Measurement 224 (2024) 113955

Contents lists available at ScienceDirect

Measurement

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

Fully automatic spinal scanning and measurement based on multi-source vision information

Cui Yang [a,b], Mianjie Chen [a], Hao Xu [a], Jianyi Li [c], Qinghua Huang [d,*]

[a] School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510640, China
[b] The Guangdong Provincial Key Laboratory of Short-Range Wireless Detection and Communication, Guangzhou, 510640, China
[c] Guangdong Provincial Key Laboratory of Medical Biomechanics, Guangdong Engineering Research Center for Translation of Medical 3D Printing Application, Department of Anatomy, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515, China
[d] School of Artificial Intelligence, OPtics and ElectroNics (iOPEN), Northwestern Polytechnical University, Xi'an, 710072, China

ARTICLE INFO

Keywords:
Fully automatic 3D spine scanning
Multi-source vision information
Landmark tracking
Multitask neural networks
Scoliosis

ABSTRACT

Ultrasound is a radiation-free alternative to X-ray in the clinical evaluation of human spine. Currently proposed spinal robot-assisted scanning systems rely on single-source vision information, which may result in scan path deviation or require human intervention. Additionally, Cobb angle measurement is almost conducted based on 3D reconstruction of spine. In this article, a fully automatic scanning system is proposed based on multi-source vision information. By fusing ultrasound and RGBD information of human back, the position and pose of the robot-held probe can be determined in real-time, ensuring that vertebrae appear in the center of ultrasound images, and the tight coupling between the probe and the human skin. We also develop an automated Cobb angle measurement method based on landmark localization. The *in-vivo* and phantom experiments demonstrate that the proposed system can enable the probe to track the human spine, and offer measurement results within error range of 5°.

1. Introduction

Spine is a crucial structure responsible for facilitating various bodily movements and functions. With the development of modern society, there is a trend of high incidence of spinal disorders. Adolescent idiopathic scoliosis (AIS) is one of the most common spinal deformities that can seriously affect the physical health of adolescents and even lead to serious psychological problems. However, its pathogenesis is still not fully known, which leaves a lack of reliable methods to prevent it [1]. Therefore, the key to effective treatment of AIS is to recognize it and intervene as soon as possible.

At present, the Cobb's method has been considered as the gold standard in scoliosis diagnosis. Doctors mostly take X-rays of the patient and calculate the Cobb angle of the spine to clinically assess the degree of scoliosis, which inevitably exposes the patient to ionizing radiation. It is important to minimize the damage caused during the examination, especially for adolescents who are in the developmental stage. With the advantage of non-radiation and convenience, 3-D spinal ultrasound imaging, which has been considered as a safe alternative to X-ray for spine diagnosis, has attracted wide attention. A series of spinal ultrasound images with spatial information were obtained by free-hand scanning or robot-assisted scanning, and the vertebral landmarks in each ultrasound image could be used to assess the degree of scoliosis.

Cheung et al. [2,3] proposed a free-hand 3D spinal ultrasound imaging method and performed a validation using spinal phantoms. Zhou et al. [4,5] developed a free-hand scoliosis assessment system for scoliosis assessment in clinics. The system records the spatial information of each spinal ultrasound image through an electromagnetic sensor, and uses the coronal image of the spine, which is sliced from the 3D ultrasound volume, to manually calculate the Cobb angle. Then they proposed several methods to realize the automatic diagnosis in [6–9]. In [10], Huang et al. proposed a novel 2.5D EFOV US imaging technique of spine by free hand scanning together with a spatial sensor recording the scanning trajectory. The group also developed a novel technique to generate the 3D structure of human spine from a tracked freehand ultrasound scanning [11].

Free-hand spinal ultrasound allows the doctors to move the hand-held probe freely to acquire ultrasound images of the area of interest. However, the high inter-observer variability in US image acquisition requires trained sonographers to ensure clinically relevant images. In order to improve the automation and intelligence of the scanning, some researchers have begun to study the feasibility of the robot-assisted spinal scanning, which can undoubtedly reduce human

* Corresponding author.
E-mail addresses: yangcui@scut.edu.cn (C. Yang), lijianyi@outlook.com (J. Li), qhhuang@nwpu.edu.cn (Q. Huang).

https://doi.org/10.1016/j.measurement.2023.113955
Received 25 July 2023; Received in revised form 24 September 2023; Accepted 27 November 2023
Available online 30 November 2023
0263-2241/

intervention and lighten the workload to prevent the occupational diseases of sonographers.

Zheng et al. [12,13] proposed a robotic-ultrasound approach based on the fusion of ultrasound and force data for automatic vertebral level detection during spinal injection. A reinforcement learning based robotic navigation method in [14] and a novel dual-agent framework in [15] were proposed to automatically determine the movement of the US probe during the standard views acquisitions in spinal sonography. Huang et al. [16–18] have developed an automated 3D ultrasound imaging system, in which a three-degree-of-freedom mechanical device drives the movement of the ultrasound probe, and an RGB-D sensor is used for path planning. The same group further improve the precision and quality of 3D imaging by equipping the system with a robotic arm to achieve a 6-degree-of-freedom scanning [19]. Recently, a robot-assisted spine ultrasound imaging system by mimicking human's manipulation has been reported in order to improve the quality of B-scan images [20]. Zhang et al. [21] designed a self-adaptive attitude-adjusting mechanism to adjust the scanning posture of the US probe and finally realize the automatic scanning of the human spine.

In [22], Zheng et al. developed an automatic spinal curvature navigation system that uses a learning-based method to locate the spinous process and a force-driven controller to fine-tune the probe's orientation. However, it is not a fully automatic system. Manually locating the initial scan point is necessary for this system due to the lack of rich visual information of the patient's back, which can actually be collected using a vision sensor. Our previous work [23] utilized a neural network called "fuse-unet" to fuse optical and depth information of the human back, enabling automatic identification of the spinal region for pre-planning scanning paths. During the scanning process, the scan plane's normal vector and two pressure sensors are used to determine the scanning posture of the probe, ensuring high-quality ultrasound imaging. After scanning, the obtained ultrasound images are reconstructed in 3D, and the Cobb angle is calculated based on the reconstructed volume. Summing up the above work, the fuse-Unet we designed actually only considers the RGBD information to detect the dent of the patient's back, based on which the scanning path was determined. However, this strategy may sometimes result in an incomplete or incorrect pre-planning path, depending on the size and feature of the human body. What is more, the sonographers must manually select a series of points in the slice extracted from the reconstructed volume to calculate the Cobb angle.

To sum up, existing robotic spine scanning systems consider only single source vision information, depending only on RGBD images or ultrasound images. Therefore, those systems that only take advantage of ultrasonic images require external devices such as force sensors and require manual manipulations to locate the initial scan position. And those systems that rely solely on RGBD images are susceptible to lighting conditions and variations in the size of the subjects' body, etc.

To address these issues, in this article, we propose a real-time path planning strategy to combine the multi-source vision information provided by RGBD images and ultrasound images for automatic and more accurate scanning of the spine. Specifically, we firstly classify the view and detect landmarks of the spinal ultrasound images by a multitask neural network to track the spine in B-scans, and then calculate the real-time position and pose of the probe using the 3D point cloud obtained from RGBD images of the patient's back. As a result, we keep the spine in the center of the probe's field of view. After scanning, the Cobb angle can be automatically calculated based on the spatial information, landmarks and view classification of each spinal ultrasound image.

We now outline our key contributions of this work as follows: (1) Considering the limitation of using RGBD images and ultrasound images separately, we take advantage of both RGBD images and ultrasound images to design a fully automatic scanning method for the spine. RGBD images are used for overall planning and together with ultrasound images for real-time navigation. Both the path and the pose of the robotic-driven probe are determined in real time using multi-source vision information. (2) In contrast to some existing semi-automatic systems, we construct a fully automatic system for both scanning and Cobb angle measurement without 3D reconstruction of the spine, providing effective adjunctive measurements for doctors. (3) Realize a multi-task network to simultaneously classify the ultrasound views and detect the spinal landmarks. View classification of ultrasound images helps to facilitate navigation in case of incorrect determination of the initial scan point.

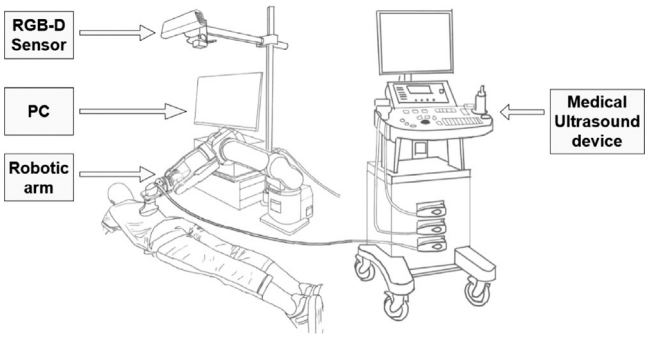


Fig. 1. The proposed system.

The remaining paper is organized as follows: in Section 2, we present an overview of our system and also give the details of our methodology, along with the experimental setup for both phantom and *in-vivo* experiments. Section 3 presents the experimental results and analysis. Section 4 summarizes the work presented in this article and discusses our future work.

2. Methodology

2.1. System design

Fig. 1 illustrates the components of the proposed system. The system is equipped with a Sonix RP medical ultrasound device from Ultrasonic Medical Corporation in Canada for obtaining ultrasound images. It also includes a convex probe (C7-3/50). Additionally, the system features a six-degree-of-freedom robotic arm (Epson C4-A601S) manufactured by Seiko Epson Corporation in Japan, known for its accuracy of up to 0.2 millimeters. Moreover, an RGBD sensor (Kinect v1) developed by Microsoft Corporation in the USA is employed to capture the RGBD images of human back. A personal computer controls and coordinates these components for seamless integration and data processing.

In the proposed system, the RGBD sensor and the RP medical ultrasound device capture the RGB images and the ultrasound images of the detection target, respectively, which are two different visual information sources for real-time planning. These two devices are designed as the eyes of a doctor in ultrasound medical scanning scenarios. The robotic arm is designed to hold the ultrasound probe and perform scanning, which plays the role of a doctor's hand. A personal computer is used to plan scanning paths in real-time based on information obtained from sensors and calculate scanning results, which plays the role of a doctor's brain. To realize automatic scanning, it is necessary to achieve good collaboration between all the independent devices. Therefore, connectivity and calibration are the foundation for precise scanning. The components communicate via TCP/IP protocol. The coordinate system of each component is defined in Fig. 2. We perform two steps of spatial calibration to achieve the transformation from the ultrasonic images to the robotic arm and also the transformation from the RGBD sensor to the robotic arm. The detailed calibration method can refer to our previous work in [23].

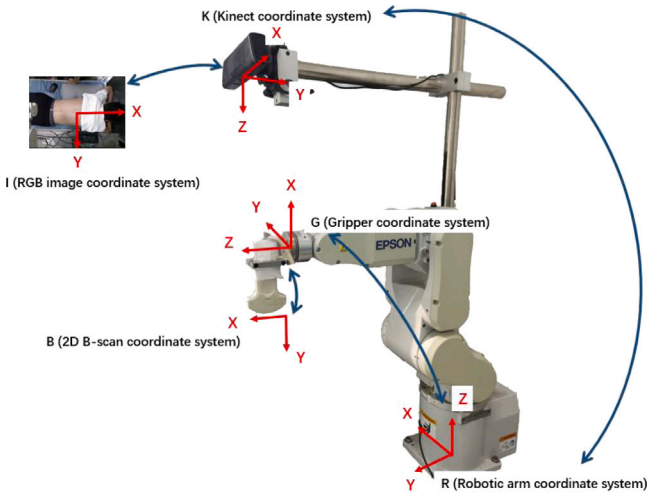


Fig. 2. The coordinate systems.

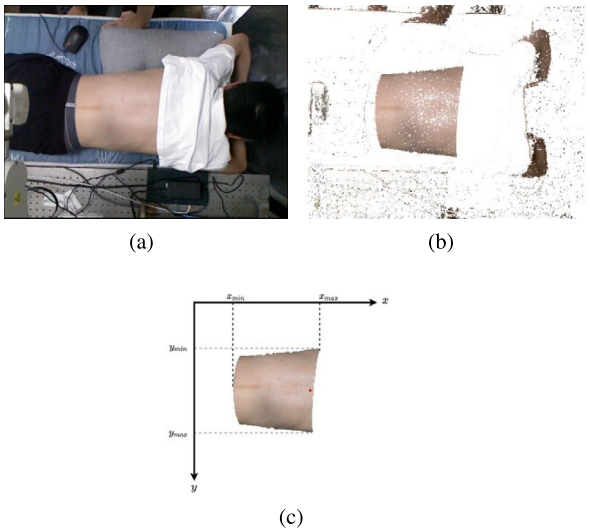


Fig. 3. Extraction of human back area. (a) Original RGB image. (b) Area satisfying the property of Red>Green>Blue. (c) The extracted area for overall planning. The red dot represents the initial scanning point. x_{min}, x_{max} are respectively the minimum and maximum values along the x axis of the human back area, and y_{min}, y_{max} are respectively the minimum and maximum values along the y axis of the human back area in the $\boldsymbol{I}$ coordinate system.

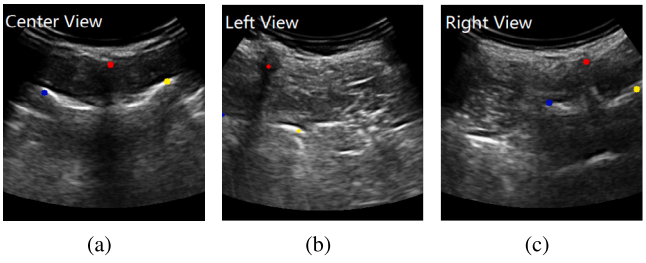


Fig. 4. Illustrations of labels for ultrasound classification and landmark detection. The red dot is TOP, the blue dot is LEFT, and the yellow dot is RIGHT.

2.2. Automatic extraction of the human back area based on RGBD images

In this section, we take full advantage of the information offered by the RGBD image to extract the human back area without user-participation, thus automatically determining the initial scan point and scan length.

Before scanning, the Kinect capture the RGBD image of the patient. We then use an experience-based rule to extract the human back area from the RGB image. Considering the skin color feature of each race, this rule first detects the human skin color, and then extracts the back area based on the different color component distributions specific to each race. As shown in Fig. 3(b), we extract the area of interest of a subject with yellow skin from the RGB image by automatically selecting pixels that satisfy the Red>Green>Blue property. The maximum connected region is recognized as the human back area, according to which we can calculate the initial scan point by $\left(x_{\max}, \frac{y_{\min}+y_{\max}}{2}\right)$, where the robot-held probe is initially placed without manual manipulation. As illustrated in Fig. 3(c), we get the scan length L by $L = x_{\max} - x_{\min}$, which helps to restrict the scan scope within the human back.

Furthermore, the human back area would also be transformed into 3D point cloud according to the transformation of the coordinate system [23] for the real-time path and pose planning.

2.3. Multi-task neural network for view classification and landmark detection of spinal ultrasound images

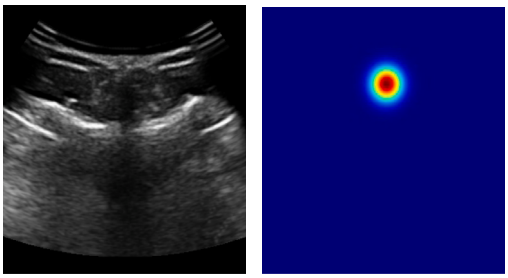
In our previous job [23], we solely relied on RGBD images for path planning, which made the algorithm heavily reliant on the physical characteristics of the subjects. Moreover, for patients with scoliosis, the position of the spine may not align perfectly with the symmetrical center of the body due to the occurrence of lateral curvature. Therefore, relying solely on surface features for scan path planning may lead to inaccuracies. Multi-sensor signals can provide more information than a single sensor, and multi-signal fusion has attracted attention in many fields [24]. Just like doctors adjust the scanning position in real time based on ultrasound images, if the system can incorporate the information provided by ultrasound images and enable the ultrasound probe to track the spinal vertebrae, ensuring that the vertebrae are present in the ultrasound images, it can greatly enhance the stability of the system. Therefore, we need to research algorithms that can automatically determine whether the spinal vertebrae are within the field of view of the ultrasound image and extract the key anatomical structure information of the vertebrae in each frame of the ultrasound image.

To do so, we have defined three types of view classifications, namely the left view, right view, and center view according to the position of the vertebra in the B-scans, as shown in Fig. 4. Specifically, if the vertebra is located on the left and right side of the ultrasound view, it is classified as a left view and a right view, respectively. If the vertebra is in the middle of the ultrasound view, it is classified as a center view.

In each view, the top of the center of the vertebra is defined as the landmark TOP. And the landmarks LEFT and RIGHT are located according to the position of the transverse process in thoracic ultrasound images and the position of the articular process in lumbar ultrasound images. We intend to design an algorithm to classify the view of each B-scan, and also detect the above anatomical structures of vertebrae.

As is widely known, ultrasound images have low contrast and intrinsic speckle noise. Moreover, With the probe placed at different positions on the vertebrae during the scan, the morphology of the vertebrae observed from the ultrasound image varies. Therefore, achieving efficient view classification and anatomical structure detection simultaneously from ultrasound images is challenging. Rather than using different networks responsible for view classification and landmark detection respectively, we adopt a multi-task network. Multi-task network can improve the performance of individual tasks as well as simplify the network architecture and reduce the time cost by fusing and sharing features between different tasks [25].

In addition, inspired by [26], unlike traditional detection algorithms, the network is designed to provide the distribution of key anatomical structures rather than a single coordinate value. We use



(a) (b)

Fig. 5. Illustrations of heatmaps regression. (a) Original spinal ultrasound image. (b) Heatmap generated by Eq. (1) at the location of landmark TOP. Blue to red represents the value from 0 to 1.

encoder–decoder architecture for heatmap regression rather than CNN for coordinate regression directly, which can drive the network to focus more on the region where the landmarks are located. As shown in Fig. 5, heatmaps are images where the Gaussian distribution is located at the position of the landmarks. The Gaussian distribution is defined as follows:

$$H(a,b)^* = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{(a-a^*)^2+(b-b^*)^2}{2\sigma^2}\right\} \tag{1}$$

where a^*, b^* are the coordinates of the landmark, $H(a,b)^*$ is the heatmap of the landmark, σ is the variance of the heatmap.

As shown in Fig. 6, to avoid over fitting and also reduce the time cost of the network, VGG11 [27], which is an effective neural network with fewer layers, is chosen as the feature extractor for our scenario. We adopt the FPN [28] as an encoder–decoder structure to realize heatmap regression, because it can fuse the multi-scale information of the image to promote landmark detection. We implement the multi-task network by sharing multiple convolutional layers in classification and landmark detection tasks. This network structure is able to achieve a balance between time cost and performance, so that both real-time and stability can be guaranteed.

2.4. Real-time path and pose planning based on multi-source vision information

After extracting information from ultrasound and RGBD images, we need to fuse the above information to calculate the position and orientation of the scan in real-time, ensuring that the vertebrae are always within the ultrasound view and also ensuring a tight coupling between the probe and the skin. The proposed probe control strategy is based on multi-source visual information, which integrates RGBD and ultrasound images.

Let P_{Rj}^{path} and P_{Rj}^{pose} separately represent the position and pose of the probe for current scan point. By analyzing the ultrasound image obtained at the current scan point, the position of the probe for next scan point P_{Rj+1}^{path} can be dynamically planned, using the following two steps.

Step1: Identify the target in the ultrasound image that needs to be positioned at the center of the ultrasound view. The center of the vertebrae detected from the current ultrasound image is defined as $P_{Bj}^{CenterV} = (x_{Bj}, y_{Bj})$, which can be achieved by considering both view classification and landmark detection results of ultrasound images. Let $P_B^{CenterI}$ denote the center of the ultrasound view, which equals to the coordinate value (161, y) given the resolution of the ultrasound image is 322*356.

To make $P_{Bj}^{CenterV}$ approach $P_B^{CenterI}$, that is, to center the vertebrae in the ultrasound view, we have so far considered a simple proportional–derivative (PD) controller. The proposed PD controller, expressed by Eq. (2), can improve the stability of the system and make human subjects feel more comfortable during scanning.

$$\begin{aligned} x_{P_{Bj}} &= x_{P_{Bj}^{CenterV}} + K_p e(k) + K_d[e(k)-e(k-1)] \\ y_{P_{Bj}} &= 0 \\ e(k) &= x_{P_B^{CenterI}} - x_{P_{Bj}^{CenterV}} \end{aligned} \tag{2}$$

where K_p and K_d are the proportional and derivative coefficients, respectively. $x_{P_{Bj}}$ is the output of the PD controller, where the imaging will appear in the center of the ultrasonic view at the next scan point.

Step2: Obtain the position of the probe based on ultrasound and RGBD information. Given P_{Bj}, the position of the probe for the next scan point can be calculated as follows.

$$P_{Ij} = (x_{Ij}, y_{Ij}) = {}^{I}T_K\,{}^{K}T_R\,{}^{R}T_G\,{}^{G}T_B P_{Bj} \tag{3}$$

$$P_{I(j+1)} = (x_{I(j+1)}, y_{I(j+1)}) = (x_{Ij}\text{-}1, y_{Ij}) \tag{4}$$

$$P_{R(j+1)}^{path} = {}^{R}T_K\,{}^{K}T_I P_{I(j+1)} \tag{5}$$

The whole conversion from B to I is shown in Eq. (3), where ${}^{I}T_K$ denotes the transformation matrix from coordinate system K to I and each coordinate system is illustrated in Fig. 2.

Fig. 7 illustrates the strategy of real-time path and pose planning. According to Eq. (3), the coordinates of P_{Bj} can firstly be transformed to the corresponding values in coordinate system I. Then the coordinates of the next scan point in I (i.e. $P_{I(j+1)}$) can be obtained using Eq. (4). The position of the next scan point $P_{R(j+1)}^{path}$, where the probe should be placed, can be calculated using Eq. (5).

To determine $P_{R(j+1)}^{pose}$ and keep the probe in close contact with the skin during scanning, we use the real-time normal-vector-based method, which was proposed in our previous work [23] and based on the information provided by RGBD images.

After the probe pose calculation is completed, the robotic arm will be driven to $P_{R(j+1)}$ until the scan length is met.

2.5. Automatic Cobb angle measurement without 3D reconstruction

The method proposed in this section does not require human intervention and requires less computation than our previous method in [23] since this method is no longer based on the spinal volume and does not need to realize 3D reconstruction. During scanning, the spatial information, landmarks, and view classification of each ultrasound image are collected, which would contribute to the automatic Cobb angle measurement.

Fig. 8 illustrates the flow of the automatic Cobb angle measurement algorithm. The landmarks TOP in the Center views are extracted and they are projected to the coordinates under the coordinate system I according to Eq. (3). We fit these points with a 6-th order polynomial curve which has been proven to be the best approximation to the physiological curvature of the spine in [29,30]. Then, the stationary points of the curve, where the partial derivatives are zero, should be obtained. Finally, after fitting the adjacent stationary points with straight lines, the slope of each straight line, i.e. k_i for the ith straight line, can be obtained. And the Cobb angle is then calculated using Eq. (6).

$$\begin{aligned} CobbAngle &= \arctan\left(\frac{k_{i-1}-k_i}{1+k_{i-1}\cdot k_i}\right) \\ i &= 2,\ldots,n-1 \end{aligned} \tag{6}$$

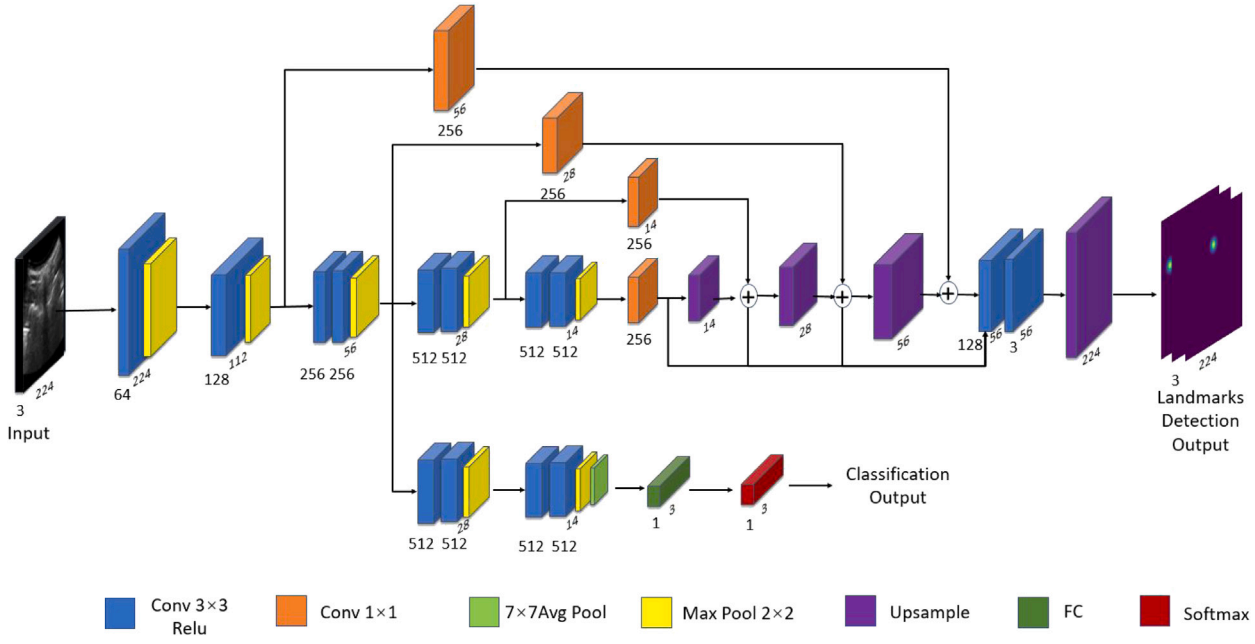


Fig. 6. Architecture of the Multi-task VGG11-FPN.

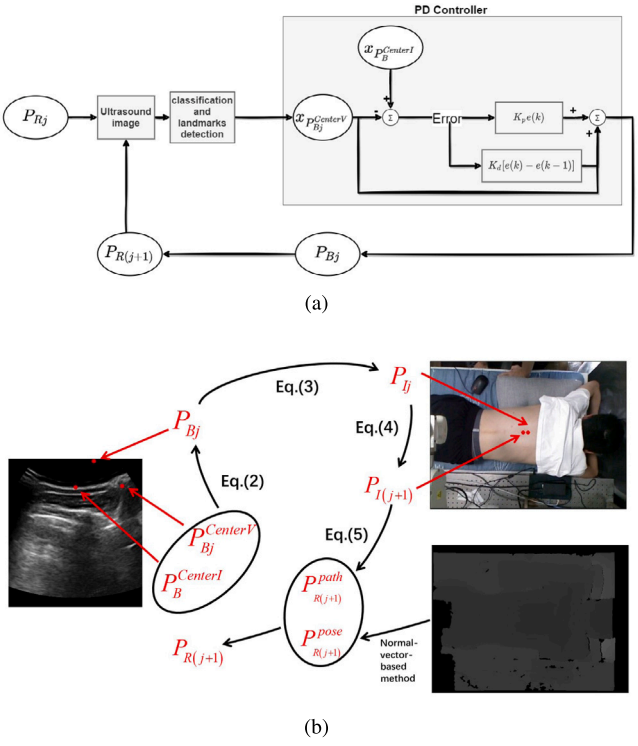


Fig. 7. Real-time path and pose planning based on multi-source vision information.

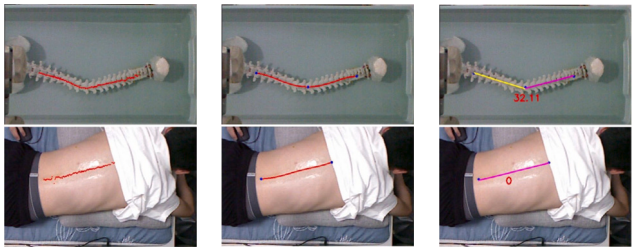


Fig. 8. Automatic measurement of Cobb angles. (a) Landmark TOP of the Center view. (b) The fitting curve and stationary points (Blue dots). (c) The fitting lines and the results of the automatic calculation.

2.6. Experimental methods

In our work, phantom and *in-vivo* experiments are performed for quantitative and qualitative analysis of automatic scanning and Cobb angle measurement.

The original image resolution provided by the ultrasound system is 488*356, and we crop it to 322*356 after extracting the region of interest. The scan depth of the ultrasound system is set to be 70 mm in all experiments. K_p and K_d of the PD controller are set to be 0.6 and 0.1, respectively.

To train the network, a total of 3298 spinal phantom ultrasound images and 4818 human spinal ultrasound images containing all the views mentioned in Section 2 are collected using the convex probe. The human data collection is based on five young and healthy male volunteers. To ensure diversity in the data, the body types of the five subjects included lean, medium and slightly overweight types. During the data collection process for each volunteer, ultrasound images of different view categories are collected by adjusting the probe to scan different regions of the spine in multiple scans. And the annotation is done by experienced ultrasound readers under the guidance of sonographers. In the annotation process of vertebral landmarks, the first step is to manually locate the positions of vertebral landmarks on ultrasound images with a size of 322 × 356 and obtain the coordinates of each landmark. Then, before inputting the network, corresponding heatmaps of size 224 × 224 are generated for each landmark according to Eq. (1). The σ in Eq. (1) is set to be 10. The size of the original ultrasound images is scaled to a uniform size of 224*224 before input to the network and restored after output.

To validate the performance improvement of multi-task networks and heatmap regression-based landmark detection, we conduct a control experiment using a single-task network and coordinate regression-based method for performance comparison with our network. Meanwhile, to examine the suitability of VGG11 as a feature extractor in

this scenario, classic lightweight networks such as AlexNet [31] and ResNet18 [32] will be selected as feature extractors to construct multi-task networks with the same architecture as our network, and their performance will be evaluated. Additionally, we use several common metrics to evaluate the detection and classification accuracy and real-time performance of the proposed network. And the comparison with other networks is conducted to verify the performance of Multi-task VGG11-FPN through ten-fold cross-validation.

We implement the network using Pytorch. The parameters of the network are randomly initialized and the pre-trained shared parameters are kept locked while another task is trained. The classification task is optimized by Adam with a learning rate of 2e−4 and the loss function is the cross entropy. And the landmark detection task uses the Adam optimizer with a learning rate of 3e−3 and the mean square error loss function. The network is trained on an NVIDIA GeForce RTX 3090 GPU(24 GB) and run on a personal computer with an NVIDIA GeForce RTX 1050 GPU(4 GB) during scanning.

Point Error(PE) and Successful Detection Rate(SDR) are used to evaluate the performance of landmark detection. PE of the nth landmark is calculated as follows:

$$PE_n = \left\| L_n^* - L_n \right\|_2 \tag{7}$$

where $n = 1, 2 \ldots M$, and M represents the number of total landmarks throughout all the ultrasound images. L_n^* and L_n represent the ground truth and the prediction result for the nth landmark. The SDR is defined as follows:

$$SDR_{dist} = \left\{ m : \left\| L_n^* - L_n \right\|_2 \le dist \right\} / M \times 100\ \% \tag{8}$$

Eq. (8) gives the percentage of the images in which landmarks are successfully detected (i.e. m) out of a total of M landmarks. The parameter $dist$ indicates the scope of successful detection. When the distance between the ground truth and the prediction result is within $dist$ in millimeters, the landmark is considered to be detected successfully. We also use accuracy to evaluate the performance of view classification, and use FPS(frames per second) as a metric to evaluate the real-time performance of the network which represents the number of images that can be processed per second.

In the automatic spine scanning and Cobb angle measurement experiment, we scan the spinal phantom 10 times for each of the three types of scoliosis. To be specific, those with a Cobb angle less than 20° are classified as mild(<20°), those with a Cobb angle between 20° and 40° are classified as moderate(20°–40°), and those larger than 40° are classified as severe(>40°). These data are used to quantitatively evaluate the automatic scanning and measurement, and to assess the reliability of the system for scenarios of different types of scoliosis. For each type of scoliosis, we randomly bend the spinal phantom into a certain angle within its corresponding range. The actual Cobb angle of the spinal phantom is measured against the gold standard measurement on X-ray and recorded prior to scanning. Then, the spinal phantom is placed in a water tank and covered with a paper simulating the human back to obtain the scanning range before water injection. Then we perform the scanning and measurement experiment to acquire the measured Cobb angle. The whole process repeat for ten times. After all the scans of ten independent experiments for each group, the statistical MAE (Mean Absolute Error), MRE (Mean Relative Error) and Error Range are calculated to indicate the angle errors between the measured values and the corresponding true values.

In our *in-vivo* experiment, we applied liquid couplant to the subject's spinal area and US probe before scanning. The performance of the real-time path planning is evaluated by varying the subject's posture and scanning along different spinal curves, after which 3D visualization of spinal landmarks will be completed to qualitatively evaluate the automatic scanning.

Table 1
Measurement results in phantom experiment. Actual angle, measured angle and MAE: statistical mean and standard deviation of ten independent experimental results for each type are calculated and presented in the form of Mean ±Standard Deviation. MRE: statistical ratio of measurement errors to corresponding true angles is presented. Error range: upper and lower limits of errors observed in ten independent experiments are presented.

Measurement	Mild (<20°)	Moderate (20°~40°)	Severe (>40°)
Actual angle (°)	14.52 ± 2.90	30.43 ± 7.93	50.22 ± 7.45
Measured angle (°)	14.81 ± 2.88	28.78 ± 8.44	49.73 ± 7.04
MAE (°)	1.10 ± 0.95	1.68 ± 1.23	1.22 ± 0.79
MRE(%)	7.94%	6.29%	2.42%
Error range (°)	0.03~3.06	0.02~3.56	0.16~2.98

3. Results

3.1. Results of phantom experiment

We train the proposed network with 3298 ultrasound images of spinal phantom and perform the automatic scanning and measurement on spinal phantom. The results are presented in Fig. 9. The black line in Fig. 9(a) represents the central line of the back area determined by RGB images, which is the central line of the paper used to model the human back. The red line represents the scan path of the robotic arm, which is determined by real-time path planning methods during scanning. As depicted in Fig. 9(a), the system is able to automatically and accurately track the spine, thus it is more conducive to the calculation of Cobb angles. Fig. 9(b) displays all the landmarks detected during the scanning process. The yellow dots represent the landmark RIGHT, the red dots represent the landmark TOP, and the blue dots represent the landmark LEFT. From the figure, it can be seen that the overall accuracy of landmarks detection is relatively high, which can present the contour of the spine. Fig. 9(c) shows a three-dimensional display of all the vertebral landmarks, which makes it possible to observe the ups and downs of the spine and the height and depth information of the vertebral key structures. And we can see that the path of the landmark TOP in Fig. 9(b) and the fitting curve in Fig. 9(d) are more stable and accurate, so the path of TOP is chosen to calculate the cobb angles and the results are shown in Fig. 9(e).

Besides, based on the data collected by automatic scanning, we calculate the cobb angles using the automatic measurement algorithm and then conduct statistical analysis between actual angles and measured angles. Table 1 gives the statistical measurement results of three different scoliosis types, i.e. mild, moderate and severe, respectively. Given the actual angles, measured angles in ten independent experiments, both mean and standard deviation of absolute error (MAE) are calculated and they are represented in the form of Mean ±Standard Deviation in Table 1. As we can see in Table 1, MAE and MRE are not much significant and the measured angles accord with the actual angles within the allowable error range of 5°, proving that automatic Cobb angle measurement algorithm can be applied to the accurate measurement of different degrees of scoliosis. In general, our system is able to realize the reliable automatic spinal scanning and Cobb angle measurement, thereby assisting doctors in the diagnosis of spine-related diseases.

3.2. Results of *in-vivo* experiment

Since the human spinal ultrasound images are more complex than phantom's. Therefore, we evaluate the performance of our network on human dataset comprising 4818 images based on ten-fold cross-validation.

The experimental results are shown in Table 2. As we can see, in the comparison between Multi-task VGG11-FPN, Multi-task AlexNet-FPN and Multi-task ResNet18-FPN, although AlexNet achieved a better

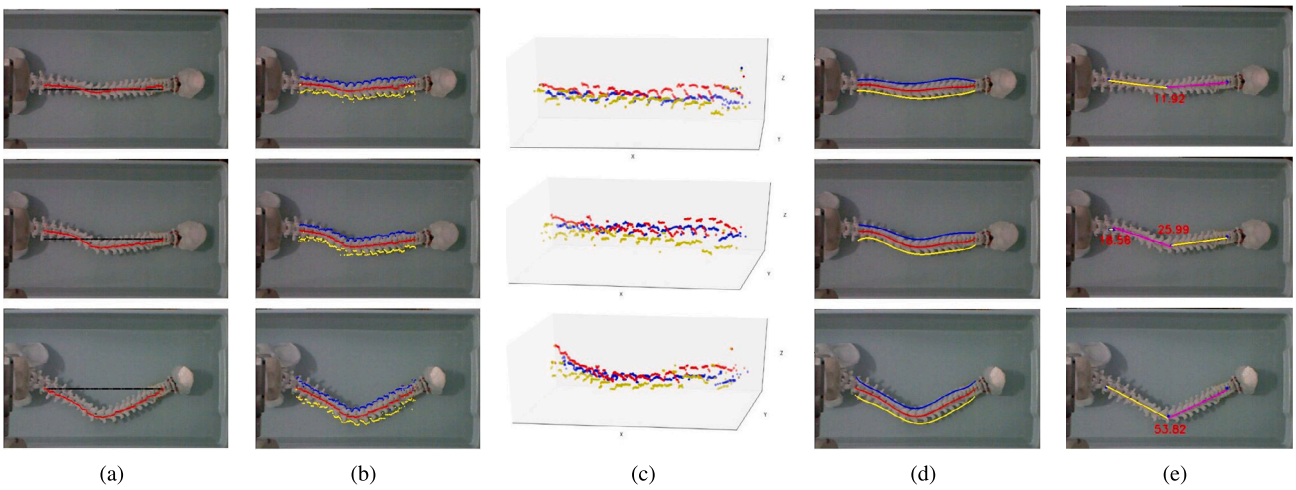


Fig. 9. Automatic scanning and measurement results of phantom. (a) Scan path (extracted based on ultrasonic landmark tracking, red line) and Central line of the back area (extracted from the RGB image, black line). (b) The detected landmarks in I. (c) The detected landmarks in R. (d) Fitting curves. (e) The results of automatic measurement.

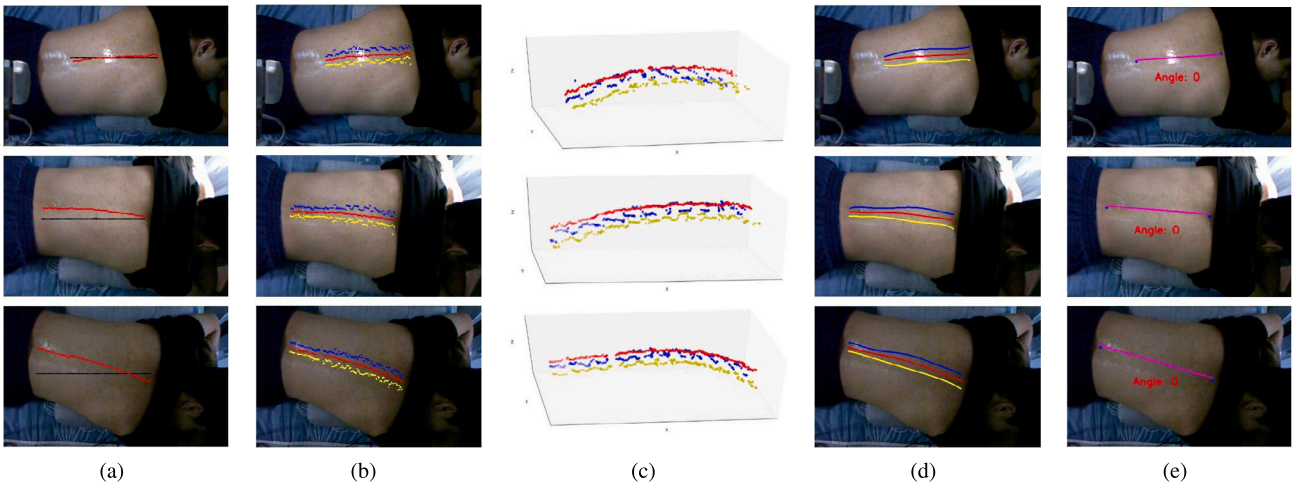


Fig. 10. Automatic scanning and Cobb angle measurement results under different path in *in-vivo*. (a) Scan path(red line) and Central line of the back area(black line). (b) The detected landmarks in I. (c) The detected landmarks in R. (d) Fitting curves. (e) The results of automatic measurement.

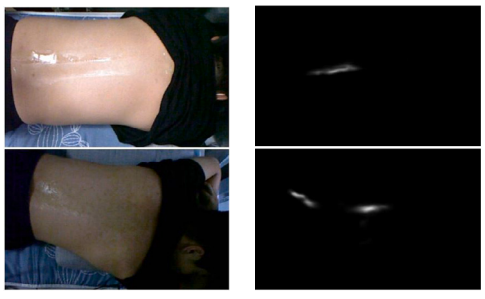


Fig. 11. Examples of incomplete or incorrect segmentation results output from fuse-Unet, which only considered RGBD images [23]. (a) The acquired RGB images before scanning. (b) The corresponding semantic segmentation results.

FPS of 104 as a feature extractor compared to VGG11 employed in a multi-task network, its classification accuracy and SDR_5 are both lower than those of Multi-task VGG11-FPN by 0.3% and 0.97%, respectively. When ResNet18 is used as the feature extractor, it achieves a 0.1% improvement in classification accuracy over Multi-task VGG11-FPN, but SDR_5 and FPS are both decreased. Considering the importance of landmark detection and real-time performance in scanning path planning, Multi-task VGG11-FPN outperforms Multi-task ResNet18-FPN in this system.

In addition, the comparison with networks Single-task VGG11 and Single-task VGG11-FPN has demonstrated the improved performance achieved by the heatmap regression method and multi-task learning. Compared to the direct regression of landmark coordinates using the Single-task VGG11 network, Multi-task VGG11-FPN achieved a 5.16% increase in SDR_5. For view classification, the multi-task approach outperformed Single-task VGG11 alone by 0.8% in accuracy. However, when Single-task VGG11-FPN was used for landmark detection alone, its SDR_5 decreased by 0.43%. These experimental results have shown that VGG11 as a feature extractor is capable of overcoming the limitations of medical image datasets and meeting real-time requirements in this scenario, while indicating that network performance can be improved through multi-task learning and the use of the heatmap regression method. As a consequence, Multi-task VGG11-FPN is selected in the proposed system.

We finally conduct an *in-vivo* experiment to evaluate the proposed system. Fig. 10 shows the automatic scanning and Cobb angle measurement results on different postures of same healthy subject based on multi-source vision information. The average time cost of automatic scanning is about 120 s and the diagnostic results can be presented by our software as soon as the scanning is finished.

Table 2
Performance comparison of Multi-task VGG11-FPN with different models on the collected human dataset. The classification accuracy is determined by computing the proportion of correctly classified image frames in ultrasound images relative to total number of ultrasound images used for prediction. *PE* (mm) is the statistical result for all landmarks and presented in the form of Mean ±Standard Deviation.

Network	Heatmap regression	Multi-task	Classification accuracy	*PE* (mm)	SDR_5	FPS
Multi-task AlexNet-FPN	√	√	99.1%	2.37 ± 2.92	92.67%	104
Multi-task ResNet18-FPN	√	√	99.5%	2.19 ± 2.82	93.62%	34
Single-task VGG11	×	×	98.6%	2.67 ± 2.32	88.48%	175
Single-task VGG11-FPN	√	×	–	2.20 ± 2.93	93.21%	70
Multi-task VGG11-FPN	√	√	99.4%	2.21 ± 2.93	93.64%	52

In the previous proposed system [23], we only consider the RGBD information to detect the dent of the patient's back, which leads to the problem that the automatic spinal area segmentation results may be affected by the light interference or the feature of human body. Fig. 11 gives several examples of incomplete or incorrect segmentation results in the system only based on RGBD images, which leads to the deviation of the pre-planning path from the spine area. Compared with the results in Fig. 11, the satisfactory scanning results obtained based on multi-source vision information are depicted in Fig. 10(a). As we can see, although the change of the subject's posture leads to the change of the initial scan point and scan length, our system is eventually able to accurately locate the position of the spine and complete the scan.

Three landmarks of Center View detected by Multi-task VGG11-FPN were transformed to 2D RGB image (Fig. 10(b)) and 3D image (Fig. 10(c)), respectively, from which we can see the general outline of the human spine. We use the sixth-order curve to fit three landmarks respectively in Fig. 10(d). And we can see that the path of the landmark TOP in Fig. 10(b) is smoother and more stable for the reason that the SDR_5 of landmark TOP can reach 99.38% while the SDR_5 of landmarks LEFT and RIGHT are only 90.71% and 90.8% respectively. So the path of TOP is chosen to measure the Cobb angles and the results are shown in Fig. 10(e).

The experimental results show that our system is able to offer a wealth of spinal information to the sonographers after the accurate and automatic scanning, which indicates the potential clinical application value of our system.

4. Conclusion

In this article, we have developed a robotic ultrasound system to realize automatic spinal scanning and diagnosis of scoliosis, which can also provide rich information of the spine in 2D and 3D. To evaluate the performance of the proposed system, we conducted both phantom experiments and *in-vivo* experiments. The proposed multi-task network can achieve classification accuracy of 99.4%, landmark detection accuracy of 93.64% and processing speed of 52 FPS, which indicate that our proposed algorithm can achieve a compromise between performance and computation. The automatic scanning can be realized in real time of single scanning time within 120s. The *in-vivo* experiment have validated that the designed system can enable the ultrasound probe to track the human spine with a tight coupling between the probe and human skin, without human intervention. The analytical analysis of the phantom experiments demonstrated that the estimation error of our automatic Cobb measurement algorithm is within 5°, which is clinically acceptable.

Although the experiments demonstrate that the system is capable of accurately performing the described functions, there are still some potential limitations. Therefore, more aspects are worth considering to drive our system towards clinical application.

One aspect is about the technologies that can be further developed. Firstly, the current *in-vivo* experiment was conducted on healthy young male volunteers, in future work, more data of scoliosis patients should be collected, experiments should be conducted on real patients, and further improvements to our system should be made based on clinical needs. Second, the gold standard for identifying scoliosis is standing anteroposterior (AP) and lateral scoliosis x-ray films. Scanning with the patient standing and subject to gravity is beneficial to improve the diagnostic accuracy [33]. So changing the subject's posture to a standing posture will improve the clinical practicality of the system. Last but not least, the SDR_5 of landmark TOP reaches 99.38% while the SDR_5 of landmarks LEFT and RIGHT are 90.71% and 90.8% respectively, which requires that we should attempt to take some measures to improve the detection accuracy of landmarks LEFT and RIGHT, such as introducing the learning of the spatial locations of the three landmarks into the neural network.

The other aspect we should further consider is the ethics of a robotic system, including but not limited to system safety, patient privacy protection, medical liability and medical service fairness. For the safety of the proposed system, we imposes restrictions on the range of motion of the robotic arm during scanning activities. Furthermore, each joint of the robotic arm operates within a predefined range of force. In the event that a body-safe force threshold is exceeded, the system will engage an emergency brake mechanism. This is implemented to ensure that the system avoids causing harm to individuals undergoing scanning procedures.

In summary, according to the literature we surveyed [22,23,34,35], this is the first time that fully automated scanning of the spine and automatic measurement of the Cobb angle have been achieved. Automatic recognition of the spinal region and classification and landmark detection of spinal ultrasound images without user participation were realized in our system. The system combines RGBD information and ultrasound information to realize automatic and real-time path planning, and provides rich 2D and 3D information of the spine. We believe that this system can assist sonographers in spinal ultrasound scanning and diagnosis and provide clearer, more intuitive spinal ultrasound information.

CRediT authorship contribution statement

Cui Yang: Conceptualization, System design, Methodology, Writing. **Mianjie Chen:** Investigation, Data collection, Software, Experiments, Writing – original draft. **Hao Xu:** Experiments. **Jianyi Li:** Implementation guide. **Qinghua Huang:** Supervision, Writing – reviewing & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No. 2022YFF1202600), the National Natural Science Foundation of China under grant 62071382, the Guangdong Provincial Key Laboratory of Short-Range Wireless Detection and Communication (Grant No. 2017B030314003) and the Guangdong Basic and Applied Basic Research Foundation (2022A1515010167, 2022A1515011604, 2020A1515010962) and the Science and Technology Planning Project of Guangzhou (202002030251).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.measurement.2023.113955.

References

[1] Jack C Cheng, René M Castelein, Winnie C Chu, Aina J Danielsson, Matthew B Dobbs, Theodoros B Grivas, Christina A Gurnett, Keith D Luk, Alain Moreau, Peter O Newton, et al., Adolescent idiopathic scoliosis, Nat. Rev. Dis. Prim. 1 (1) (2015) 1–21.

[2] Chung-Wai James Cheung, Siu-Yin Law, Yong-Ping Zheng, Development of 3-D ultrasound system for assessment of adolescent idiopathic scoliosis (AIS): and system validation, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2013, pp. 6474–6477.

[3] Chung-Wai James Cheung, Guang-Quan Zhou, Siu-Yin Law, Ka-Lee Lai, Wei-Wei Jiang, Yong-Ping Zheng, Freehand three-dimensional ultrasound system for assessment of scoliosis, J. Orthopaedic Transl. 3 (3) (2015) 123–133.

[4] Chung-Wai James Cheung, Guang-Quan Zhou, Siu-Yin Law, Tak-Man Mak, Ka-Lee Lai, Yong-Ping Zheng, Ultrasound volume projection imaging for assessment of scoliosis, IEEE Trans. Med. Imaging 34 (8) (2015) 1760–1768.

[5] Yong-Ping Zheng, Timothy Tin-Yan Lee, Kelly Ka-Lee Lai, Benjamin Hon-Kei Yip, Guang-Quan Zhou, Wei-Wei Jiang, James Chung-Wai Cheung, Man-Sang Wong, Bobby King-Wah Ng, Jack Chun-Yiu Cheng, et al., A reliability and validity study for Scolioscan: a radiation-free scoliosis assessment system using 3D ultrasound imaging, Scoliosis Spinal Disord. 11 (2016) 1–15.

[6] Guang-Quan Zhou, Wei-Wei Jiang, Ka-Lee Lai, Yong-Ping Zheng, Automatic measurement of spine curvature on 3-D ultrasound volume projection image with phase features, IEEE Trans. Med. Imaging 36 (6) (2017) 1250–1262.

[7] Guang-Quan Zhou, Dong-Sheng Li, Ping Zhou, Wei-Wei Jiang, Yong-Ping Zheng, Automating spine curvature measurement in volumetric ultrasound via adaptive phase features, Ultrasound Med. Biol. 46 (3) (2020) 828–841.

[8] Zixun Huang, Rui Zhao, Frank HF Leung, Kin-Man Lam, Sai Ho Ling, Juan Lyu, Sunetra Banerjee, Timothy Tin-Yan Lee, De Yang, Yong-Ping Zheng, DA-GAN: Learning structured noise removal in ultrasound volume projection imaging for enhanced spine segmentation, in: 2021 IEEE 18th International Symposium on Biomedical Imaging, ISBI, IEEE, 2021, pp. 770–774.

[9] Rui Zhao, Zixun Huang, Tianshan Liu, Frank HF Leung, Sai Ho Ling, De Yang, Timothy Tin-Yan Lee, Daniel PK Lun, Yong-Ping Zheng, Kin-Man Lam, Structure-enhanced attentive learning for spine segmentation from ultrasound volume projection images, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2021, pp. 1195–1199.

[10] Qinghua Huang, Qifeng Deng, Le Li, Junlin Yang, Xuelong Li, Scoliotic imaging with a novel double-sweep 2.5-dimensional extended field-of-view ultrasound, IEEE Trans. Ultrasonics Ferroelectr. Freq. Control 66 (8) (2019) 1304–1315.

[11] Qinghua Huang, Hao Luo, Cui Yang, Jianyi Li, Qifeng Deng, Peng Liu, Maoqing Fu, Le Li, Xuelong Li, Anatomical prior based vertebra modelling for reappearance of human spines, Neurocomputing 500 (2022) 750–760.

[12] Maria Victorova, David Navarro-Alarcon, Yong-Ping Zheng, 3D ultrasound imaging of scoliosis with force-sensitive robotic scanning, in: 2019 Third IEEE International Conference on Robotic Computing, IRC, IEEE, 2019, pp. 262–265.

[13] Maria Tirindelli, Maria Victorova, Javier Esteban, Seong Tae Kim, David Navarro-Alarcon, Yong Ping Zheng, Nassir Navab, Force-ultrasound fusion: Bringing spine robotic-us to the next "level", IEEE Robot. Autom. Lett. 5 (4) (2020) 5661–5668.

[14] Hannes Hase, Mohammad Farid Azampour, Maria Tirindelli, Magdalini Paschali, Walter Simson, Emad Fatemizadeh, Nassir Navab, Ultrasound-guided robotic navigation with deep reinforcement learning, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2020, pp. 5534–5541.

[15] Keyu Li, Yangxin Xu, Jian Wang, Dong Ni, Li Liu, Max Q-H Meng, Image-guided navigation of a robotic ultrasound probe for autonomous spinal sonography using a shadow-aware dual-agent framework, IEEE Trans. Med. Robot. Bionics 4 (1) (2021) 130–144.

[16] QingHhua Huang, Jiulong Lan, Xuerong Li, Automatic ultrasound scanning system based on robotic arm, Sci. China Inf. Sci. 62 (5) (2019) 50211:1–50211:3.

[17] Qinghua Huang, Bowen Wu, Jiulong Lan, Xuelong Li, Fully automatic three-dimensional ultrasound imaging based on conventional B-scan, IEEE Trans. Biomed. Circuits Syst. 12 (2) (2018) 426–436.

[18] Zhaohong Chen, Yongdong Chen, Qinghua Huang, Development of a wireless and near real-time 3D ultrasound strain imaging system, IEEE Trans. Biomed. Circuits Syst. 10 (2) (2015) 394–403.

[19] Qinghua Huang, Jiulong Lan, Xuelong Li, Robotic arm based automatic ultrasound scanning for three-dimensional imaging, IEEE Trans. Ind. Inform. 15 (2) (2018) 1173–1182.

[20] Qinghua Huang, Yingchen Wang, Hao Luo, Jianyi Li, On mimicking human's manipulation for robot-assisted spine ultrasound imaging, Robot. Intell. Autom. 2023 (2023) 373–381.

[21] Jingyu Zhang, Tianjian Liu, Yunjiang Wang, Weiwei Jiang, Keji Yang, Haoran Jin, Yongjian Zhu, Self-adaptive ultrasound scanning system for imaging human spine, IEEE Trans. Ind. Electron. 69 (1) (2021) 570–581.

[22] Maria Victorova, Michael Ka-Shing Lee, David Navarro-Alarcon, Yongping Zheng, Follow the curve: Robotic ultrasound navigation with learning-based localization of spinous processes for scoliosis assessment, IEEE Access 10 (2022) 40216–40229.

[23] Cui Yang, Mingyao Jiang, Mianjie Chen, Maoqing Fu, Jianyi Li, Qinghua Huang, Automatic 3-D imaging and measurement of human spines with a robotic ultrasound system, IEEE Trans. Instrum. Meas. 70 (2021) 1–13.

[24] Cui Jian, Xie Ping, Wang Xiao, Wang Jing, He Qun, Jiang Guoqian, M2FN: An end-to-end multi-task and multi-sensor fusion network for intelligent fault diagnosis, Measurement 204 (30) (2022) 1–12.

[25] Zhoubing Xu, Yuankai Huo, JinHyeong Park, Bennett Landman, Andy Milkowski, Sasa Grbic, Shaohua Zhou, Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II. Vol. 11, Springer, 2018, pp. 711–719.

[26] Jingyuan Xu, Hongtao Xie, Chuanbin Liu, Fang Yang, Sicheng Zhang, Xun Chen, Yongdong Zhang, Hip landmark detection with dependency mining in ultrasound image, IEEE Trans. Med. Imaging 40 (12) (2021) 3762–3774.

[27] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[28] Alexander Kirillov, Ross Girshick, Kaiming He, Piotr Dollár, Panoptic feature pyramid networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6399–6408.

[29] Jakub Krejci, Jiri Gallo, Petr Stepanik, Jiri Salinger, Optimization of the examination posture in spinal curvature assessment, Scoliosis 7 (2012) 1–9.

[30] Tri Arief Sardjono, Michael HF Wilkinson, Albert G Veldhuizen, Peter MA van Ooijen, Ketut E Purnama, Gijsbertus J Verkerke, Automatic Cobb angle determination from radiographic images, Spine 38 (20) (2013) E1256–E1262.

[31] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[33] Patrick Knott, Eden Pappo, Michelle Cameron, Jean Claude deMauroy, Charles Rivard, Tomasz Kotwicki, Fabio Zaina, James Wynne, Luke Stikeleather, Josette Bettany-Saltikov, et al., SOSORT 2012 consensus paper: reducing x-ray exposure in pediatric patients with scoliosis, Scoliosis 9 (1) (2014) 4.

[34] Qinghua Huang, Jiakang Zhou, ZhiJun Li, Review of robot-assisted medical ultrasound imaging systems: technology and clinical applications, Neurocomputing 559 (2023) 126790.

[35] Jiakang Zhou, Haozhe Tian, Wei Wang, Qinghua huang, Fully automated thyroid ultrasound screening utilizing multi-modality image and anatomical prior, Biomedical Signal Processing and Control (ISSN: 1746-8094) 87 (2024) 105430.