A Vision-based Target Localization Method for Robot-assisted Sonography

Jinyan Tian, Weiyong Si, Ning Wang, Chenguang Yang*

***Abstract*—This article proposes a vision-based ultrasound scanning localization method that solves the issue of human pose constraints. Specifically, by proposing an improved AlphaPose network, the neck positioning for individuals with different body types is addressed. The interference caused by other moving human subjects in the scene is resolved through an improved DeepSORT tracking algorithm. The impact of environmental noise interference is reduced through depth information processing techniques. Additionally, a robotic platform for neck ultrasound scanning is constructed. This system is capable of automatically identifying and localizing moving human necks, planning trajectories for the ultrasound probe approach, and achieving real-time interaction.**

***Index Terms*—Robotic ultrasound system, Scanning target localization, Robot vision**

I. Introduction

Manual ultrasound examination is widely used in the field of medical diagnosis due to its non-invasive nature, safety, and real-time capability. In order to improve the repeatability of ultrasound examination, shorten the examination time (at least 20 minutes each time [1]), and reduce excessive reliance on ultrasound physician experience [2], we need to introduce corresponding technologies for improvement. This can increase the examination accuracy and consistence, and improve examination efficiency [3], [4].

Robotic ultrasound (RUS) uses a robotic arm equipped with an ultrasound scanner to perform ultrasound scanning(US), and has been shown to be a viable solution to these problems through human-robot collaboration for ultrasound scanning tasks [5]. Recently, RUS has been actively studied to solve the application of medical examination in various clinical settings, including lung [6], cardiac [7] and spine [8] imaging. Also, RUS has be used in the diagnosis and evaluation of thyroid nodules [9], which could identify small lesioned nodules that are not noticed during physical examination [10].

This paper is dedicated to the study of automatic ultrasonic region identification and localization techniques in RUS technology. The RUS localization process is usually achieved by teleoperation ultrasound robots based on shared control technology [11], [12]. Although it relieves the technician, it has learning costs and sometimes results in task failure due to oddities in the robot arm joints [13]. In recent years, robot-assisted ultrasound scanning has applied traditional vision methods for feature extraction to achieve localization in specific scanning regions [14]. Some scholars have used RGB image perception to segment the patient's scan site from the background and locate the desired 2D scan target on the patient's body. And depth images have been used to resolve the 3D normal pose of the probe for RUS [6]–[8]. However, it does not plan the scanning trajectory of the process to reach a fixed position and has only been experimented on dummy models to assess the results. Multiple studies mentioned above do not have servo functions to dynamically adjust control strategies in real scenarios according to the subject's body proportions, position movement, and posture changes [15].

Convolutional neural networks (CNN) are popular for their ability to automatically learn features in images and to obtain feature structures at different scales by changing the network structure to achieve better region localization. It has been widely used in tasks such as target detection. Both the two-stage detection region convolutional neural networks(RCNN) [16]–[18] series and the one-stage detection led by YOLO [19]–[22] have achieved good accuracy in the field of image recognition. In this paper, we intend to introduce the neural network of dynamic human recognition detection into ultrasound scanning and localization to solve the problem of ultrasound region localization failure caused by age and body shape proportion.

The main contributions of this paper can be summarized as follows:

1) In response to the issue of human posture constraints in ultrasound scanning, a vision-based framework for ultrasound region localization is proposed.
2) By proposing an improved AlphaPose network and DeepSORT tracking algorithm, the environmental interference (environmental noise and other dynamic patients) is effectively suppressed.
3) To address the challenges posed by dynamic and complex scenarios in ultrasound scanning, a trajectory generation method based on learning from demonstration is introduced. A visual servoing system for neck ultrasound scanning is developed, enabling interaction with real human subjects.

II. Methodology

In this work, we studied a region localization algorithm for robot ultrasound scanning tasks, as shown in Fig. 1. The main approaches in this section include the study of coarse-grained human skeleton segmentation, tracking framework, fine-grained neck segmentation and spatial localization algorithms.

J. Tian is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, China.

W. Si, N. Wang and C. Yang are Bristol Robotics Lab at the University of the West of England, Bristol, BS16 1QY, UK.

Corresponding author is C. Yang (E-mail: cyang@ieee.org).

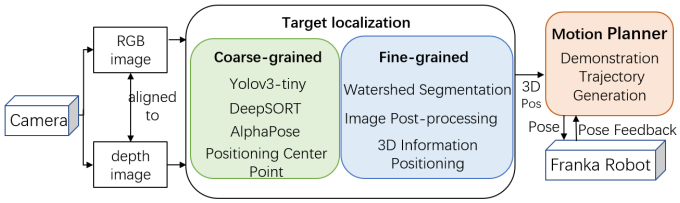


Fig. 1: Block diagram of the neck ultrasound scan control. Referring to the coarse-grained and fine-grained ideas in VJ Det [23] for the design of the target localization algorithm.

In the target detection and region localization, referring to the coarse-grained and fine-grained ideas in Viola Jones Detectors(VJ Det) [23], the whole human body region is extracted using a large framework in the background window, and then referring to the weakly supervised learning method in Deformable Part-based Model(DPM) [24], the initial YOLOv3-tiny and Alpha pose are regarded as human feature point detectors, and the subsequent neck detector and related operations take on the role of the complementary filter, solving the problem of the human neck The problem of low labeling data is solved. By training the YOLOv3-tiny and Alpha pose networks, all configurations of the subsequent neck detectors as part filters can be automatically learned as latent variables.

A. *Coarse-grained human skeleton segmentation and tracking*

1) Targeted human tracking based on YOLOv3-tiny and DeepSORT: It is difficult to directly segment and track the human skeleton in complex environments and lacks robustness. Using a one-step end-to-end deep network to simultaneously draw candidate frames and perform multiple tasks on a complex environment increases the computer's computational power and is less realistic. To make the system more versatile and adaptable to various scenes and individuals, transfer learning is employed by retraining the network using YOLOv3-tiny Net weights pre-trained on ImageNet.

A new YOLOv3-tiny network was trained to detect a single class model of the human body, and the network structure is shown in Fig. 2.

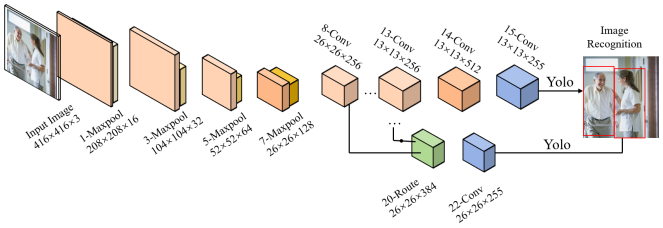


Fig. 2: YOLOv3-tiny network.

The dataset used is a selection of images from the person category of the MS COCO 2017 target detection (1200 images), overlaid with a selection of images of the human body that would appear in the actual ultrasound scene map (manually annotated by the Labelme software). Compared with yolov3 network, the structure of yolov3-tiny net is simpler, the parameters that need to be trained are greatly reduced.

The introduction of a DeepSORT-based patient tracking algorithm improves the tracking performance of the network, preventing misjudgements when multiple patients are present in the scene, which could lead to a sudden increase in robot joint acceleration and safety issues. Also, since the DeepSORT algorithm will match new Detections three times as they appear, the false recall rate of the original YOLOv3-tiny can be improved with the DeepSORT algorithm.

The DeepSORT-based major patient tracking algorithm adds a major id locking module to the traditional DeepSORT algorithm, which will only perform estimation on candidate frames of major patients in the scene. As is shown in Fig. 3, the primary id module is a standard non-maximum suppression NMS, which considers each patient present in the scene and orders all detections in order of their distance from the centre, from closest to farthest. Then all prediction detections are traversed in turn, if the overlap area with the previously selected predictor frame is greater than a certain threshold, it is suppressed (to prevent multiple patients from overlapping and interfering with the main target experiment). All the suppressed frames are compared and the one with the largest intersection ratio with the central image IOU is output as the primary target.

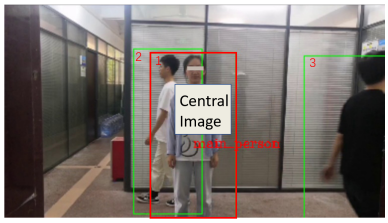


Fig. 3: The primary id module. The detection frame with serial number 1 has the largest IOU with the center image and is the main target.

2) ROI positioning based on improved AlphaPose: With the target detection algorithm, the human body region frame of the main patient was obtained and the next step was to input this region frame into the AlphaPose [25] network for human skeleton extraction. The network structure of the improved AlphaPose is shown in Fig. 4. The improved AlphaPose algorithm adds a regional center localization technique after the PPNMS module to solve specific ultrasound area localization problems.

The network structure of the Spatial Transformer Networks (STN) [26] is shown in Fig. 5. A spatial transformation is applied to all input images and acts on the entire feature input. The significance of the STN proposed is the ability to correct the transformed image by translation, rotation, scaling and other mapping into the desired image output desired by the subsequent network, and then identify and obtain candidate frames.

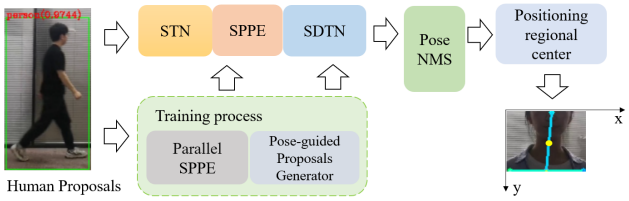


Fig. 4: The improved AlphaPose network. Where STN network refer to Fig. 5, SDTN network refer to Fig. 6 and positioning regional center network refer to Fig. 7.

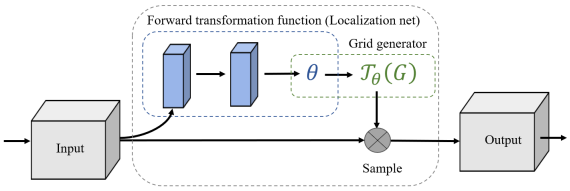


Fig. 5: STN network.

Localization net performs operations such as rotation, translation and scaling of 2D images, as shown in Eq. (1), resulting in a mapping θ of the input (x, y) to the output (x', y'). Grid generator performs a spatial data transformation, transforming a normal image or feature map into a more familiar look for network training and recognition, as shown in Eq. (2). The Sample operation solves the problem of θ being a small number by bilinear interpolation.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \tag{1}$$

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \mathcal{T}_\theta(G_i) = \mathrm{A}_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \tag{2}$$

The network structure of the Inverse Spatial DeTransformer Networks (SDTN) is illustrated in Fig. 6 and is used to map the estimated pose back to the original image coordinates to produce the pose proposal.

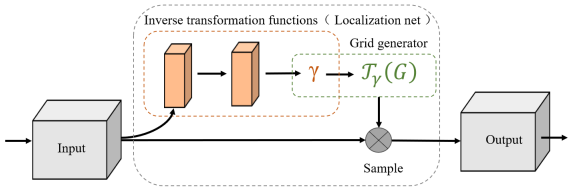


Fig. 6: SDTN network.

The SDTN is the inverse mapping of the STN, and the two are related by the Eq. (3). The parallel single human pose detection (Parallel SPPE, a four-layer hourglass network) in between the two then determines whether the centroid of the human pose is in the centre of the human detection frame, and returns a larger error if it is not, which in turn leads to an optimised STN, prompting the STN to extract a more accurate human detection frame.

$$\gamma = \theta^{-1} \tag{3}$$

Positioning regional center network focuses on the relationship between the main anatomical landmark points of the human body and the ROI of the region of interest, roughly locating the ultrasound scan area through the mapping relationship between the human skeleton segmentation and the main anatomical landmark points of the human body, and then obtaining the average ratio of the distance between the main landmark points and the ROI through the medical human body model scale [27], thereby locating the central position of the diagnostic region of the ultrasound scan. The specific algorithm process is shown in Fig. 7.

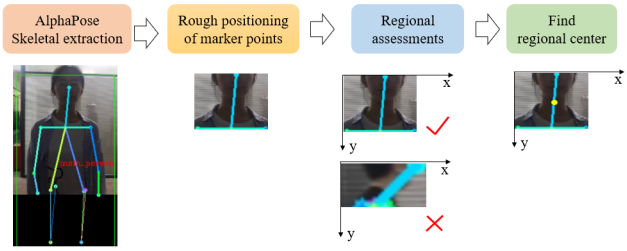


Fig. 7: Regional positioning process. The neck region of a normal patient can be roughly located by the nose marker point and both shoulder marker points, as shown in the second process. Then, the rectangular region containing the neck needs to be evaluated (the third process) to determine whether it conforms to the posture. Finally the center point of the region is located according to the relationship between the region and the center point of the region.

Determine whether it conforms to the posture includes the following requirements:

(1) The human body is located in the middle of the picture, both the shoulder marker point and the nose marker point appear in the picture.

(2) The human body is facing forward, the nose mark point on the horizontal axis is located in the middle of the double shoulder mark point, the nose mark point is lower than the double shoulder mark point.

(3) The neck of the human body is more square (the nose mark on the horizontal axis is located in the middle of the double shoulder mark within the threshold of 50%).

B. Fine-grained Watershed-based neck segmentation

The fine-grained image segmentation and spatial localization algorithm consists of watershed neck segmentation algorithm module, image post-processing module, and depth data information spatial localization module. The subsequent study will also continue the algorithm design with the neck as the experimental region. This paper will use the watershed algorithm to segment the neck area and incorporate the feature

points for watershed post-processing to segment the neck edge.The specific experimental procedures are shown below:

(1) Select the image and perform grayscale conversion, binarization, and median filtering.

(2) Calculate the gradient of the image using the Sobel operator, and apply the region growing algorithm to determine the segmented regions.

(3) Watershed transform is performed on the labeled image to segment the image into different regions.

(4) Filter the labeled edges to leave the region image containing the centroid.

(5) Broad search along the centroid toward the horizontal and vertical axes to find the neck edges.

Also, considering the influence of ambient lighting and other effects, there is noise in the point cloud acquired by RGB-D, which may lead to poor coupling between ROI and ultrasound probe poor coupling between them. Therefore, data processing of depth data is required after obtaining depth information. In this paper, we propose a spatial positioning module for depth data information by improving the data normalization method. The depth of the tiny rectangle near the edge and 5 pixels inward (pix) is taken as the specific depth of the neck edge, as shown in the red tiny rectangle in Fig. 8 of the point cloud image.

Fig. 8: Deep information processing.

In addition, the obtained edge position depth information will have outliers. It is necessary to perform outlier processing before finding the mean value of the location, thus making the localization method more accurate and robust. The method for detecting outliers is a box plot. The interquartile range(IQR) is based on statistical knowledge of the data set as a whole. Outliers are typically identified using box plot criteria, where data points that fall beyond 1.5 times the interquartile range from the upper and lower quartiles are considered as outliers, as shown in Eq. (4).

$$\begin{cases} \text{Outliers.value} & < (\text{Q1} - 1.5 * \text{IQR}) \\ \text{Outliers.value} & > (\text{Q3} + 1.5 * \text{IQR}) \end{cases} \tag{4}$$

III. Experiment

A. Experimental platforms and control methods

In order to verify the proposed method, a neck robotic ultrasound imaging system was built for validation and the experimental platform is shown in Fig. 9. A 7 DoFs Franka Emika Panda robotic arm is equipped with the sonon300l3 ultrasound probe to perform the scanning task. A Realsense camera (D435i) was used to record RGB-D images and provide visual feedback. The whole experiment was carried out in agreement with the requirements of the Academy's Research Ethics Committee. The experimental protocol was approved by the UWE Research Ethics Committee (UWE REC REF No:FET212259).

Fig. 9: Neck robotic ultrasound imaging platform.

Meanwhile, to enhance the interactivity of the robot, dynamic movement primitives (DMP) were introduced in the experiment to learn manipulation skills from demonstrations. The original DMP model was used for offline encoding of manipulation skills and was not suitable for online skill updates. Therefore, a discrete DMP model was selected using a known method [28]. The basic formula for DMP is expressed in Eq. (5).

$$\begin{aligned} \tau_s \dot{v} &= \alpha_z(\beta_z(p_g - p) - v) + F(x) \\ \tau_s \dot{p} &= v \end{aligned} \tag{5}$$

Where p_g indicates the target position, p indicates the current position; τ_s is the proportional parameter, v is the velocity. α_z and β_z need to be designed separately, but these parameters generally satisfy $\alpha_z = 4\beta_z$. $F(x)$ is a nonlinear forcing term that affects the trajectory. And $F(x)$ can be defined as a Gaussian function, as shown in Eq. (6) and Eq. (7).

$$F(x) = \frac{\sum_{i=1}^{N} \psi_i(x) w_i}{\sum_{i=1}^{N} \psi_i(x)} x(p_g - p_0) \tag{6}$$

$$\psi_i = \exp\left(-h_i \left(x - c_i\right)^2\right) \tag{7}$$

Where p_0 denotes the initial position of the expected trajectory. The function $F(x)$ is the result of locally weighted regression using N basis functions ψ_i. The variable x guarantees that the constraint effect of $F(x)$ decreases as x decreases. The parameters c_i and h_i represent the center and width of ψ_i. And i denotes the weight assigned to the corresponding basis function ψi.

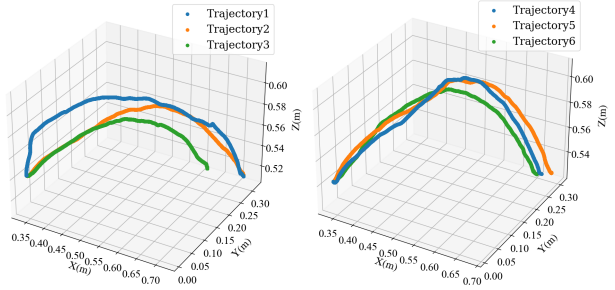
A regularized system can coordinate different degrees of freedom, which can be described as Eq. (8).

$$\tau_s \dot{x} = -\alpha_x x, \quad x \in [0, 1]; \quad x(0) = 1 \tag{8}$$

The scaling parameter τ_s and the positive coefficient α_x are denoted, where $x(0) = 1$ represents the initial value of x, which exponentially converges to zero.

B. Experimental results

During the experiment, the skill of moving the robotic arm from the zero point to the edge of the object using teleoperation human demonstration [29] was transferred to the robot. In order to enrich the skill primitive library, six task demonstrations were performed, and the demonstration trajectory is shown in Fig. 10.



(a) Trajectories 1-3 (b) Trajectories 4-6

Fig. 10: Demonstration trajectories.

In the subsequent process, the decision of which specific model in the primitive library the robot chooses is determined by the decision tree that has been built in advance [30]. The program will calculate the endpoint position of the neck based on the previous algorithms, and enter the DMP model. The experimental results are respectively shown in Fig. 11, Fig. 12 and Fig. 13.

Fig. 11 shows the RGB video stream image of the robot arm approaching the neck and the corresponding B-mode ultrasound image. The yellow edge line in the RGB frame is the identified neck edge. It can be seen that the robotic arm gradually approaches the human neck and finally presses and scans.

Fig. 12 shows the comparison between the experimental trajectory of the robotic arm and the two demonstration trajectories. It can be seen that even when there is a large deviation in the target position on the Y axis, specific end points can still be smoothly approached. The experimental result shows that the proposed algorithm performs well in the ultrasound approach scanning region task and has good generalization ability.

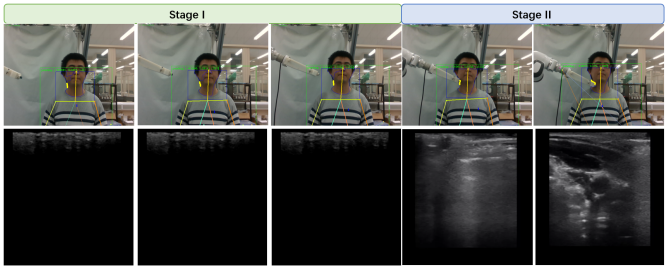


Fig. 11: The approaching process of robotic ultrasound scanning on neck. In stage I, robot approaches the neck. In stage II, robot reaches the target position to press and get the ultrasound image.

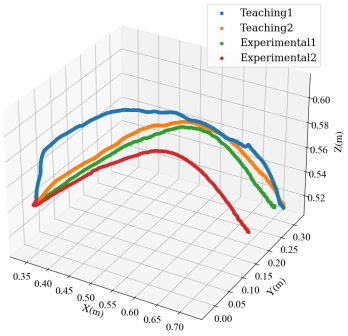
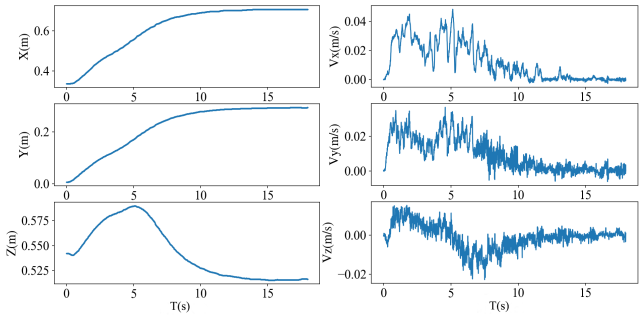


Fig. 12: Comparison of demonstration trajectory and experimental trajectory.

Fig. 13 shows the changes in the position and velocity of the ultrasound probe during the experiment. It can be seen that the position change of the ultrasound probe in various directions is relatively gentle and has flexibility.



(a) Changes in position (b) Changes in velocity

Fig. 13: Position and velocity changes of the ultrasound probe.

Fig. 14: The process and trajectory of human body movement during visual servoing.

During the visual servoing period, if the human body moves (the spatial position of the neck changes dynamically), a comparative experiment is conducted to approach the neck in real time using the traditional non-visual servoing method and the interactive visual servoing method proposed in our paper. The process and trajectory of human body movement during visual servoing are shown in the Fig. 14. The comparison results are shown in Fig. 15. It can be seen that the ultrasound scan without visual servo can only reach the initial neck position. When the body adjusts its posture and position, it cannot track in real time to achieve the arrival of the scan point.

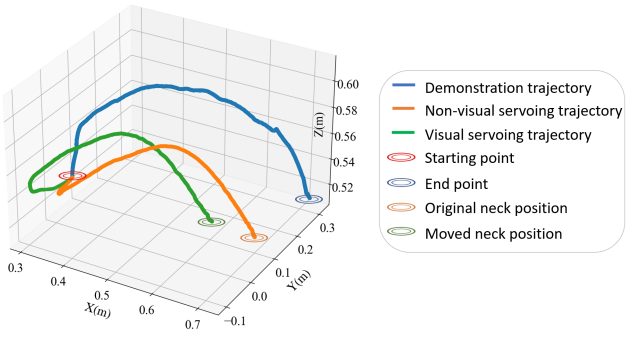


Fig. 15: A comparative experiment to approach the neck in real time using the traditional non-visual servoing method and the interactive visual servoing method proposed in our paper.

IV. Conclusion

The main contribution of this research is the proposal of a vision-based ultrasound scanning localization method and the development of a robotic neck ultrasound system for validation. This system is capable of automatically identifying and locating the human neck, as well as planning the trajectory of the ultrasound probe approaching process without human intervention. In addition, by addressing the posture constraints associated with traditional ultrasound scanning, this study offers new insights into vision-based interactive ultrasound scanning systems for dynamic human subjects.

We have considered the generation of ultrasound images that can be guaranteed in a visual servo dynamic scene. Future research work will investigate region-specific servo control based on ultrasound image characteristics, which hopefully will assist physicians in ultrasound image quality judgment and observation of pathological changes. Additionally, incorporating force sensors into the ultrasound scanning robot system to assist visual servoing is also one of the directions for our future research.

References

[1] M. Muir, P. Hrynkow, R. Chase, D. Boyce, and D. Mclean, "The nature, cause, and extent of occupational musculoskeletal injuries among sonographers: recommendations for treatment and prevention," *Journal of Diagnostic Medical Sonography*, vol. 20, no. 5, pp. 317–325, 2004.

[2] Sofka and M. Carolyn, "Sonography of the painful shoulder," *Ultrasound Quarterly*, vol. 24, no. 4, pp. 243–244, 2008.

[3] G. Harrison and A. Harris, "Work-related musculoskeletal disorders in ultrasound: Can you reduce risk," *Ultra Sound*.

[4] M. Mclean, "The nature, cause, and extent of occupational musculo skeletal injuries among sonographers recommendations for treatment and prevention," *Journal of diagnostic medical sonography: JDMS*, vol. 20, no. 5, 2004.

[5] X. Fong, S. Natarajan, and M. O. Culjat, "Robotic ultrasound systems in medicine," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, no. 3, pp. 507–523, 2013.

[6] X. Ma, Z. Zhang, and H. K. Zhang, "Autonomous scanning target localization for robotic lung ultrasound imaging," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 9467–9474.

[7] R. Tsumura, Y. Koseki, N. Nitta, and K. Yoshinaka, "Towards fully automated robotic platform for remote auscultation," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 19, no. 1, p. e2461, 2023.

[8] M. Tirindelli, M. Victorova, J. Esteban, S. T. Kim, D. Navarro-Alarcon, Y. P. Zheng, and N. Navab, "Force-ultrasound fusion: Bringing spine robotic-us to the next "level"," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5661–5668, 2020.

[9] R. Kojcev, A. Khakzar, B. Fuerst, O. Zettinig, C. Fahkry, R. DeJong, J. Richmon, R. Taylor, E. Sinibaldi, and N. Navab, "On the reproducibility of expert-operated and robotic ultrasound acquisitions," *International journal of computer assisted radiology and surgery*, vol. 12, pp. 1003–1011, 2017.

[10] J. T. Kaminski, K. Rafatzand, and H. K. Zhang, "Feasibility of robot-assisted ultrasound imaging with force feedback for assessment of thyroid diseases," in *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11315. SPIE, 2020, pp. 356–364.

[11] D. Zhang, W. Si, W. Fan, Y. Guan, and C. Yang, "From teleoperation to autonomous robot-assisted microsurgery: A survey," *Machine Intelligence Research*, vol. 19, no. 4, pp. 288–306, 2022.

[12] J. Luo, W. Liu, W. Qi, J. Hu, J. Chen, and C. Yang, "A vision-based virtual fixture with robot learning for teleoperation," *Robotics and Autonomous Systems*, vol. 164, p. 104414, 2023.

[13] K. Kamali, I. A. Bonev, and C. Desrosiers, "Real-time motion planning for robotic teleoperation using dynamic-goal deep reinforcement learning," in *2020 17th Conference on Computer and Robot Vision (CRV)*. IEEE, 2020, pp. 182–189.

[14] M.-S. Pan, J.-T. Tang, and X.-L. Yang, "Movement invariants-based algorithm for medical image tilt correction," *International Journal of Automation and Computing*, vol. 7, no. 4, pp. 543–549, 2010.

[15] S. E. Salcudean, W. H. Zhu, P. Abolmaesumi, S. Bachmann, and P. D. Lawrence, "A robot system for medical ultrasound," 2002.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[17] R. Girshick, "Fast r-cnn. arxiv 2015," *arXiv preprint arXiv:1504.08083*, 2015.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[20] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[21] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[23] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. Ieee, 2001, pp. I–I.

[24] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *2008 IEEE conference on computer vision and pattern recognition*. Ieee, 2008, pp. 1–8.

[25] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2334–2343.

[26] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," *Advances in neural information processing systems*, vol. 28, 2015.

[27] K. S. Saladin, *Human anatomy*. Rex Bookstore, Inc., 2005.

[28] Y. Zhang and C. Yang, "Automatic regrouping of trajectories based on classification and regression tree," *International Journal of Modelling, Identification and Control*, vol. 35, no. 3, pp. 217–225, 2020.

[29] W. Si, N. Wang, and C. Yang, "A review on manipulation skill acquisition through teleoperation-based learning from demonstration," *Cognitive Computation and Systems*, vol. 3, no. 1, pp. 1–16, 2021.

[30] Y. Zhang and C. Yang, "Automatic regrouping of trajectories based on classification and regression tree," *International Journal of Modelling, Identification and Control*, vol. 35, no. 3, pp. 217–225, 2020.