Systems With Applications 269 (2025) 126410

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Automatic ultrasound curve angle measurement via affinity clustering for adolescent idiopathic scoliosis evaluation

Yihao Zhou [a], Timothy Tin-Yan Lee [a], Kelly Ka-Lee Lai [a], Chonglin Wu [a], Hin Ting Lau [a], De Yang [a], Zhen Song [a], Chui-Yi Chan [a], Winnie Chiu-Wing Chu [d], Jack Chun-Yiu Cheng [b,c], Tsz-Ping Lam [b,c], Yong-Ping Zheng [a,*]

[a] Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[b] Department of Orthopaedics and Traumatology, The Chinese University of Hong Kong, Hong Kong, China
[c] SH Ho Scoliosis Research Lab, Joint Scoliosis Research Center of the Chinese University of Hong Kong and Nanjing University, Hong Kong, China
[d] Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Ultrasound volume projection imaging
Intelligent scoliosis diagnosis
Vertebrae
Landmark detection

ABSTRACT

The current clinical gold standard for evaluating adolescent idiopathic scoliosis (AIS) is X-ray radiography, specifically through Cobb angle measurement. However, frequent monitoring of AIS progression using X-rays presents a significant challenge due to the risks associated with cumulative radiation exposure. Although 3D ultrasound offers a validated radiation-free alternative, it relies on manual spinal curvature assessment, leading to inter and intra-rater angle variation. In this study, we propose an automated ultrasound curve angle (UCA) measurement system that utilizes a dual-branch network to simultaneously perform landmark detection and vertebra segmentation on ultrasound coronal images. The system incorporates an affinity clustering algorithm within vertebral segments to establish landmark relationships, enabling efficient line delineation for UCA measurement. Our method, specifically optimized for UCA calculation, demonstrates superior performance in landmark and line detection compared to existing approaches. The high correlation between the automatic UCA and Cobb angle (R^2=0.858) suggests that our proposed method can potentially replace manual UCA measurement in ultrasound scoliosis assessment. This advancement could significantly enhance the accuracy and reliability of scoliosis monitoring while reducing the need for manual measurement.

1. Introduction

Adolescent idiopathic scoliosis (AIS) is the most prevalent spinal deformity in children, affecting approximately 0.47–5.2% of teenagers (Konieczny et al., 2013). Clinical diagnosis and progression monitoring of scoliosis rely on the radiographic Cobb spine measurement. However, frequent use of X-rays is not viable due to the potential harm from cumulative radiation exposure, especially in patients who require regular curve monitoring (Himmetoglu et al., 2015; McArthur et al., 2015; Simony et al., 2016). Though the EOS imaging system offers low-dose radiographs, its high setup cost limits its widespread adoption (Jeon et al., 2018). In addition, it is impractical for resource-constrained healthcare facilities to utilize such a system. Therefore, it is essential to explore alternative imaging modalities that are cost-effective, safe, and easily accessible for more regular scoliosis monitoring.

3D ultrasound imaging has emerged as a promising complementary modality for tracking scoliosis, providing a radiation-free solution to reveal pathology. As shown in Fig. 1, the subject is being scanned using an ultrasound probe to capture a series of B-mode ultrasound images and their corresponding 3D spatial information, thereby forming volume data. Volume projection imaging (VPI) generates 2D coronal-plane images from the volume data through non-planar volume rendering (Cheung, Zhou, Law, Lai et al., 2015; Cheung, Zhou, Law, Mak et al., 2015). The shadow of the superficial bone surface enables clinicians to observe spinal deformity due to the nature of ultrasound imaging. Chen et al. were the first to evaluate scoliosis in VPI images by manually identifying the spinous column profile (Fig. 1.(b)) (Chen et al., 2013). Zhou et al. achieved automatic spinous curvature evaluation by utilizing prior knowledge of vertebral anatomical structures (Zhou et al., 2020). However, for patients with severe

* Corresponding author.
E-mail addresses: yihao.zhou@connect.polyu.hk (Y. Zhou), timothy.ty.lee@polyu.edu.hk (T.T.-Y. Lee), kelly.lai@polyu.edu.hk (K.K.-L. Lai), chonglin.wu@polyu.edu.hk (C. Wu), ting-er.lau@polyu.edu.hk (H.T. Lau), de-derek.yang@polyu.edu.hk (D. Yang), zhen0212.song@connect.polyu.hk (Z. Song), stella-chui-yi.chan@polyu.edu.hk (C.-Y. Chan), winniechu@cuhk.edu.hk (W.C.-W. Chu), jackcheng@cuhk.edu.hk (J.C.-Y. Cheng), tplam@cuhk.edu.hk (T.-P. Lam), yongping.zheng@polyu.edu.hk (Y.-P. Zheng).

https://doi.org/10.1016/j.eswa.2025.126410
Received 15 May 2024; Received in revised form 18 December 2024; Accepted 2 January 2025
Available online 7 January 2025

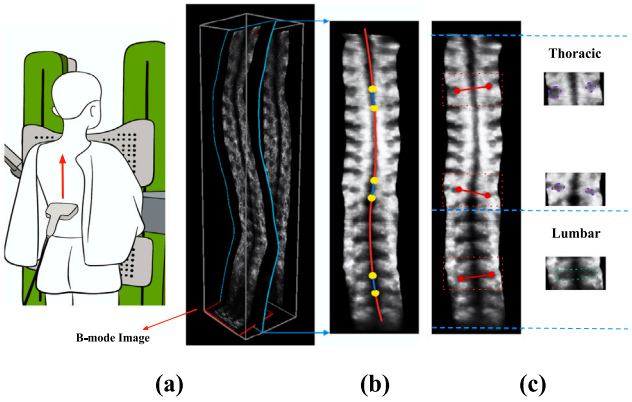

Fig. 1. (a) An Illustration of the generation of a volume projection imaging (VPI). The probe constantly moves from bottom to top along the spine curve on the patient's skin. B-mode images combined with recorded spatial information are grouped to generate 3D ultrasound volume. The coronal ultrasound image is then generated using the VPI method, which incorporates a customized depth profile based on the distance from the skin to the laminae. (b) Ultrasound spinous process angle (SPA). The scoliotic curve on the medial shadow of the spinous processes is used to measure the angle for AIS diagnosis. Cheung, Zhou, Law, Mak et al. (2015) (c) Ultrasound curve angle (UCA). For thoracic region, line is placed on the center of the shadow of a transverse process (purple dotted line). For lumbar region, lines are drawn towards the center of the bilateral sides of lump (green dotted line) (Lee et al., 2021).

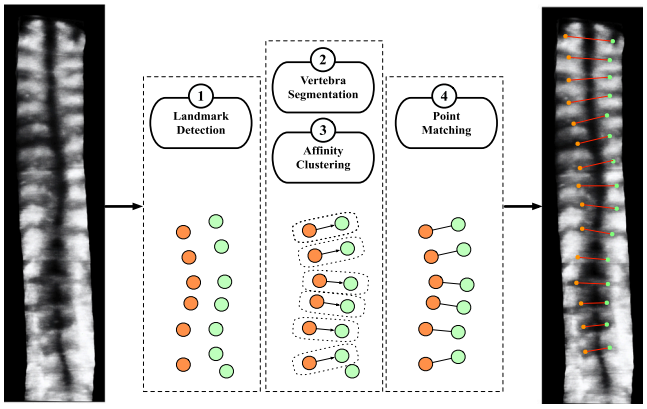

Fig. 2. The model identifies all potential anatomical landmarks along both sides of the bone curvature. The landmarks corresponding to the same vertebrae are connected based on the clustered affinity map. The most tilted lines in different regions are selected to form the UCA for assessing scoliosis.

scoliosis, their spinal processes may deform and rotate significantly. Thus, the spinal profile formed by the spinous process cannot accurately represent the actual lateral deformity of the spine, leading to underestimation of spinal deformity. To evaluate the spine deformity more accurately via VPI, the ultrasound curve angle (UCA) has been proposed (Fig. 1 (c)) (Lee et al., 2021). Shadows corresponding to the transverse processes and ribs in the thoracic region, as well as those from the superior and inferior articular processes in the (thoraco)lumbar region, can be identified similarly. Precise recognition of these structures' foundations is crucial for line delineation. However, manual landmark identification in ultrasound images presents significant challenges, including operator dependency and interpretation difficulties in low-quality images. These limitations necessitate an automated, reliable solution for landmark detection and angle measurement. In this study, we reconceptualizes the angle measurement problem as a pose estimation task (Dang et al., 2019; Wang, Zhang et al., 2021), exploiting the natural pairing pattern of vertebral landmarks along the spinous profile. As illustrated in Fig. 2, our proposed method includes the following steps:

(1) Identify anatomical landmarks using a landmark detection network.
(2) Establish the affinity relationship between landmarks requiring a line connection on the vertebrae segmentation map.
(3) Utilize a grouping strategy to associate detected landmarks with desired connections. This involves categorizing landmarks belonging to the same vertebra and differentiating them from landmarks of other vertebrae.
(4) Draw lines between connected landmarks to visualize and measure the overall spine deformity.

Building upon these concepts, we present an estimation model for automatic UCA measurement. The model architecture comprises a dual-branch network for landmark detection and vertebra discrimination. The detection decoder predicts the heatmap of landmark locations at both thoracic and lumbar regions. Vertebra discrimination is achieved through segmentation, followed by an affinity clustering strategy to establish the point-affinity correspondence via the segmented image for candidate landmark alignment. This innovative framework addresses the fundamental challenges of manual measurement while enhancing accuracy and reproducibility in UCA assessment. The main contributions of this study are summarized as follows:

- We have successfully achieved the automatic measurement of UCA by assembling the points to the line directly, which is reported for the first time.
- We have introduced an innovative affinity clustering strategy designed to capture the affinity relationships among candidate landmarks within the vertebrae segmentation map. This approach facilitates the grouping of landmarks belonging to the same vertebra, forming the angle through optimal parsing with the clustered affinity map.
- We have conducted quantitative experiments on a dataset of ultrasound coronal images with corresponding biplanar radiographs. The strong correlations with the Cobb angle illustrate that our proposed automatic method holds the potential to replace manual UCA measurements in the ultrasound assessment of scoliosis.

2. Related work

Several previous studies have explored the use of ultrasound in diagnosing scoliosis. Cheung et al. first reported using VPI method on a sequence of 2D B-mode ultrasound images to visualize spine anatomy (Cheung, Zhou, Law, Mak et al., 2015). The VPI-SP, midline shadow curve generated by spinous processes (SPs), has demonstrated a good correlation with the Cobb angle (shun Wong et al., 2019; Zheng et al., 2016). Huang et al. developed a method for real-time tracking of SPs in the ultrasonic video to establish a 3D spinal profile for deformity assessment (Huang et al., 2023). As the spine rotates, however, the curvature of SPs might be underestimated. An alternative and more accurate method, UCA, has been demonstrated to be comparable to the conventional Cobb angle (Lee et al., 2021). It computes spinal deformity using the lateral shadow features of transverse processes (TPs), articular processes, and laminae. The prevailing method for performing UCA is through manual measurement, which relies on human discretion. Some research has been conducted on spine segmentation to achieve automatic UCA measurement. Yang et al. proposed a semi-automatic measurement workflow that utilizes the contoured mask of TPs-related features (Yang et al., 2022). Banerjee et al. proposed a hybridized, multi-scale feature fusion U-net to extract semantically rich features and fuse multi-scale features (Banerjee et al., 2022). Huang et al. investigated a joint network for spine segmentation with the interaction of noise-removing work. A selective feature-sharing strategy has been employed to filter out irrelevant features (Huang et al., 2022). Xie et al. utilize a structure-affinity transformer to extract accurate regions of vertebrae (Xie et al., 2024). However, ultrasound images are

inherently susceptible to noise and speckle artifacts, adversely affecting segmentation accuracy, particularly in the thoracolumbar region where anatomical features appear disintegrated (Banerjee et al., 2024). Due to this, the importance of straightforward landmark identification for line placement is self-evident, regardless of the segmentation performance of vertebral bodies.

Instead of drawing lines within the segmentation region, we approach measuring UCA as a pose estimation-like task, i.e., matching the landmark to form the line. Top-down approaches have achieved great success in the pose estimation task, which first detects object bounding boxes in the image and then localizes landmarks for detected individual targets (Khirodkar et al., 2021). Fueled by the transformer's explosion, attention mechanisms have been implemented to capture the global-local correspondence within the image. TokenPose and TransPose introduce extra tokens and a sophisticated decoder, respectively, to predict the concealed landmarks (Li et al., 2021a; Yang et al., 2021). To further exploit the capability of the transformer for feature extraction, multi-resolution parallel transformers are proposed to concentrate on the information across resolutions throughout the whole process (Wang et al., 2022; Yuan et al., 2021). ViTPose and its variant explore the potential of plain and nonhierarchical vision transformers for pose estimation (Xu et al., 2022, 2023). Although these methods perform well in the natural image field, they exhibit certain limitations in medical image analysis, especially in ultrasound spine imaging. Precise detection of the spinal structure's bounding box remains challenging, primarily due to the spine's irregular morphology, homogeneous tissue textures, and poorly defined anatomical boundaries.

On the other hand, bottom-up approaches are more practical as they first identify specific landmarks and then assemble them into objects. OpenPose, a method that utilizes convolutional neural networks and part affinity fields to detect key points and estimate human poses, demonstrates excellent performance in multi-person scenarios and can handle occlusion and overlapping body parts (Cao et al., 2018). Another bottom-up approach is using associate embeddings (Newell et al., 2017). This method introduces an additional channel in the CNN output, called the "tag map", which contains an embedding value for each keypoint. The tag map is used to group detected key points belonging to the same instance by minimizing the distance between their embedding values. This approach simplifies the process of assembling key points into specific objects and has shown promising results in terms of accuracy and efficiency (Geng et al., 2021; Kreiss et al., 2019; Li et al., 2021b; Wang & Zhang, 2022). In this study, we adopt a density-based clustering algorithm to represent individual vertebral regions' orientation and positional information. The identified landmarks are easily assembled using the clustered affinity information to form a line.

3. Method

3.1. Problem statement

In this study, we propose UCA measurement as a landmark match problem. Given vertebral line segments $\hat{v}_1, \ldots, \hat{v}_N$, N is the number of vertebrae, our system simultaneously performs two tasks: (1) detects landmark coordinates on the left and right sides of the spinous process, where $\{\mathbf{p}_i^l \in \mathbb{R}^2\}_{i=1}^{M_l}, \{\mathbf{p}_j^r \in \mathbb{R}^2\}_{j=1}^{M_r}$ represent the left and right point sets respectively, with M_l and M_r denoting their respective cardinalities, and (2) generates vertebrae segmentation masks $\mathbf{Y} \in \mathbb{R}^{H\times W}$. The system then establishes point-affinity correspondences through an affinity clustering mechanism that leverages the directional information embedded in the segmentation masks, resulting in vertebral line predictions $v_k = (\mathbf{p}_{i_k}^l, \mathbf{p}_{j_k}^r)_{k=1}^N$, where i_k and j_k are the matched indices. The line segment of matched landmarks can be calculated as the slope of vertebrae, enabling the measurement of UCA by identifying the most tilted lines in local regions.

3.2. Model architecture

The pipeline of automatic UCA measurement is graphically illustrated in Fig. 3. Our proposed framework consists of five consecutive processes: (a) image feature extraction, (b) landmark location detection, (c) vertebra segmentation, (d) affinity clustering on segmentation, and (e) vertebrae parsing and selection. The subsequent sections offer a comprehensive description of each module.

Feature Encoder: Initially, we employ a convolution neural network as the feature encoder to extract high and low-dimensional features from the input image $x \in \mathbb{R}^{1\times H\times W}$. The feature encoder is implemented using a High-Resolution Network (HRNet) (Wang, Sun et al., 2021). We present the detailed structure of HRNet used in the Appendix. HRNet consists of parallel branches of features at different resolutions, enabling the capture of multi-scale spatial information. The encoder progressively fuses information across these branches, where lower-resolution features from each stage are integrated with higher-resolution features from previous stages. Then, the multi-resolution features are exchanged via a channel-wise summation operation. We upsample the lower-resolution features to match the original resolution using bilinear interpolation. The resulting multi-branch feature representations are concatenated in channel dimensions. To prepare for the transformer in the decoder, the image representation from the feature encoder $\mathbb{F} \in \mathbb{R}^{C\times H\times W}$ is divided into non-overlapping patches $\mathbb{F}_{patch} \in \mathbb{R}^{N\times d}$, where C, N, and d are the feature channel, patch number, and patch channel, respectively. A patch embedding is then added to the patches to retain spatial information.

Vertebrae Segmentation Decoder: The encoded features through the feature extractor serve as the input for the vertebrae segmentation decoder to predict the vertebra region. The structure of the decoder is based on the transformer block (Dosovitskiy et al., 2021; Payer et al., 2019). It contains a layer-normalization and a multi-head self-attention module, followed by a layer-normalization and a multi-layer perception (MLP). The multi-head self-attention (MHSA) mechanism effectively captures non-local dependencies across batches by transforming value features $\mathbf{V} \in \mathbb{R}^{N\times d}$ based on the dot-product similarity computed between query $\mathbf{Q} \in \mathbb{R}^{N\times d}$ and key $\mathbf{K} \in \mathbb{R}^{N\times d}$:

$$\mathbf{MHSA}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathbf{Softmax}(\frac{\mathbf{Q}(\mathbf{K})^T}{\sqrt{h}})\mathbf{V} \tag{1}$$

where h is the head number in the attention block. Residual connections are implemented after the attention and MLP computations to enhance the informative interaction between the input and output of the block. The landmark decoder predicts the location of each landmark via a probability heatmap where the value at the coordinate of the landmark is maximal, and the sum of all values in one channel is equal to 1.

Landmark Detection Decoder: The structure of the landmark decoder is consistent with the architecture in the segmentation decoder except for the input for the attention computation. Specifically, we observe that the semantic information of vertebra, i.e., vertebral segmentation, can make point detection more focused on the vertebra region and suppress irrelevant information. Based on this observation, we replace the multi-head self-attention module in the landmark encoder with the multi-head cross-attention (MHCA) module, where the key value is from the segmentation decoder. The multi-head cross-attention can be formulated as follows:

$$\mathbf{MHCA}(\mathbf{Q},\mathbf{K}_{seg},\mathbf{V}) = \mathbf{Softmax}(\frac{\mathbf{Q}(\mathbf{K}_{seg})^T}{\sqrt{h}})\mathbf{V} \tag{2}$$

The segmentation decoder predicts the pixel-wise vertebrae segmented map in which the directional information of the vertebrae can be decoupled to establish the affinity relationship among the detected points.

For each decoder's output, we reshape the decoded patches to the original image resolution. Finally, we have the segmentation of every

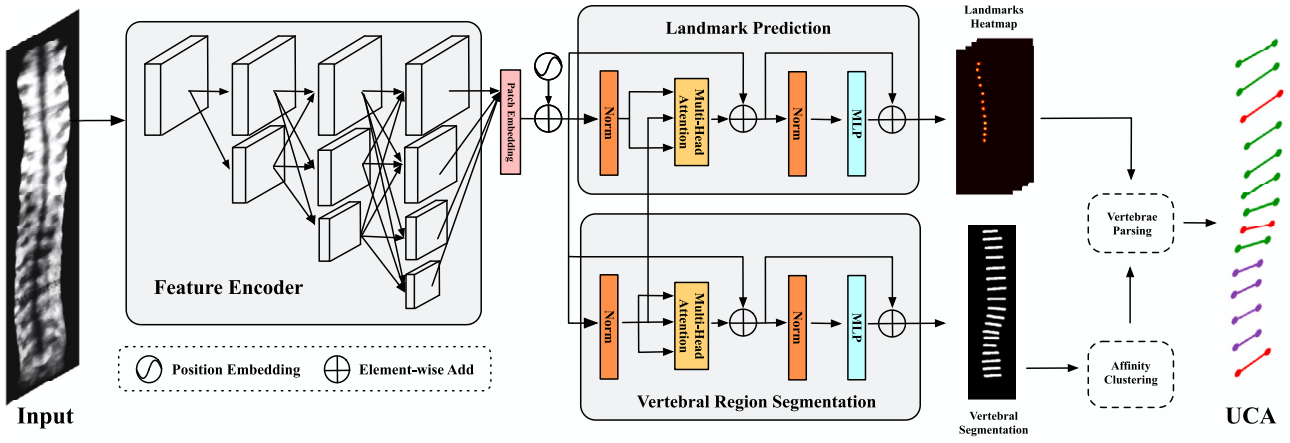

Fig. 3. Overview pipeline of automatic UCA measurement. The model extracts latent features through a feature backbone for landmark detection and spine segmentation. The features in the segmentation decoder are shared with the landmark detection decoder. In the reference stage, the vertebrae segmentation map is parsed to represent the affinity relationship among the detected candidate points. Points belonging to the same vertebra are grouped to form lines.

Algorithm 1 Clustering Affinity Pipeline

Require: Predicted vertebra segmentation map; Threshold γ.

Initialize clustered affinity map $A = 0$
for each foreground point **do**
  if point is not assigned as a neighborhood point **then**
    Search neighborhood points where the value of the path connected to the point is non-zero
    Estimate density as the number of points in the neighborhood
    if density equal or larger than γ **then**
      Grouped the point into the same cluster along with their neighborhood points
      Divide clusters into two subsequences according to the x-coordinate of the clustered points.
      Calculate the centroid of subsequences, c_l and c_r.
      for each point in the cluster **do**
        $A = \frac{c_r - c_l}{\|c_r - c_l\|^2}$
      end for
    else
      Classify the point with its neighborhood points as noise point group
    end if
  else
    Continue
  end if
end for
return Clustered affinity map A

vertebra and the multi-class keypoint-wise heatmap (left thoracic, right thoracic, left lumbar, right lumbar).

Vertebral Affinity Clustering The segmentation map reveals the position of each vertebra and illustrates the spine's orientation, setting the stage for subsequent point matching. Algorithm.1 outlines the clustering affinity pipeline. For each point in the foreground segment, we calculate the density of its neighborhood. A point's neighborhood is defined by the set of points connected to it through non-zero value paths. The density is quantified by counting the number of points within the neighborhood. We then apply a threshold value γ. Points with neighborhood densities equal to or exceeding γ are grouped into the same cluster along with their connected path points. Conversely, points with densities below γ, together with their neighborhood points, are classified as noise.

After completing the clustering process, multiple distinct clusters are identified. Within each cluster, points are sorted in ascending order by their x-coordinates to form an ordered coordinate sequence. Then, the cluster is divided into two subsequences based on the distances to its leftmost and rightmost points. For each subsequence, we compute its centroid, yielding left and right centroid positions. Finally, the orientation of each clustered point is determined through a two-dimensional vector mapping.

$$A = \begin{cases} \frac{c_r - c_l}{\|c_r - c_l\|^2} & \text{if point in a cluster} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Here, A represents the clustered affinity map. If the point belongs to a cluster, the point value is a unit vector from the left centroid $\mathbf{c_l}$ to the right centroid $\mathbf{c_r}$, indicating the direction of the cluster (vertebra). For points outside the cluster, the value is zero.

Vertebrae Parsing and Selection In the reference stage (Fig. 5), we adopt the non-maximum suppression on the predicted keypoint-wise heatmap to acquire the local maximum response of the target point for each channel. These candidate points define a set of potential line connections representative of the vertebrae. We utilize the established point-affinity correspondence in the segmentation map to connect the points that belong to one vertebra. In detail, we have two sets of points distributed on the left and right sides of the spinous process profile. We then perform the line integral computation along the path of two candidates $\vec{A}$ on the clustered affinity map.

$$\mathbf{C}_{ij} = \int_{u=0}^{u=1} \vec{A}((1-u)\,\mathbf{p}_i^l + u\mathbf{p}_j^r)du \tag{4}$$

The $\mathbf{C}_{ij}$ indicates the confidence of whether the $\mathbf{p}_i^{\mathbf{l}}$ and $\mathbf{p}_j^{\mathbf{r}}$ are connected to form the line. The optimal matching of candidates is achieved with the Hungarian algorithm to acquire all the vertebrae connections (Kuhn, 1955). After obtaining the optimal matching, we filter out low-confidence matches based on the criteria that confidence significantly lower than the average of finished matches are ignored, as they might intersect with other line segments or be less reliable. The line of interest for UCA measurements is based on the horizontal slope between each pair of line segments and their adjacent counterparts to identify local extrema (peaks and valleys) (Fig. 4). Additionally, if the absolute value of the angle formed by the detected uppermost or lowermost vertebral bodies in the global space and the most inclined vertebral body obtained in the local space exceeds 10 degrees, the UCA is calculated. Consequently, all detected vertebral bodies are considered for UCA computations.

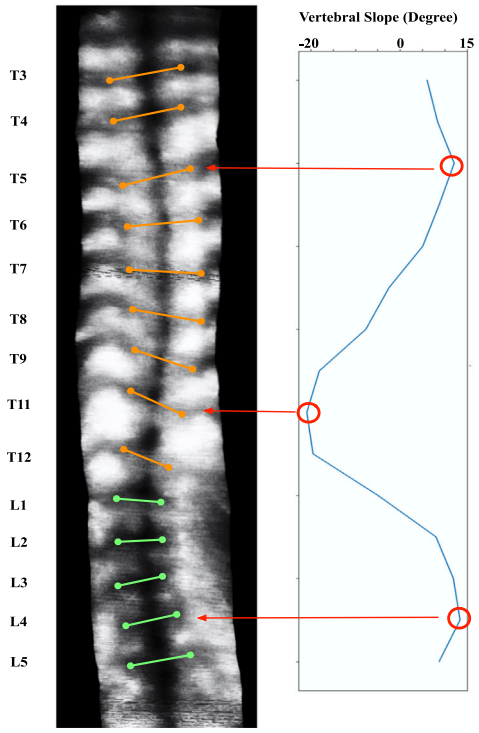

Fig. 4. An example of vertebral-level line detection. The local extreme values are used for angle measurement.

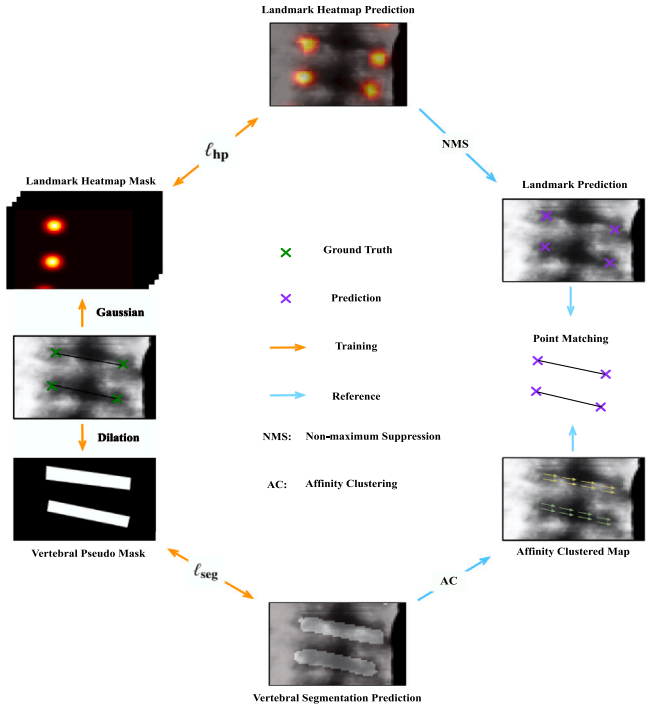

Fig. 5. The training and reference pipeline of automatic UCA measurement.

3.3. Training

For landmark heatmap training, we initially use a Gaussian kernel applied to the annotated points to produce ground truth keypoint-wise heatmap, which is denoted as $\hat{\mathbf{H}} = Gaussian(\sigma)$. σ is a learnable variable that controls the standard deviation of the Gaussian kernel. We use a mean square error loss to compute the difference between the predicted and ground truth heatmap. The heatmap loss function $\ell_{\mathbf{hp}}$ is as follows:

$$\ell_{\mathbf{hp}} = -\min_{\sigma} \sum_{\mathbf{p}} (\left\| \hat{\mathbf{H}}(\sigma) - \mathbf{H} \right\|^2) + \|\sigma\|_2^2 \tag{5}$$

where $\mathbf{H} \in \mathbb{R}^{4\times H\times W}$ is the predicted heatmap, where each channel corresponds to one of four landmark points: left thoracic, right thoracic, left lumbar, and right lumbar regions. The regularization term in the L2 norm encourages the values of σ to be minimized, while the former objective function favors larger σ values. This creates a balance where larger σ results in over-smoothed predictions that may be inaccurate. Conversely, smaller sigma values lead to highly accurate responses but with multiple peaks nearby.

For vertebrae segmentation training, we create pseudo-masks $\hat{\mathbf{Y}}$ as the ground truth segmented map by leveraging annotated UCA line segments. We define a fixed-size convolution kernel K and apply a dilation operation to line segment v:

$$(v \oplus K)(x, y) = \bigcup_{(i,j)\in K} v(x - i, y - j) \tag{6}$$

$$\hat{\mathbf{Y}} = \bigcap_{i=1}^{n} (v_i \oplus K) \tag{7}$$

The operation $\oplus$ represents dilation. The K controls the degree of dilation over the lines. For each pixel location (x, y), the result of $(v \oplus K)$ is the union of the pixels from $l(x - i, y - j)$ for all offsets (i, j) within the kernel K. The intersection operation $\bigcap$ is then applied to all these dilated line segments. The $\hat{\mathbf{Y}}$ represent the areas of dilation across all the ground truth line segments. After acquiring the pseudo-masks, the segmentation loss is based on the Dice loss as expressed by:

$$\ell_{\mathbf{seg}} = 1 - \frac{2\left|\hat{\mathbf{Y}} \bigcap \mathbf{Y}\right|}{\left|\hat{\mathbf{Y}}\right| + |\mathbf{Y}|} \tag{8}$$

The total loss ℓ_{total} is:

$$\ell_{total} = \lambda_1 \times \ell_{\mathbf{hp}} + \lambda_2 \times \ell_{\mathbf{seg}} \tag{9}$$

We display the training and reference pipeline of the proposed framework in Fig. 5. In our experiment, we empirically set $\lambda_1 = 1$ and $\lambda_2 = 0.2$ to stabilize the training process.

4. Experiments

4.1. Materials

All participants were recruited from the Department of Orthopedics and Traumatology of The Chinese University of Hong Kong. Informed consents are obtained before the scanning session. Patients with a Cobb angle greater than 60° and BMI indices above 25.0 kg/m^2 were excluded. VPI images were acquired by two 3D ultrasound imaging systems: Scolioscan 801 and Scolioscan Air (Lai et al., 2021). The Scolioscan 801 system achieves 3D ultrasound imaging using an ultrasound probe (a custom-designed linear probe with a frequency of 7.5 MHz and width of 10 cm) combined with an electromagnetic spatial sensor. The Scolioscan Air is a portable 3D ultrasound imaging system that consists of a palm-sized linear ultrasound module with a transducer of 75 mm in width and a central frequency of 7.5 MHz, a Realsense depth tracking camera, and a tablet PC installed with a customized software using volume project imaging (VPI) to process, reconstruct, and visualize the image. For model development, three experts with more than 5 years of ultrasound experiments manually annotated the line between the spinal feature points on both sides of the spinous process profile as ground truth. 1212 cases were included, with 970 cases used to train the model and 242 used for the in-house validation dataset to evaluate the performance. 386 prospective cases with biplanar radiographs were

Table 1
Comparison of different advanced methods on line prediction in the thoracic and lumbar regions.

Method		Params (M)	Speed (ms)	Thoracic		Lumbar	
				AP	AR	AP	AR
Top-Down	HRFormer	56	203	0.831 ± 0.14	0.947 ± 0.06	0.763 ± 0.25	0.923 ± 0.12
	TransPose	68	112	0.842 ± 0.10	0.943 ± 0.07	0.785 ± 0.24	0.913 ± 0.15
	TokenPose	58	96	0.863 ± 0.13	0.936 ± 0.07	0.734 ± 0.21	0.899 ± 0.13
	ViTPose	92	72	0.856 ± 0.12	0.958 ± 0.05	0.746 ± 0.20	0.921 ± 0.11
	ViTPose++	92	75	0.864 ± 0.13	0.952 ± 0.06	0.752 ± 0.21	0.914 ± 0.10
Bottom-up	Hourglass[a]	290	396	0.871 ± 0.12	0.788 ± 0.16	0.839 ± 0.20	0.789 ± 0.22
	HRNet-w32[a]	35	89	0.863 ± 0.13	0.875 ± 0.13	0.842 ± 0.24	0.791 ± 0.22
	HRNet-w48[a]	72	123	0.871 ± 0.12	0.875 ± 0.16	0.870 ± 0.18	0.799 ± 0.21
	HigherHRNet-w32[a]	35	95	0.909 ± 0.09	0.941 ± 0.11	0.874 ± 0.18	0.784 ± 0.21
	HigherHRNet-w48[a]	72	128	0.934 ± 0.08	0.928 ± 0.11	0.901 ± 0.14	0.858 ± 0.16
	DEKR	32	117	0.945 ± 0.07	0.938 ± 0.06	0.902 ± 0.12	0.903 ± 0.13
	CID	34	63	0.915 ± 0.08	0.942 ± 0.06	0.843 ± 0.14	0.901 ± 0.14
	Ours	51	95	0.963 ± 0.08	0.950 ± 0.08	0.918 ± 0.10	0.907 ± 0.11

[a] Indicates using associate embedding for point grouping, AP: Average precision, AR: Average recall.

used to test the model's performance after the model was developed. The Cobb angles were measured by two radiograph experts.

The model is implemented based on PyTorch and trained on a 48 GB NVIDIA RTX A6000 GPU. The data augmentation includes random horizontal flipping, rotation ranging from −30 to 30 degrees, brightness, and contrast transformation. We resize the input images into 256×512, keeping the aspect ratio. The initial learning rate is $1e^{-5}$, and an Adam optimizer with a momentum of 0.9 is employed for model development. The total training epoch is $1e^4$, with a decrease in learning rate to half every 2000 epoch. We set the hyper-parameters of γ and K to 10 and 3, respectively. γ is designed to filter out the less-than-ideal segmentated vertebral region. K controls the range of dilation of line segments for pseudo-segmentation mask generation. In the ablation study section, we provide comprehensive experiments to investigate the contribution of these hyper-parameters to the model performance.

4.2. Evaluation metrics

The performance evaluation of the model was conducted using vertebral-level line prediction and UCA measurement. Traditional metrics such as Mean Euclidean Distance (MED) and Mean Manhattan Distance (MMD) are often used to assess the disparity between predicted points and ground truth (Meng et al., 2023). However, these metrics may not fully account for the possibility of redundant or missing predicted lines. To provide a more comprehensive evaluation of the line prediction performance, we integrate the point-based metrics with the concept of Endpoint Distance Error (EDE) as follows:

$$EDE = \frac{\sum_{l,r} exp\left(-d_{l,r}/s\right)}{2}. \tag{10}$$

Here $d_{l,r}$ stands for the Euclidean Distance of the left and right endpoint of the predicted line with its corresponding ground truth. We set s equal to 100, which is a scaling factor. The EDE measures the localization accuracy of a single line, where correct prediction is defined as the distance between the predicted endpoint and the ground truth less than 3.5 mm (EDE > 0.5) according to the posterior vertebral body heights obtained from a study using the human cadaver (Kunkel et al., 2011). The perfect prediction would yield an EDE of 1. We define the average precision (AP) and average recall (AR) scores as the ratio of corrected lines to the total ground truth and the ratio of corrected lines to the total of predicted lines in the scan, respectively. These metrics consider both the redundancy and absence of predicted lines. To further verify the validity of the UCA measurement, we used linear regression and Bland–Altman analysis to investigate the agreement between predicted UCA and Cobb angle.

4.3. Comparison with advanced networks

The performance of line prediction is estimated by comparing it with other state-of-the-art estimation methods, including the Hourglass (Newell et al., 2016), HRNet (Wang, Sun et al., 2021), HigherHRNet (Cheng et al., 2020), CID (Wang & Zhang, 2022), and DEKR (Geng et al., 2021). We reimplement the methods according to the mmPose.[1] For baselines (Cheng et al., 2020; Newell et al., 2016; Wang, Sun et al., 2021), we assemble detected key points whose tags have a small L_2 distance into a line using associative embedding. Additionally, we use CenterNet (Duan et al., 2019) as the vertebrae detector for the comparison of Top-down transformed-based methods, including TransPose (Yang et al., 2021), HRFormer (Yuan et al., 2021), TokenPose (Li et al., 2021a), ViTPose (Xu et al., 2022), and ViTPose++ (Xu et al., 2023). We use ViT-B, HRFormer-B, and HRNet-w48 as the backbones for ViTPose/ViTPose++, HRFormer, and TokenPose/TransPose, respectively. Table 1 summarize the comparison results. Top-down methods show strong AR performance but suffer significant drop in AP metrics. This limitation is primarily from their dependence on vertebrae detection, which becomes unreliable in the presence of artifacts and poor image quality. Moreover, in severe scoliosis cases, where multiple vertebrae may occupy a single bounding box due to spinal deformity, the proximity of vertebrae often results in ambiguous landmark assignments. Conversely, bottom-up methods achieve more robust performance in AP. However, it is worth noting that these methods fail to effectively address the challenge of erroneous landmark connections due to the structural similarity between the current vertebral body and its neighboring vertebrae. Consequently, the key point parsing process generates numerous intersecting line segments, significantly reducing accuracy. Furthermore, relying exclusively on the Gaussian heatmap generated from the labels leads to inadequate performance, particularly in accurately detecting all landmarks in the presence of image blurriness on the vertebral region. Our proposed method achieved superior performance in terms of precision and recall. This implies that the predicted line segments are accurate within the specified range, and the occurrence of redundant line segments is minimal. We attribute this notable performance difference to the limitations of other methods, particularly in their regression of association between key points. Our proposed method benefits from segmentation supervision, enabling precise landmark prediction on both sides of the spine. We employ a clustering strategy to address the ambiguity in certain regions predicted by the model. This strategy effectively reduces the generation of redundant line segments by classifying these ambiguous areas as noise regions, preventing their impact on identifying vertebral regions.

[1] https://github.com/open-mmlab/mmpose.

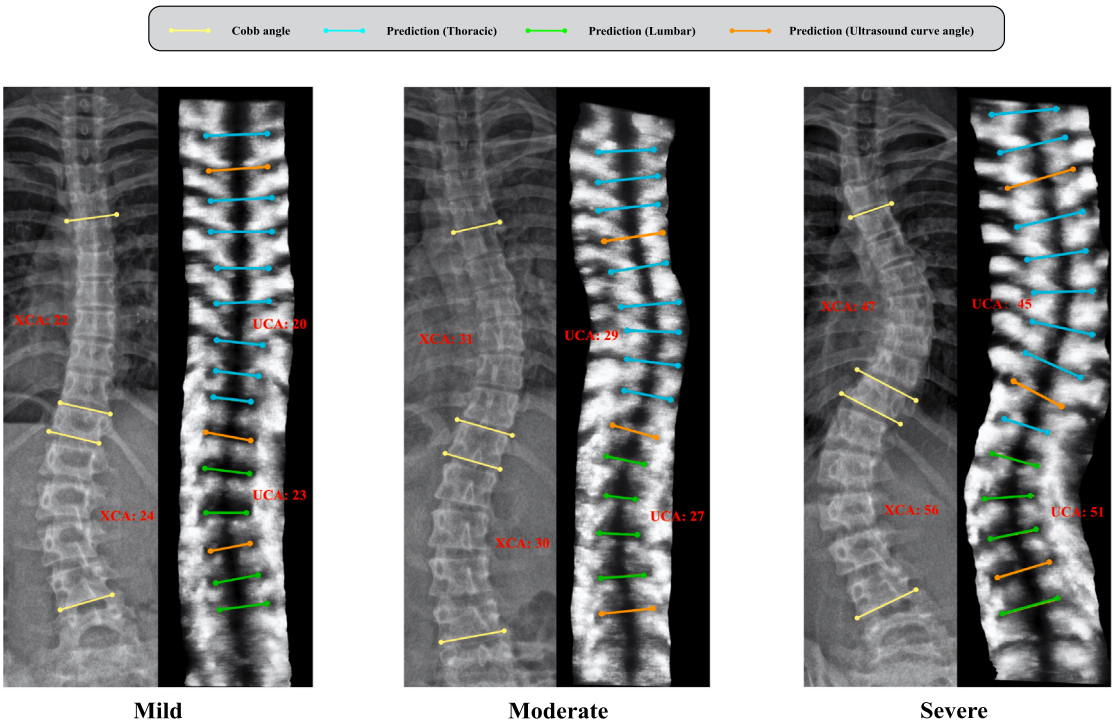

Fig. 6. Visual comparison between Ultrasound Curve Angle (UCA) and X-ray Cobb angle (XCA) in different degrees of scoliosis.

Table 2
Ablation study of the size of dilated kernel and segmentation feature transformation.

FF	K	Thoracic		Lumbar	
		AR	AP	AR	AP
	1	0.951(0.12)	0.930(0.10)	0.905(0.09)	0.891(0.09)
✓	1	0.955(0.10)	0.940(0.10)	0.910(0.10)	0.890(0.09)
	3	0.956(0.10)	0.941(0.10)	0.911(0.10)	0.906(0.11)
✓	3	0.963(0.08)	0.950(0.08)	0.918(0.10)	0.907(0.11)
	5	0.958(0.10)	0.940(0.09)	0.917(0.10)	0.910(0.09)
✓	5	0.961(0.09)	0.947(0.08)	0.915(0.11)	0.906(0.08)

Table 3
Ablation study of noise threshold.

Threshold γ	Thoracic		Lumbar	
	AR	AP	AR	AP
1	0.881 (0.12)	0.901 (0.10)	0.837 (0.12)	0.891 (0.12)
10	0.907 (0.11)	0.899 (0.09)	0.846 (0.11)	0.903 (0.11)
50	0.267 (0.21)	0.326 (0.17)	0.146 (0.27)	0.196 (0.23)

For computational complexity analysis, we adopt HRNet as the backbone to extract image features and achieve model efficiency without introducing excessive parameters. The resulting reference speed indicates that our model architecture design meets the real-time requirements for practical large-scale scoliosis screening.

4.4. Ablation studies

This section aims to verify the effectiveness of the proposed components, including the affinity clustering (AC) strategy for point grouping, different sizes of dilated kernels for vertebral region discrimination, and noise threshold. To assess the contribution of our proposed components for landmark detection, we chose a different size of the dilated kernel to create the mask of the vertebral region. We also investigate the effectiveness of feature transformation (FF) from segmentation to the landmark detection encoder. The comparison results are shown in Table 2. The AR and AP are optimal when the kernel size is set to 3. A too-large, dilated kernel may lead to overlapping spatial position predictions of different vertebrae, resulting in a biased affinity representation towards the direction of the upper and lower vertebrae. Conversely, a too-small, dilated kernel makes model convergence more challenging, with an increased likelihood of predicting regions outside the vertebrae. On the other hand, experimental results suggest that introducing representations of vertebral regions aids in landmark localization, especially when one side's shadow of the spinous process is unclear or missing due to the discontinuity in the scanning process.

The definition of noise regions in the segmentation map significantly impacts the establishment of the point-affinity relationship, as demonstrated in Table 3. A threshold that is too small ($\gamma = 1$) may result in uncertainty of affinity representation of the vertebra due to fragmentary vertebral segmentation, particularly in the lumbar region of low vertebral visibility. Conversely, an overly high threshold ($\gamma = 50$) may mistakenly classify valid vertebral segments as noise, preventing decoding landmark correspondence at that region.

We investigate the performance of two bottom-up strategies for keypoint grouping: associate embedding (AE) and part affinity field (PAF). We modify the output of the head of segmentation to predict the PAF and the embedding heatmap for each candidate, respectively. The results are shown in Table 4. We visualize the predicted landmarks from different grouping strategies with their corresponding line connection in Fig. 6. The model performs point grouping based on the part affinity field. Still, it is more powerful in ultrasound coronal images. Different from directly regressing the affinity field of key points, clustered affinity information from segmented spines performed more efficiently for the following reasons. Firstly, spinal structures do not exhibit feature overlap, simplifying the prediction of location information across the region of a vertebra. Secondly, we notice that the prediction of affinity or embedding heatmap from the neural network is affected by the similar profile of adjacent vertebrae. This results in detected key points that tend to be connected to adjacent vertebrae rather than the exact vertebra where the landmark is located.

Table 4
Ablation study of different point grouping strategies.

Grouping	Thoracic		Lumbar	
	AR	AP	AR	AP
AE	0.922 (0.11)	0.910 (0.09)	0.875 (0.11)	0.879 (0.11)
PAF	0.931 (0.10)	0.904 (0.09)	0.885 (0.11)	0.890 (0.12)
AC	0.963 (0.08)	0.950 (0.08)	0.918 (0.10)	0.907 (0.11)

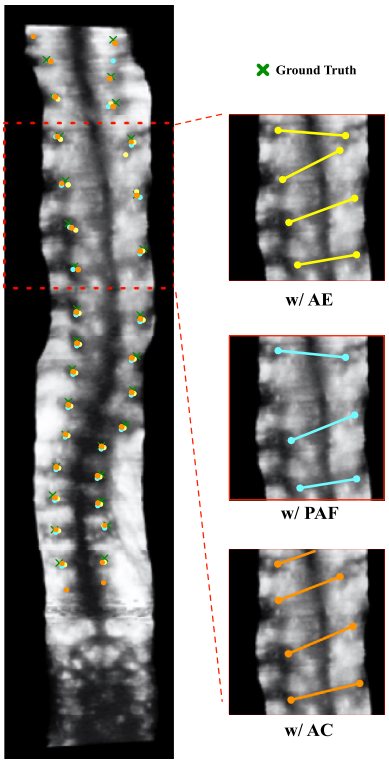

Fig. 7. Sample image in ablation study of using different point grouping strategy. The similarity in the features of vertebrae could result in incorrect line connections.

4.5. Comparison with Cobb angle

In this section, we evaluate the correlation between automatic ultrasound curve angle and X-ray Cobb angle. The linear regression analysis and Bland-Altman plots are shown in Fig. 7. Fig. 8 visualizes the result between predicted UCAs and Cobb angles for the same patients. The results reveal a strong correlation between the automated UCA and the Cobb angle, evidenced by an R^2 value of 0.858. The study results closely align with the previously reported results of comparing manual UCA and Cobb angle ($R^2 = 0.888$) (Lee et al., 2021). The Bland-Altman plots indicate an overall mean difference of 1.31 degrees, exhibiting good agreement between the predicted UCA and the Cobb angle. Seventy-six percent of UCA (441 out of 580 curves) exhibit a difference within 5° compared to the Cobb angle. The scaling factor, derived from the linear equation, is determined to be 1.02, indicating a great agreement between automated UCA and the Cobb angle. We observed that the vertebral bodies contributing to the angle measurement may not align consistently between ultrasound and X-ray imaging. This discrepancy arises from the different approaches used in the two modalities. In X-ray, the apex position is first determined, and then line segments are selected on either side of the apex. In contrast, we directly calculate the inclination of each vertebral body in ultrasound, thereby avoiding the potential occurrence of adjacent vertebral bodies being more inclined than the measured vertebral body. Our approach achieves a more accurate and reliable assessment of the vertebral inclination.

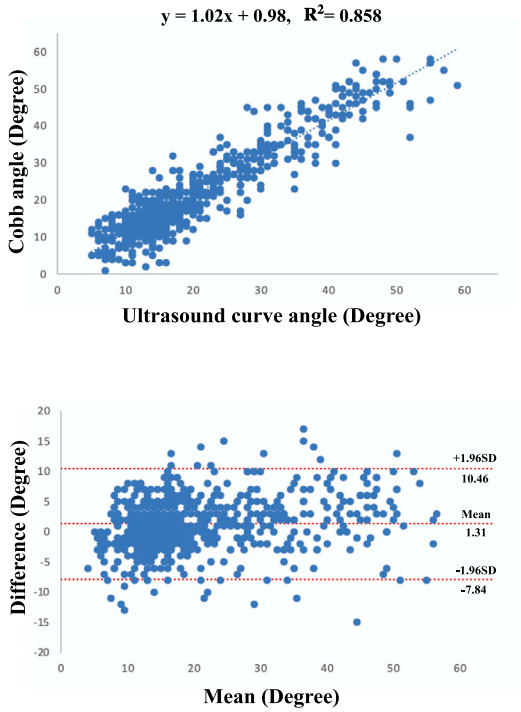

Fig. 8. Linear regression analysis and Bland-Altman plots of predicted UCA on the test data.

5. Discussion

Ultrasound imaging has been validated as a diagnostic tool for scoliosis assessment, utilizing various spinal anatomical structures as a reference to quantify the degree of spinal curvature. Previous methods have researched on features of the spinous process for assessing spinal curvature (Brignol et al., 2020; Chen et al., 2024; Huang et al., 2023; Ungi et al., 2020). However, axial vertebral rotation, common in mild and moderate scoliosis, can lead to an underestimation of spinal curvature when using the spinous process. Recently, many segmentation-based methods have been developed for measuring the angle based on the segmented transverse process, and lumbar lump (Banerjee et al., 2022; Huang et al., 2022; Wong et al., 2022). However, these methods can be affected by artifacts in ultrasound images, which may arise due to the discontinuity in probe movement during scanning. In addition, due to the similarity of vertebrae features in ultrasound images, previous landmark-based algorithms struggle to establish the point correlation of detected landmarks, leading to connections between different vertebrae landmarks. As a result, a fully automatic angle measurement solution for scoliosis analysis remains unachievable.

The development of an automated measurement system for ultrasound-based scoliosis assessment is of paramount clinical importance as it eliminates inter- and intra-rater measurement variability and standardizes the diagnostic workflow. In this study, we first achieve automatic ultrasound curve angle measurement by employing a bottom-up strategy to establish point-affinity correspondence among detected landmarks of the vertebrae. Our method is compared with other state-of-the-art detection networks and X-ray Cobb angle, with a high correlation of $R^2 = 0.835$. By leveraging prior knowledge of vertebrae segmentation, our method effectively captures the direction of each vertebra. Notably, our method does not rely on identifying the apex due to the invisibility of the upper and lower edges of vertebrae in ultrasound images; instead, it directly compares the slope of each vertebra to obtain UCA. By automating the angle measurement process

Table 5
Detailed architecture of feature encoder.

Layer name	Input size	Output size	Details
Input	1 × 512 × 256	–	Raw input
Stem network			
Stem Conv1	1 × 512 × 256	64 × 256 × 128	Conv(3 × 3, 64), stride = 2, BN, ReLU
Stem Conv2	64 × 256 × 128	64 × 256 × 128	Conv(3 × 3, 64), stride = 1, BN, ReLU
Stage 1 (4 Residual Units)			
Bottleneck1_1	64 × 256 × 128	256 × 256 × 128	Conv(1 × 1, 64) Conv(3 × 3, 64) Conv(1 × 1, 256)
Bottleneck1_2–4	256 × 256 × 128	256 × 256 × 128	Same as Bottleneck1_1
Stage 2 (Parallel Branches)			
Branch1	256 × 256 × 128	32 × 256 × 128	4 × Residual Units Each: Conv(1 × 1, 32) Conv(3 × 3, 32 Conv(1 × 1, 32)
Branch2	256 × 256 × 128	64 × 128 × 64	Strided Conv(3 × 3, 64) 4 × Residual Units Each: Conv(1 × 1, 64) Conv(3 × 3, 64) Conv(1 × 1, 64)
Stage 3 (Multi-Resolution Fusion)			
Branch1	32 × 256 × 128	32 × 256 × 128	4 × Residual Units + Multi-Scale Fusion
Branch2	64 × 128 × 64	64 × 64 × 32	4 × Residual Units + Multi-Scale Fusion
Branch3	64 × 128 × 64	128 × 64 × 32	Strided Conv(3 × 3, 128) 4 × Residual Units + Multi-Scale Fusion
Stage 4 (Multi-Resolution Fusion)			
Branch1	32 × 256 × 128	32 × 256 × 128	4 × Residual Units + Multi-Scale Fusion
Branch2	64 × 128 × 64	64 × 128 × 64	4 × Residual Units + Multi-Scale Fusion
Branch3	128 × 64 × 32	128 × 64 × 32	4 × Residual Units + Multi-Scale Fusion
Branch4	128 × 64 × 32	256 × 32 × 16	Strided Conv(3 × 3, 256) 4 × Residual Units + Multi-Scale Fusion
Final Multi-Resolution Fusion			
Final Fusion	32 × 256 × 128 64 × 128 × 64 128 × 64 × 32 256 × 32 × 16	480 × 512 × 256	Bilinear Upsampling + Feature Concatenation

for ultrasound spine images, our approach could reduce the time and resources required for manual analysis, potentially leading to cost savings in clinical settings. Additionally, the increased accuracy and reliability of our method may contribute to improved patient outcomes, which can have further economic implications. Our approach, based on vertebrae segmentation and landmark detection, can be easily extended to Cobb angles measurement and segmentation problems in X-ray (Jaszcz et al., 2022). Our method also supports vertebral-level analysis, potentially useful for surgical navigation and treatment monitoring (Gueziri et al., 2020; Zhang et al., 2021).

However, notable angle deviations are observed in cases with poor image quality, primarily due to uncertainties in landmark detection. This sub-optimal image quality could be attributed to insufficient contact between the probe and the skin during ultrasound scanning. In addition, the VPI images are generated based on the average skin-to-laminae distances, which can vary among individuals. Consequently, the generated VPI images may not optimally visualize all vertebral features necessary for UCA measurement. This issue also impedes our method's ability to accurately distinguish between the thoracic and lumbar regions in the ultrasound VPI images, a crucial step in the automatic UCA process. This is because thoracic and lumbar UCAs are computed using different anatomical features derived from the VPI images. Our current approach involves identifying the last pair of ribs on the 12th vertebra to separate the thoracic and lumbar regions, thereby enabling the application of different strategies for assigning UCA lines. However, VPI images generated based on average skin-to-laminae distances may not visualize the 12th ribs. In future studies, we plan to use a gel pad to minimize the likelihood of insufficient contact between the skin and the probe. Additionally, we aim to incorporate information about the ribs into the model to improve the accuracy of distinguishing between the thoracic and lumbar regions. It is noted that our ultrasound protocol's primary targets are preoperative cases; therefore, we did not recruit subjects with large Cobb angles. In future studies, we plan to incorporate additional case studies with more severe scoliosis to comprehensively evaluate the model's performance across a broader spectrum of scoliosis severity. This expansion will allow us to assess the robustness and adaptability of our approach in handling more complex cases, ultimately enhancing the generalizability and clinical applicability of our model. In addition, the 7.5 MHz probe used in this study may not be optimal for acquiring good-quality US images, so high BMI subjects are not included. In future studies, ultrasound probes with lower frequencies will be adopted to investigate the feasibility of using these probes on high BMI subjects.

6. Conclusion

We have proposed a framework for automatic ultrasound curve angle measurement, identifying potential landmarks, and performing line delineation. Our method addresses the challenge of localizing points to perform line delineation automatically. This is meaningful in the clinical setting because the manual process of drawing lines is both time-consuming and operator-dependent. Different from previous segmentation-based networks on volume data or 2D B-mode images, our approach does not rely on the accurate annotation of vertebral structures. In addition, transferring segmentation features into landmark detection allows the model to focus more on specific target areas, making the prediction of the precise location of potential landmarks accurate. Beyond angle measurement, our approach supports vertebral-level analysis, providing a comprehensive understanding of spinal morphology. The superior performance of our method compared to other advanced methods indicates the effectiveness of the proposed network. Furthermore, experiments on a VPI image dataset with radiographs demonstrated the reliability of the proposed network. With minimal operator interaction and skills required, clinicians can efficiently acquire the angle from ultrasound coronal images, eliminating intra-rater and inter-rater operator variation. This holds great potential for replacing manual UCA measurements.

CRediT authorship contribution statement

Yihao Zhou: Conceptualization, Methodology, Writing – original draft. **Timothy Tin-Yan Lee:** Supervision, Validation, Writing – review & editing. **Kelly Ka-Lee Lai:** Investigation, Data curation. **Chonglin Wu:** Software. **Hin Ting Lau:** Investigation. **De Yang:** Project administration. **Zhen Song:** Validation, Formal analysis. **Chui-Yi Chan:** Data curation. **Winnie Chiu-Wing Chu:** Resources. **Jack Chun-Yiu Cheng:** Resources. **Tsz-Ping Lam:** Resources. **Yong-Ping Zheng:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yongping Zheng reports a relationship with Telefield Medical Imaging Limited that includes: board membership, consulting or advisory, and equity or stocks. Yongping Zheng has patent #A three-dimensional (3D) ultrasound imaging system for assessing scoliosis (US8900146B2) issued to The Hong Kong Polytechnic University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was partially supported by The Research Grant Council of Hong Kong (R5017-18).

Appendix

The detailed structure of the feature encoder used in this study is illustrated in Table 5.

Data availability

The authors do not have permission to share data.

References

Banerjee, S., Huang, Z., Lyu, J., Leung, F. H., Lee, T., Yang, D., Zheng, Y., McAviney, J., & Ling, S. H. (2024). Automatic assessment of ultrasound curvature angle for scoliosis detection using 3-D ultrasound volume projection imaging. *Ultrasound in Medicine & Biology*, *50*(5), 647–660.

Banerjee, S., Lyu, J., Huang, Z., Leung, F. H., Lee, T., Yang, D., Su, S., Zheng, Y., & Ling, S. H. (2022). Ultrasound spine image segmentation using multi-scale feature fusion skip-inception U-net (SIU-net). *Biocybernetics and Biomedical Engineering*, *42*, 341–361.

Brignol, A., Gueziri, H.-E., Cheriet, F., Collins, D. L., & Laporte, C. (2020). Automatic extraction of vertebral landmarks from ultrasound images: A pilot study. *Computers in Biology and Medicine*, *122*, Article 103838.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2018). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. URL http://arxiv.org/abs/1812.08008.

Chen, W., Lou, E. H., Zhang, P. Q., Le, L. H., & Hill, D. (2013). Reliability of assessing the coronal curvature of children with scoliosis by using ultrasound images. *Journal of Children's Orthopaedics*, *7*.

Chen, H., Qian, L., Gao, Y., Zhao, J., Tang, Y., Li, J., Le, L. H., Lou, E., & Zheng, R. (2024). Development of automatic assessment framework for spine deformity using freehand 3D ultrasound imaging system. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*.

Cheng, B., Xiao, B., Wang, J., Shi, H., Huang, T. S., & Zhang, L. (2020). HigherhrNet: Scale-aware representation learning for bottom-up human pose estimation. http://dx.doi.org/10.1109/CVPR42600.2020.00543.

Cheung, C. W. J., Zhou, G. Q., Law, S. Y., Lai, K. L., Jiang, W. W., & Zheng, Y. P. (2015). Freehand three-dimensional ultrasound system for assessment of scoliosis. *Journal of Orthopaedic Translation*, *3*.

Cheung, C. W. J., Zhou, G. Q., Law, S. Y., Mak, T. M., Lai, K. L., & Zheng, Y. P. (2015). Ultrasound volume projection imaging for assessment of scoliosis. *IEEE Transactions on Medical Imaging*, *34*, 1760–1768.

Dang, Q., Yin, J., Wang, B., & Zheng, W. (2019). Deep learning based 2D human pose estimation: A survey. *Tsinghua Science and Technology*, *24*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: transformers for image recognition at scale.

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). Centernet: Keypoint triplets for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6569–6578).

Geng, Z., Sun, K., Xiao, B., Zhang, Z., & Wang, J. (2021). Bottom-up human pose estimation via disentangled keypoint regression. http://dx.doi.org/10.1109/CVPR46437.2021.01444.

Gueziri, H.-E., Santaguida, C., & Collins, D. L. (2020). The state-of-the-art in ultrasound-guided spine interventions. *Medical Image Analysis*, *65*, Article 101769.

Himmetoglu, S., Guven, M. F., Bilsel, N., & Dincer, Y. (2015). DNA damage in children with scoliosis following X-ray exposure. *Minerva Pediatrica*, *67*.

Huang, Y., Jiao, J., Yu, J., Zheng, Y., & Wang, Y. (2023). Si-MSPDNet: A multiscale siamese network with parallel partial decoders for the 3-D measurement of spines in 3D ultrasonic images. *Computerized Medical Imaging and Graphics*, *108*.

Huang, Z., Zhao, R., Leung, F. H., Banerjee, S., Lee, T. T. Y., Yang, D., Lun, D. P., Lam, K. M., Zheng, Y. P., & Ling, S. H. (2022). Joint spine segmentation and noise removal from ultrasound volume projection images with selective feature sharing. *IEEE Transactions on Medical Imaging*, *41*, 1610–1624.

Jaszcz, A., Połap, D., & Damaševičius, R. (2022). Lung x-ray image segmentation using heuristic red fox optimization algorithm. *Scientific Programming*, *2022*(1), Article 4494139.

Jeon, C.-H., Kwack, K.-S., Park, S., Lee, H.-D., & Chung, N.-S. (2018). Combination of whole-spine lateral radiograph and lateral scanogram in the assessment of global sagittal balance. *The Spine Journal*, *18*(2), 255–260.

Khirodkar, R., Chari, V., Agrawal, A., & Tyagi, A. (2021). Multi-instance pose networks: Rethinking top-down pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3122–3131).

Konieczny, M. R., Senyurt, H., & Krauspe, R. (2013). Epidemiology of adolescent idiopathic scoliosis.

Kreiss, S., Bertoni, L., & Alahi, A. (2019). *Vol. 2019-June, PifPaf: Composite fields for human pose estimation*.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*.

Kunkel, M. E., Herkommer, A., Reinehr, M., Böckers, T. M., & Wilke, H.-J. (2011). Morphometric analysis of the relationships between intervertebral disc and vertebral body heights: an anatomical and radiographic study of the human thoracic spine. *Journal of Anatomy*, *219*(3), 375–387.

Lai, K. K. L., Lee, T. T. Y., Lee, M. K. S., Hui, J. C. H., & Zheng, Y. P. (2021). Validation of scolioscan air-portable radiation-free three-dimensional ultrasound imaging assessment system for scoliosis. *Sensors*, *21*.

Lee, T. T. Y., Lai, K. K. L., Cheng, J. C. Y., Castelein, R. M., Lam, T. P., & Zheng, Y. P. (2021). 3D ultrasound imaging provides reliable angle measurement with validity comparable to X-ray in patients with adolescent idiopathic scoliosis. *Journal of Orthopaedic Translation*, *29*, 51–59.

Li, Y., Zhang, S., Wang, Z., Yang, S., Yang, W., Xia, S.-T., & Zhou, E. (2021a). Tokenpose: Learning keypoint tokens for human pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 11313–11322).

Li, Y., Zhang, S., Wang, Z., Yang, S., Yang, W., Xia, S. T., & Zhou, E. (2021b). TokenPose: Learning keypoint tokens for human pose estimation. http://dx.doi.org/10.1109/ICCV48922.2021.01112.

McArthur, N., Conlan, D. P., & Crawford, J. R. (2015). Radiation exposure during scoliosis surgery: A prospective study. *Spine Journal*, *15*.

Meng, N., Wong, K. Y. K., Zhao, M., Cheung, J. P., & Zhang, T. (2023). Radiograph-comparable image synthesis for spine alignment analysis using deep learning with prospective clinical validation. *eClinicalMedicine*, *61*.

Newell, A., Huang, Z., & Deng, J. (2017). *Vol. 2017-December*, *Associative embedding: End-to-end learning for joint detection and grouping*.

Newell, A., Yang, K., & Deng, J. (2016). *LNCS: vol. 9912*, *Stacked hourglass networks for human pose estimation*.

Payer, C., Štern, D., Bischof, H., & Urschler, M. (2019). Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Medical Image Analysis*, *54*, 207–219.

shun Wong, Y., lee Lai, K. K., ping Zheng, Y., ning Wong, L. L., wah Ng, B. K., hang Hung, A. L., kei Yip, B. H., wing Chu, W. C., hung Ng, A. W., Qiu, Y., yiu Cheng, J. C., & ping Lam, T. (2019). Is radiation-free ultrasound accurate for quantitative assessment of spinal deformity in idiopathic scoliosis (IS): A detailed analysis with EOS radiography on 952 patients. *Ultrasound in Medicine & Biology*, *45*.

Simony, A., Hansen, E. J., Christensen, S. B., Carreon, L. Y., & Andersen, M. O. (2016). Incidence of cancer in adolescent idiopathic scoliosis patients treated 25 years previously. *European Spine Journal*, *25*.

Ungi, T., Greer, H., Sunderland, K. R., Wu, V., Baum, Z. M., Schlenger, C., Oetgen, M., Cleary, K., Aylward, S. R., & Fichtinger, G. (2020). Automatic spine ultrasound segmentation for scoliosis visualization and measurement. *IEEE Transactions on Biomedical Engineering*, *67*(11), 3234–3241.

Wang, R., Geng, F., & Wang, X. (2022). MTPose: Human pose estimation with high-resolution multi-scale transformers. *Neural Processing Letters*, *54*(5), 3941–3964.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2021). Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*.

Wang, D., & Zhang, S. (2022). *Vol. 2022-June*, *Contextual instance decoupling for robust multi-person pose estimation*.

Wang, C., Zhang, F., & Ge, S. S. (2021). A comprehensive survey on 2D multi-person pose estimation methods. *Engineering Applications of Artificial Intelligence*, *102*.

Wong, J., Reformat, M., Parent, E., & Lou, E. (2022). Convolutional neural network to segment laminae on 3d ultrasound spinal images to assist cobb angle measurement. *Annals of Biomedical Engineering*, *50*(4), 401–412.

Xie, H., Huang, Z., Leung, F. H., Law, N., Ju, Y., Zheng, Y.-P., & Ling, S. H. (2024). SATR: A structure-affinity attention-based transformer encoder for spine segmentation. In *2024 IEEE international symposium on biomedical imaging* (pp. 1–5). IEEE.

Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in Neural Information Processing Systems*, *35*, 38571–38584.

Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2023). Vitpose++: Vision transformer for generic body pose estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yang, D., Lee, T. T. Y., Lai, K. K. L., Lam, T. P., Chu, W. C. W., Castelein, R. M., Cheng, J. C. Y., & Zheng, Y. P. (2022). Semi-automatic ultrasound curve angle measurement for adolescent idiopathic scoliosis. *Spine Deformity*, *10*, 351–359.

Yang, S., Quan, Z., Nie, M., & Yang, W. (2021). Transpose: Keypoint localization via transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 11802–11812).

Yuan, Y., Fu, R., Huang, L., Lin, W., Zhang, C., Chen, X., & Wang, J. (2021). Hrformer: High-resolution transformer for dense prediction. arXiv preprint arXiv:2110.09408.

Zhang, J., Wang, Y., Liu, T., Yang, K., & Jin, H. (2021). A flexible ultrasound scanning system for minimally invasive spinal surgery navigation. *IEEE Transactions on Medical Robotics and Bionics*, *3*(2), 426–435.

Zheng, Y. P., Lee, T. T. Y., Lai, K. K. L., Yip, B. H. K., Zhou, G. Q., Jiang, W. W., Cheung, J. C. W., Wong, M. S., Ng, B. K. W., Cheng, J. C. Y., & Lam, T. P. (2016). A reliability and validity study for scolioscan: A radiation-free scoliosis assessment system using 3D ultrasound imaging. *Scoliosis and Spinal Disorders*, *11*.

Zhou, G. Q., Li, D. S., Zhou, P., Jiang, W. W., & Zheng, Y. P. (2020). Automating spine curvature measurement in volumetric ultrasound via adaptive phase features. *Ultrasound in Medicine & Biology*, *46*.