IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

A-SEE2.0: Active-Sensing End-Effector for Robotic Ultrasound Systems with Dense Contact Surface Perception Enabled Probe Orientation Adjustment

Yernar Zhetpissov, Xihan Ma, *Student Member, IEEE,* Kehan Yang, Haichong K. Zhang, *Member, IEEE*

***Abstract*—Conventional freehand ultrasound (US) imaging is highly dependent on the skill of the operator, leading to inconsistent results and increased physical burden on sonographers. Robotic Ultrasound Systems (RUSS) aim to address these limitations by providing standardized and automated imaging solutions, especially in environments with limited access to skilled operators. This paper presents the development of a RUSS system that employs a novel end-effector, A-SEE2.0, which uses dual RGB-D depth cameras to maintain the US probe normal to the skin surface, a default starting configuration for anatomical landmarks identification. Our RUSS integrates RGB-D camera data with robotic control algorithms to maintain orthogonal probe alignment on uneven surfaces without preoperative data. Validation tests using a phantom model show that the system achieves robust normal positioning accuracy. A-SEE2.0 demonstrates 2.47 ± 1.25 degrees normal positioning error on a flat surface and 12.19 ± 5.81 degrees error on a mannequin surface. This work highlights the clinical potential of A-SEE2.0 by demonstrating that, during in-vivo forearm ultrasound examinations, it achieves image quality comparable to manual scanning by a human sonographer.**

***Index Terms* — Computer vision for medical robotics, medical robots and systems, robotics and automation in life sciences, sensor-based control.**

I. INTRODUCTION

ULTRASOUND (US) imaging has established itself as a pivotal tool in medical diagnostics, finding applications in obstetrics [1], cardiology [2], guiding interventional procedures [3], and assisting in radiotherapy treatments [4]. Renowned for their cost-effectiveness, real-time capabilities, and safety [5], US examinations, however, frequently cause pain leading to musculoskeletal disorders among sonographers [29]. Professionals are required to exert significant pressure when positioning the US probe on the patient's body and to fine-tune the probe's position for an optimal image in an ergonomically challenging manner.

US images can vary significantly with slight changes in probe position, orientation, or contact force, making it challenging for sonographers to achieve consistent and repeatable imaging outcomes. Achieving consistent and reliable results demands highly skilled personnel with substantial experience. Unfortunately, such resources are becoming scarce globally [6]. Additionally, the close proximity between sonographers and patients poses infection risks, especially in the context of contagious diseases [7].

To address these challenges, researchers have explored the development of an autonomous Robotic US Systems (RUSS). These innovative systems utilize robot arms to manipulate the US probe, alleviating sonographers of physical burdens. The number of RUSS-related publications has grown exponentially, reaching 125,110 from 2001 to 2022 [8]. Importantly, RUSS allow for remote diagnosis, eliminating the need for direct contact between sonographers and patients [2]. The robot arm can precisely control the probe's pose (position and orientation) and the applied force, ensuring high motion precision and, consequently, securing examination accuracy and repeatability.

Most autonomous RUSS implementations adopt a two-step strategy. First, a scan trajectory is defined based on preoperative data, such as Magnetic Resonance Imaging (MRI) or a vision-based point cloud of the patient's body [9], [10]. In the second step, the robot travels along the trajectory, continuously updating the probe pose and applied force based on intraoperative inputs, such as force/torque sensing and real-time US images.

However, challenges remain in maintaining consistent ultrasound image quality suitable for diagnostic interpretation during autonomous scanning [26]. Involuntary patient movements, errors in scan trajectory registration to patient, and the highly deformable nature of the skin pose significant hurdles. Real-time adjustment of probe positioning and orientation is crucial, especially for maintaining appropriate acoustic coupling between the transducer and the body [5]. Proper probe orientation ensures a clearer visualization of pathological clues in US images, but achieving near real-time reliable adjustment remains a formidable challenge in the field.

A. Related Works

The orthogonal placement of the US probe with respect to the patient body is considered as natural pose for maximum acoustic coupling. Thus, RUSS imaging related works attempt

Manuscript received: March 6, 2025; Revised June 13, 2025; Accepted July 14, 2025.

This paper was recommended for publication by Editor Jessica Burgner-Kahrs upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by Worcester Polytechnic Institute internal funding and the National Institutes of Health under Grant DP5 OD028162, Grant R01 DK133717, and Grant R01 EB030539. *(Yernar Zhetpissov and Xihan Ma are co-first authors) (Corresponding author: Haichong K. Zhang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the institutional research ethics committee at Worcester Polytechnic Institute, under Application No. IRB-21-0613.

Yernar Zhetpissov, Xihan Ma, and Kehan Yang are with the Department of Robotics Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: ytzhetpissov@wpi.edu; xma4@wpi.edu; kyang3@wpi.edu).

Haichong K. Zhang is with the Department of Robotics Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA, and also with the Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: hzhang10@wpi.edu).

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

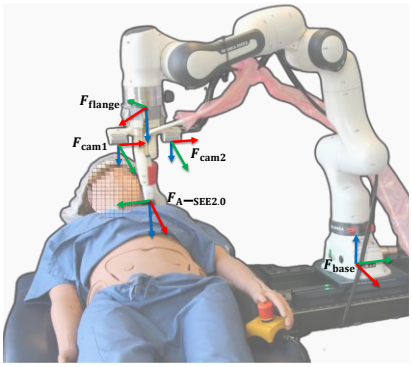

Fig. 1. A-SEE2.0 setup with coordinate frames. F_{base} is robot base frame; F_{flange} is frame end-effector attached to; F_{cam1} and F_{cam2} are RGB-D cameras' frames; $F_{A\text{-}SEE2.0}$ is US probe tip frame.

to place the transducer perpendicular to the skin surface [11], [12], [13], [28]. RGB-depth (RGB-D) cameras have been widely utilized for medical procedures [14], and in RUSS particularly [10], [11], [12], [13]. Despite their popularity, they are mainly used for offline preoperative scan planning and probe placement.

For full orientation control, i.e. simultaneous in- and out-of-plane motion, depth cameras are often used to estimate the normal direction of a constrained surface though the main limitation of relatively low accuracy remains [13]. State-of-the-art RUSS works frequently use external depth perception [10], [26]. For example, Su *et al.* present a fully autonomous RUSS for thyroid scanning using external RGB-D camera for patient localization and initial path planning [25]. Such external placement simplifies the setup and avoids the spatial constraints imposed by end-effector-mounted sensors. However, it may be more susceptible to occlusions and patient movement and require precise calibration depending on the US procedure performed. In contrast, end-effector-mounted RGB-D sensors, as used in our work, offer direct alignment with the scanning region and allow for more compact, self-contained deployment, albeit with more limited coverage and stricter mechanical integration.

Most of the previous works that used probe-mounted based RUSS setups focused on offline preoperative collection of the patient's region of interest to be scanned for path planning. G. Ma *et al.* [15] used eye-in-hand RGB-D camera for RUSS of a forearm and relied on the point cloud surface of a forearm for orientation adjustment during the scan. In addition, Zhang *et al.* [16] proposed a spatial and temporal probe orientation compensation strategy. However, the calculation of the normal vector relied only on the average of five vectors and sampling time is relatively high (0.2s) for real-time applications which may compromise the system accuracy and responsiveness.

To achieve real-time orthogonal positioning simultaneously in-plane and out-of-plane, X. Ma *et al.* [17] proposed the active-sensing end-effector (A-SEE) for RUSS using four in-hand range lasers. Although it is a cost-effective and time-efficient approach for self-normal-positioning for lung ultrasound (LUS), the usage of sparsely arranged, point target laser sensors compromises its perception capability [18]. As a result, scanning narrow anatomical regions, such as the upper and lower extremities and the neck becomes problematic. To summarize, the usage of end-effector-placed RGB-D depth cameras for probe orientation control is not fully explored. Additionally, more sophisticated perception of the probed area during US examination may be essential to advance fully autonomous RUSS in an innovative and impactful manner, where local deformation of the tissue offers an additional source of information for sonography skill learning. Thus, a real-time, full-field-of-view, depth perception-based approach is needed.

B. Contributions

The current work builds upon the further improvement of the A-SEE [17]. The means of sensing mechanism that can perceive the surface structure to be probed is lacking in the previous works. Although A-SEE enables the robot to accurately orient US probe relative to the probed tissue during a relatively simple lung sonography task, due to the large probing area and small required probe inclination angles, it would fail in cases like forearm scanning, where even a single laser sensing outside the human tissue area would result in incorrect orientation. In this work, we aim to lay the foundation for increased robustness and intelligence of the autonomous RUSS systems by addressing the essential problem of reliable alignment of the US probe relative to a patient body in real-time. This is accomplished by introducing a novel robot end-effector where two RGB-D cameras are amounted alongside the US probe to continuously perceive the probed body surface. The required rotation adjustments towards the desired surface-probe orientation are computed based on the fused depth camera point clouds. A human-shared autonomy system is implemented where the probe can automatically land on a patient's body, autonomously adjust its orientation and contact force during imaging, and be slid on the patient's body and rotated about its long axis via teleoperation. The main contributions of this paper are summarized as follows:

- We propose the second generation (the first generation was introduced in [17]) of the compact and cost-effective end tool for RUSS, referred to as the active-sensing end-effector 2.0 (A-SEE2.0). A-SEE was designed to provide real-time information on the rotation adjustment required for achieving normal positioning. It is the first form factor that allows simultaneous in-plane and out-of-plane US probe orientation adjustment. Compared with the previous generation, A-SEE2.0 further implements a dual RGB-D sensing strategy that, for the first time, enables full-field perception of the local tissue surface, allowing for reliable and optimal orthogonal probe placement in real-time. The two RGB-D cameras, rigidly mounted on either side of the ultrasound probe in an eye-in-hand configuration, provide a fused point cloud with 360-degree coverage of the probed area. This configuration ensures local surface perception from a close distance, effectively eliminating line-of-sight issues and capturing the full geometry surrounding the probe tip, even in anatomically constrained or curved regions.
- We integrate A-SEE2.0 with a RUSS system and implement a complete US imaging workflow to demonstrate its self-normal-positioning capability.

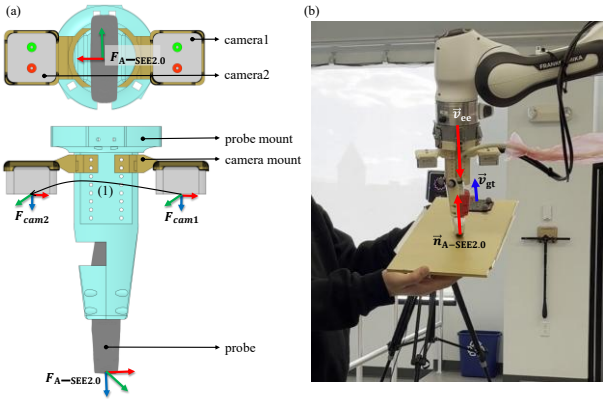

Fig. 2. ASEE2.0 design and probe normal positioning vectors definition. (a) Bottom- and side-views of CAD model. Cameras are on the opposite sides of out-of-plane US imaging plane. (1) is the transformation $\boldsymbol{F}_{\boldsymbol{cam2}}^{\boldsymbol{cam1}}$ between $\boldsymbol{F}_{\boldsymbol{cam1}}$ and $\boldsymbol{F}_{\boldsymbol{cam2}}$; (b) Flat plane following experiment and ASEE2.0 vector and frame definitions.

II. Materials and Methods

This section describes the system implementation details. The proposed A-SEE2.0-integrated RUSS (shown in Fig. 1) performs intraoperative probe self-normal-positioning with contact force control. The shared control scheme facilitates teleoperated sliding of the probe along the patient's body surface and enables rotation of the probe about its axis during imaging.

A. Technical Approach Overview

The implementation of the system was accomplished through the following tasks: **Task 1)** A-SEE2.0 fabrication, RGB-D cameras calibration and integration with RUSS; **Task 2)** Orientational positioning via the 2-DoF rotational motion of the probe; **Task 3)** US probe contact force control via the 1-DoF translational motion; **Task 4)** Teleoperation control using a joystick for the remaining 3-DoF.

A-SEE2.0's perception component consists of two short-range RGB-D cameras (D405, RealSense, Intel, USA) attached to side-mounted brackets (see Fig. 2). These cameras are chosen due to their sub-millimeter accuracy at close-range and compactness with an ideal depth sensing range between 7 cm and 50 cm. Point cloud processing and normal estimation are performed on workstation "PC-1" (see Fig. 3).

The wired US probe (C1-6 Curved Array Probe, General Electric, USA) is encapsulated in a 3D-printed mount. A clamping mechanism connects A-SEE2.0 to the 7-DoF robotic manipulator (Panda Research 3, Franka Emika, Germany). The robot motion control and the teleoperation pipeline are implemented on workstation "PC-2". System communication is established using Robot Operating System (ROS).

The robot's self-normal-positioning is activated before landing on the patient body. In-plane and out-of-plane rotation of the US probe, ensuring real-time orthogonal placement, is based on the fused point cloud. Constant force adaptation allows automatic translational motion in probe z-axis. The remaining 3 DoF (rotation about z-axis, translation along x- and y-axis) are controlled by the operator via a 3-DoF flight joystick

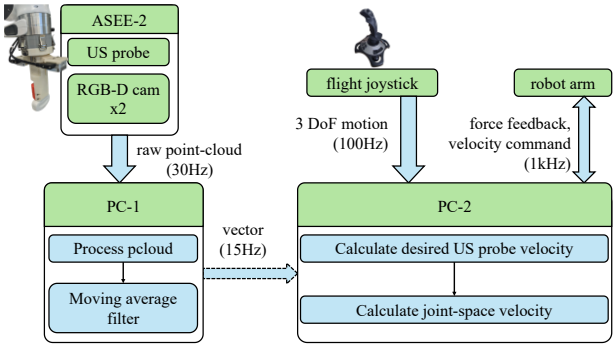

Fig. 3. RUSS equipped with ASEE2.0 imaging workflow diagram. Green boxes represent hardware, blue boxes - calculations, arrows - data and command communication.

(PXN-2113-SE, PXN, Japan).

B. System Calibration

Knowing the target body position in 3D space relative to the world coordinates is crucial for accurate US probe placement tasks. For our setup, both intrinsic and extrinsic camera calibrations were performed to ensure the accurate acquisition of point cloud data relative to the cameras and the robot end-effector. For intrinsic calibration, each camera was calibrated using its on-chip self-calibration feature. For camera-to-camera calibration, defined as a static homogeneous transformation $F_{\text{cam2}}^{\text{cam1}}$, the relative location of the two RGB-D cameras measured from the CAD model of the 3D-printed camera mount was used.

The Park and Martin eye-in-hand calibration method [19] was used to obtain a homogeneous transformation between the camera and the end-effector [20]. The forward kinematics of the robot then allows the transformation of a point cloud to the robot base frame. Thus, the normal vector of the probed surface relative to the robot is known, allowing for further self-normal positioning.

C. Probe Normal-Positioning

The probe's angular positioning in A-SEE2.0 relies on the fused point cloud generated from the RGB-D cameras and robot kinematics. This section describes the point cloud processing pipeline, normal vector estimation, and the probe normal positioning by actuating the robot arm to align with the estimated normal vector.

Point cloud processing is required for precise patient skin sensing and real-time normal vector estimation. Fig. 2(b) illustrates schematically the normal vector estimation. The raw point cloud from each camera is cropped along the z-axis of F_{cam1} and F_{cam2}, preserving points in [0.02, 0.25] cm range. The proximity to the sensed surface inherent in US procedures leads to an excessively high density of points near the sensed surface. To reduce redundancy, the data is down-sampled to 35000 points per camera.

The fused point cloud is obtained by concatenating the point clouds from cam1 and cam2 using the rigid body transformation $F_{\text{cam2}}^{\text{cam1}}$, calculated in Section II-B.

The Point Cloud Library (PCL) is used for further denoising of point clouds and normal estimation. The points

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

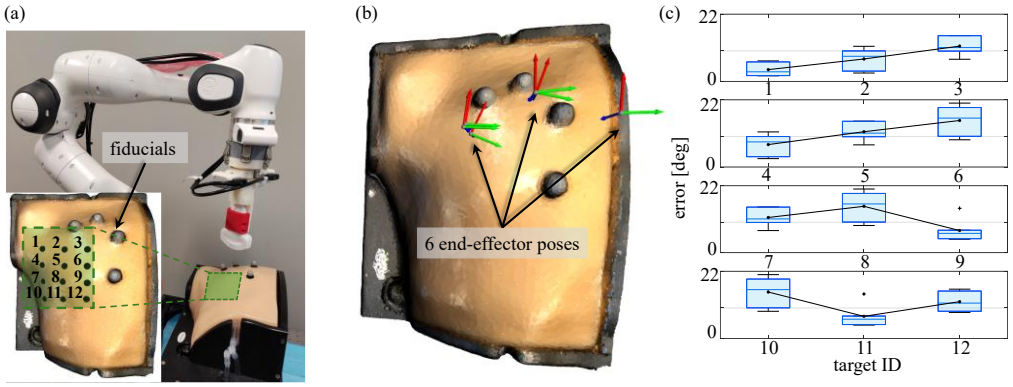

Fig. 4. Normal estimation accuracy experiment A2. (a) Setup image depicting 12 targets and fiducials configuration; (b) Six $F_{A-SEE2.0}$ frame location snapshots during data collection process. (c) The normal estimation errors calculated at 12 targets on lung phantom.

corresponding to the US probe were box-cropped using the known probe and mount dimensions, as well as their relative position to the cameras. Then, voxel grid downsampling, followed by statistical outlier removal were performed.

For point cloud data streamed at every timestamp, the normal vector is computed as the average of local normals over a 10x10 cm rectangular region centered around the probe tip. Due to the ambiguity of the estimated vector direction, the z-axis component of $\vec{n}_{\mathrm{A-SEE2.0}}$ is kept positive as follows:

$$\vec{n}_{\mathrm{A-SEE2.0}} = [x, y, z] = \begin{cases} [-x', -y', -z'] & z' < 0 \\ [x', y', z'] & z' \geq 0 \end{cases}, \quad (1)$$

where $\vec{n}_{\mathrm{A-SEE2.0}}$ is the vector normal to the probed surface, $[x', y', z']$ is the original output vector of normal estimator.

A moving average filter is applied to the estimated normal vector to ensure smooth motion. The size of the moving window was empirically set to 7 for better robustness and minimal latency.

With accurate normal estimation in real-time, A-SEE2.0 can be integrated with the robot to enable motion that tilts the US probe, aligning it with the normal direction of the skin surface. As depicted in Fig. 2(b), upon normal positioning of the probe, the angular difference between the vectors $\vec{n}_{\mathrm{A-SEE2.0}}$ and $F_{\mathrm{A-SEE2.0}}$ frame's z-direction vector $\bar{v}_{\mathrm{ee}}$ is minimized. Thus, we address **Task 2** by simultaneously applying in-plane rotation ω_y, and out-of-plane rotation ω_x. The angular velocities about the two axes at timestamp t are given by a PD control law

$$\begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} = \begin{bmatrix} K_p & K_d & 0 & 0 \\ 0 & 0 & K_p & K_d \end{bmatrix} \begin{bmatrix} proj_{xz}\bar{v}_{\mathrm{ee}} \\ \frac{\Delta proj_{xz}\bar{v}_{\mathrm{ee}}}{\Delta t} \\ proj_{yz}\bar{v}_{\mathrm{ee}} \\ \frac{\Delta proj_{yz}\bar{v}_{\mathrm{ee}}}{\Delta t} \end{bmatrix}, \quad (2)$$

where K_p and K_d are empirically tuned control gains; $proj_{xz}\bar{v}_{\mathrm{ee}}$ and $proj_{yz}\bar{v}_{\mathrm{ee}}$ are projections of $\bar{v}_{\mathrm{ee}}$ on xz and yz planes, respectively;
$\Delta proj_{xz}\bar{v}_{\mathrm{ee}} = proj_{xz}\bar{v}_{\mathrm{ee}}(t) - proj_{xz}\bar{v}_{\mathrm{ee}}(t-1)$;
$\Delta proj_{yz}\bar{v}_{\mathrm{ee}} = proj_{yz}\bar{v}_{\mathrm{ee}}(t) - proj_{yz}\bar{v}_{\mathrm{ee}}(t-1)$; Δt is control time interval. A PD controller was chosen for its simplicity, real-time capability, and proven performance in the previous generation. Thus, its use allowed a direct comparison with A-SEE to evaluate the impact of dense depth sensing introduction in A-SEE2.0. The angular velocity control response can reach 30 Hz.

D. US Probe Contact Force Control

To prevent loose contact between the probe and the skin, which causes acoustic shadows in the image, a force control strategy is implemented to maintain consistent pressure throughout the imaging process. This strategy also ensures that the probe maintains gentle contact with the body, prioritizing patient safety. We adapted a force control method from A-SEE that regulates the linear velocity along the z-axis within the A-SEE2.0 framework. Velocity adaptation follows a two-stage approach, handling landing and scanning motions separately. During landing, the probe's velocity asymptotically decreases as it approaches the body surface. During scanning, the velocity is adjusted based on deviations between the measured desired force values. The velocity along the z-axis at time t is calculated as

$$v_{fz}(t) = w \cdot v + (1-w) \cdot v_{fz}(t-1), \quad (3)$$

where w is a constant in the range *(0,1)* to maintain the smooth velocity profile, and v is computed as

$$v = \begin{cases} K_{p1}(\tilde{d} - \min(\overrightarrow{d_z})) & \min(\overrightarrow{d_z}) \geq \tilde{d} \\ K_{p2}(\tilde{F} - \bar{F}_z) & \min(\overrightarrow{d_z}) < \tilde{d} \end{cases}, \quad (4)$$

where $\overrightarrow{d_z}$ is the vector of the z-components of the processed point cloud; $\bar{F}_z$ is the measured tip force along the z-axis of $F_{\mathrm{A-SEE2.0}}$, estimated from joint torque readings provided by the robot software; $\tilde{F}$ is the desired reference contact force; K_{p1}, K_{p2} are empirically determined gains; $\tilde{d}$ is the threshold distance distinguishing the landing stage from the scanning stage, set to the distance from camera lens to the tip of the probe (150 mm).

E. RUSS equipped with A-SEE2.0 Imaging Workflow

The A-SEE2.0 system, providing 7-DoF control of the US probe, integrates self-normal positioning, contact force regulation, and teleoperation capabilities. Fig. 3 illustrates the overall A-SEE2.0 system block diagram for RUSS.

During the landing stage, the probe is gradually lowered until it reaches the reference contact force. Once contact is established, the operator gains control over the probe's lateral movements in the x-y plane and its rotation along the long axis. Using joystick teleoperation, the operator can slide the probe along the patient's body surface, generating end-effector

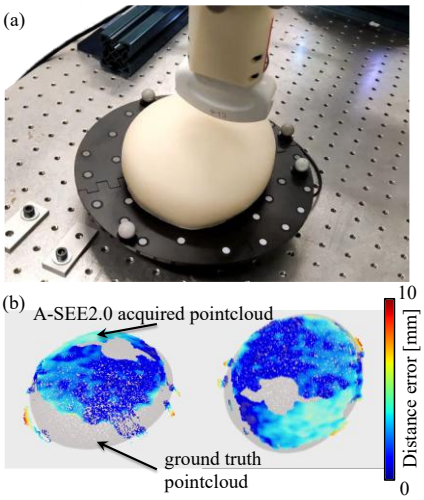

Fig. 5. Surface reconstruction experiment A3. (a) Setup image showing the breast phantom, fiducials and end-effector probe; (b) Breast phantom surface perception distance error heatmap of the fused point cloud. Grey colored points represent scanned ground-truth.

velocities in the $F_{\mathrm{A-SEE2.0}}$ body frame. These velocities are then transformed into joint velocities commanded to the robot.

Throughout the procedure, US images are continuously streamed to PC2. After the operator completes the scanning, the robot automatically returns to its home configuration.

III. Experiment Setup

A. Validation of Self-Normal-Positioning

Three experiments were designed to evaluate the self-normal positioning and surface reconstruction capabilities of A-SEE2.0. Experiment A1 (expt. A1) evaluates the system's ability to track the normal vector on a flat surface (ideal case) and its response time. Experiment A2 (expt. A2) assesses normal vector estimation on complex uneven surfaces, simulating anthropomorphic structures. Experiment A3 (expt. A3) evaluates system calibration and point cloud fusion accuracy by measuring phantom surface reconstruction performance.

In expt. A1, four reflective optical trackers were placed on the robot end-effector, coplanar with *xy* plane of $F_{\mathrm{A-SEE2.0}}$, to determine $\bar{v}_{\mathrm{ee}}$. Another set of four trackers were placed on the flat surface to define $\bar{v}_{\mathrm{gt}}$ (See Fig. 2(b)). Markers were tracked using a motion capture system (Vantage, Vicon Motion Systems Ltd, U.K.). The normal positioning error is defined as the angular difference between $\bar{v}_{\mathrm{ee}}$ and $\bar{v}_{\mathrm{gt}}$. In the experiment, rotations about the x-and y-axis of $F_{\mathrm{A-SEE2.0}}$ were enabled. The plane board was held against the US probe and rotated in four phases. Force control was disabled in this experiment to isolate the evaluation of self-normal-positioning capability. During each phase, the operator altered $\bar{v}_{\mathrm{gt}}$ by tilting the plane to a certain angle. Meanwhile, the robot was aligning $\bar{v}_{\mathrm{ee}}$ and $\bar{v}_{\mathrm{gt}}$ using the self-normal-positioning feature. Once the robot reached a steady state, the next phase was introduced. The four phases were repeated four times (16 tilts total), during which synchronized $\bar{v}_{\mathrm{ee}}$ and $\bar{v}_{\mathrm{gt}}$ values were recorded. The normal positioning errors and response times were calculated.

In expt. A2, phantom upper torso (COVID-19 Live Lung Ultrasound Simulator, CAE Healthcare™, USA) mimicking patient body texture was utilized as the experiment subject for evaluation of local surface normals estimation. The purpose of the experiment was to examine A-SEE2.0's ability to estimate accurate surface normals. Target regions were chosen to match anterior, anterior-lateral, and lateral regions of the chest examined during a lung US procedure as per BLUE lung US protocol [21]. This allows testing of normal angle estimation across a variety of tilt angles. The phantom was scanned using the 3D Scanner app to obtain the ground truth surface point cloud. To register the ground truth phantom point cloud with A-SEE2.0 acquired phantom point cloud, four fiducials were placed on the phantom surface. The target regions were manually selected on the left side of the torso, as shown in Fig. 4(a). Targets 1, 4, 7, 10 are in the anterior region; targets 3, 6, 9, 12 are in the lateral region; targets 2, 5, 8, 11 are in the middle. Each target region is represented by a circular area with a radius of 3 cm. The ground truth normal $\bar{v}_{\mathrm{gt}}$ for each target are estimated by averaging local normal vectors inside each circular area. The robot was manually moved to six positions (see Fig. 4(b)) with RGB-D cameras facing the phantom to collect A-SEE2.0 point clouds, ensuring visibility of the whole region-of-interest with targets and the fiducials. Source normal vectors were calculated as described in Section II-C for each region-of-interest. The point clouds were registered using fiducial correspondences between the ground truth and A-SEE2.0 acquired point clouds. The resultant registration was used to transform targets into $F_{\mathrm{A-SEE2.0}}$ frame. The angle errors of the normal estimation were calculated accordingly for each target. The mean angle error across 6 robot positions was recorded.

In expt. A3, a breast elastography phantom (Part Number 1552-01, CIRS) was used to evaluate point clouds fusion by estimating surface reconstruction accuracy. Chamfer Distance (CD) metric was used as it is a commonly used surface similarity measure in computer vision research [31]. It represents the average distance of paired nearest neighbors between the two point clouds. CD effectively captures the overall geometrical similarity of two point clouds and is insensitive to slight local misalignment, making it suitable for this evaluation. The robot end-effector was placed above the phantom and was capturing the whole phantom surface (see Fig. 5a) from a perpendicular viewing angle. Similar to expt. A2, breast phantom ground truth surface representation was acquired using 3D Scanner app and registered with A-SEE2.0 fused point cloud using four fiducials.

B. Validation of Autonomous Probe Adjustment in Narrow Anatomical Regions

To demonstrate the superiority of the A-SEE2.0 system over the previous generation A-SEE, a human subject study was conducted to evaluate its ability to acquire high-quality images during forearm US scanning, a scenario involving narrow anatomical region. Three volunteers were recruited for the right forearm US procedure for both robotic and manual experiments. Written informed consent was given to the volunteers prior to all test sessions. The body mass index (BMI)

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

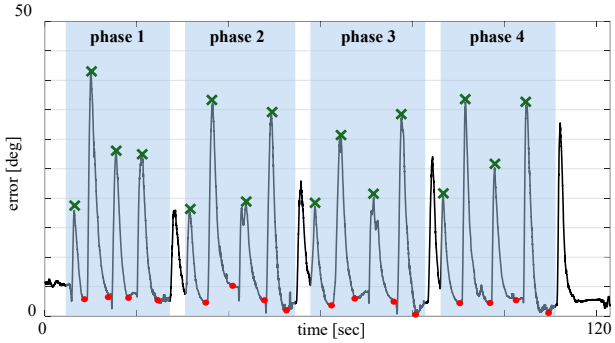

Fig 6. Recorded normal positioning errors on the flat surface over time for self-normal-positioning validation experiment expt. A1. The red dots indicate the residual errors (i.e., the minimum normal positioning error achievable) after normal positioning, the green crosses indicate maximum recorded errors for each side tilt.

of the human subjects ranged from 22.4 to 29.2, with a mean BMI of 25.75. The overall setup is shown in Fig. 7(a).

When collecting robotically acquired data, the operator teleoperated the probe to scan over the target's forearm, while A-SEE2.0 dynamically optimized its orientation. US images were continuously acquired during the whole procedure. Next, the operator performed manual scanning following same trajectory along the forearm. Ten representative images where the subjects' vessel walls were visible were selected from both robotic and manual procedures for image quality evaluation. Since forearm vessels are commonly examined in forearm US [22], we used their visibility, quantified by the contrast-to-noise ratio (CNR), as a measure of image quality. For a given a rectangular region of interest (ROI) on the landmark, CNR is defined as:

$$CNR = \frac{|\mu_{\text{roi}} - \mu_{\text{bg}}|}{\sqrt{\sigma_{\text{roi}}^2 + \sigma_{\text{bg}}^2}}, \quad (5)$$

where μ_{roi} and σ_{roi} are the mean and the standard deviation of the pixel intensities in the ROI, μ_{bg} and σ_{bg} are the mean and the standard deviation of the pixel intensities in the image background. The CNR of the vessel wall was calculated for both A-SEE2.0 and manually acquired US images. The ROI area was fixed at 10 x 10 pixels for CNR computation.

In the human study, we also validated the effectiveness of the force control strategy. The force exerted at the probe tip, recorded by the robot's sensors, was sampled at 30 Hz throughout the experiment. The target contact force was empirically set to a constant value of 3.5 N. This value was selected to balance anatomical visibility and safety, consistent with prior studies that involved imaging in the forearm region [27]. To quantify the control performance, we used the force control error, defined as the difference between the measured and desired forces as the evaluation metric.

IV. Results

A. Validation of Self-Normal-Positioning

Fig. 6 shows the results of flat surface tracking in expt. A1 using A-SEE2.0. The error peaked when the operator rotated the plane in four distinct directions, then decreased rapidly as the robot aligned the probe with the normal. The mean normal positioning error across sixteen manually positioned planes, calculated by averaging the residual errors (see red dots in Fig.6), was 2.47 ± 1.25 degrees (40.77% less than A-SEE). The response time measures the robot's responsiveness to the sudden changes in the imaging surface orientation. It is calculated as the time interval between the previous peak and the corresponding valley. The average response time was 3.34 ± 0.56 seconds, outperforming the response time of A-SEE with its 3.67 seconds response time by 8.99% [17]. Additionally, the authors successfully applied A-SEE to human lung ultrasound by demonstrating equivalent CNR in comparison with manual ultrasound, even in the presence of respiratory motion [30]. Therefore, the improved A-SEE2.0 response time is considered sufficient to handle minor patient movements in many clinical applications, such as lung US examinations.

Fig. 4(c) shows the normal estimation errors when sensing the phantom in expt. A2. Errors when scanning the anterior targets (1, 4, 7, 10), anterior-lateral targets (2, 5, 8, 11), and lateral targets (3, 6, 9, 12) were 13.46 ± 10.72 degrees, 8.43 ± 1.43 degrees, and 12.10 ± 0.80 degrees, respectively. The mean error for all 12 targets was 12.19 ± 5.81 degrees, indicating generally acceptable normal estimation performance across the upper torso regions. The higher normal estimation error observed in expt. A2 is attributed to the increased geometric complexity of the lung phantom surface. The approximate normal estimation for the mannequin is improved by 16.9% when compared to the normal positioning accuracy of 14.67 degrees on A-SEE. A-SEE2.0 exhibited higher accuracy for anterior-lateral targets, as most positions were aligned perpendicularly to this region. Self-normal positioning and force control were disabled in this experiment to isolate the evaluation of normal direction estimation capability.

Fig. 5(b) presents the surface reconstruction error via a heatmap. The error used for the heatmap generation is the absolute distance from each point in the fused point cloud to the closest point in the ground-truth point cloud. The average CD of ten fused point clouds collected from the same probe position is 7.5 ± 0.42 mm which is considered sufficiently low when compared with the literature [31]. As expected, the error was the highest at the boundary with larger surface curvatures from the camera's perspective. Overall, the point cloud fusion produced a reliable surface reconstruction outcome, validating the system's calibration accuracy.

B. Validation of Autonomous Probe Adjustment in Narrow Anatomical Regions

To evaluate whether the A-SEE2.0 system can perform effective autonomous ultrasound scanning, particularly on complex and narrow anatomical regions, we conducted human the forearm imaging experiment. The goal of this validation is to show that A-SEE2.0 can achieve diagnostically valid image quality (compared to manual scanning) and stable contact force control.

The assessment of US image quality and contact force control was carried out during human subject forearm US

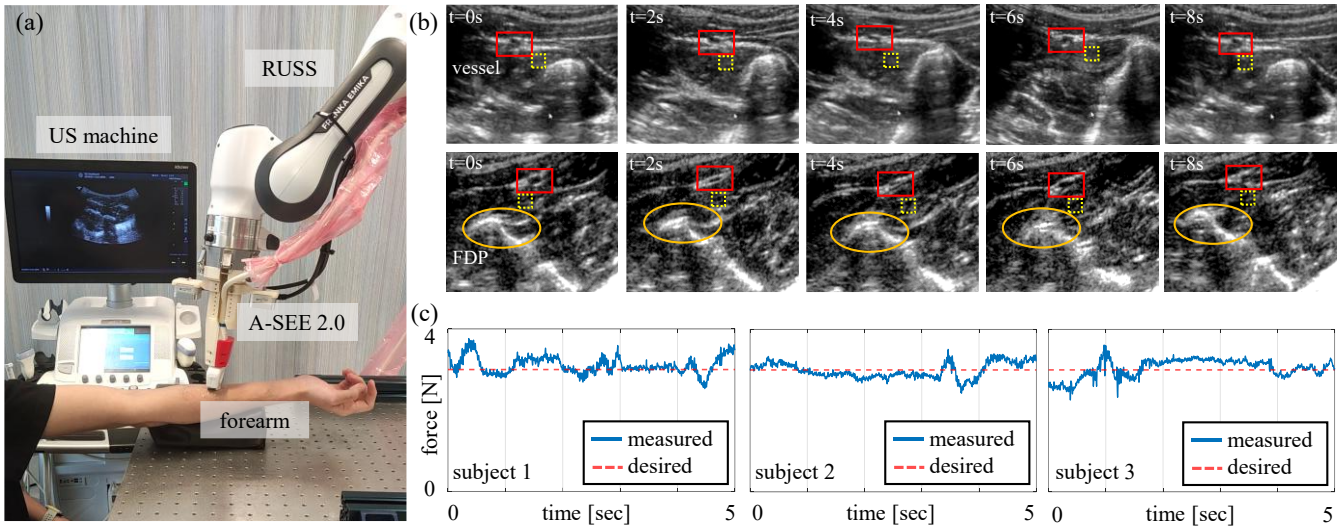

Fig. 7. Human subjects experiment. (a) Human subject experimental setup; (b) First row - manual, second row – A-SEE2.0 acquired example scans of subject1. Red boxes show vessel walls as ROI and yellow dashed boxes show background regions chosen for CNR metric of image quality evaluation step. Yellow ellipses show flexor digitorum profundus (FDP) muscle boundary; (c) Force profiles during forearm US experiment for all 3 subjects.

experiment performed using A-SEE2.0. Fig. 7(b) illustrates exemplary US images captured during the experiments with ROI being depicted. Vessel walls have been chosen as the ROI, where CNR was calculated. Overall, vessel walls were clearly visible and differentiable from the surrounding tissue. Table I shows the image quality results corresponding to each subject. The mean CNR across all subjects was 3.78 ± 1.78 (manual) vs. 3.91 ± 1.08 (A-SEE2.0), indicating a similar image quality of A-SEE2.0 and manually acquired images. Moreover, the noticeable drop in standard deviation for all subjects indicates that with A-SEE2.0-equipped robot system, the image quality was more consistent.

A pain level (scale of 10) associated with the probe pressure during A-SEE2.0 RUSS experiment has been queried 3 times for each subject throughout imaging. 2 out of 3 subjects reported 0 level of pain, 1 subject reported a minor discomfort of 1.67, showing the system's overall safety.

Fig. 7(c) presents the probe's z-direction force profiles recorded during A-SEE2.0 imaging for each subject. The force errors remained mostly within ± 0.5 N, with an average force error of 0.041 ± 0.215 N. The results demonstrate that the contact force controller effectively maintained the applied force near the target value during real forearm US imaging using A-SEE2.0.

These results show the system's effectiveness in maintaining appropriate probe orientation and pressure while acquiring high-quality images in constrained anatomical regions like the human forearm. Notably, the previous generation A-SEE lacked sufficient sensing coverage, thus was not capable of scanning anatomically constrained regions like the human upper and lower extremities.

TABLE I
COMPARISONS OF MANUAL AND ROBOTIC FOREARM ULTRASOUND SCANS

Manual		**A-SEE2.0**		
Subject	CNR	Subject	CNR	Average Pain level
1	3.16 ± 1.57	1	3.29 ± 0.69	1.67
2	3.29 ± 1.01	2	4.02 ± 0.80	0
3	4.88 ± 1.96	3	4.41 ± 1.39	0
mean	**3.78 ± 1.78**	mean	**3.91 ± 1.08**	**0.56**

V. Discussion and Conclusions

A novel end-effector A-SEE2.0 for RUSS with real-time probe normal-positioning capability has been presented. This work extends the previous version of the active-sensing end-effector by incorporating RGB-D cameras for a more comprehensive perception of the scanned surface. This feature brings advantages to RUSS by enabling dense sensing of the contact area. The use of in-hand RGB-D cameras can enhance the understanding of probe-tissue interactions during US scanning, potentially enabling a more intelligent and generalizable framework for autonomous RUSS. Although only orthogonal positioning has been evaluated in this paper, omnidirectional orientation control of the US probe is theoretically possible using A-SEE2.0. This flexibility is valuable for various US imaging applications such as liver, heart, and gallbladder [23] that requires significant probe tilting. One limitation of the previous A-SEE was the constrained sensing field-of-view. The addition of depth data in this work expands the US scanning capabilities beyond simple front-chest LUS to more advanced applications, such as forearm US. Preliminary experiment results show that A-SEE2.0 has improved in normal positioning accuracy, with a 40.77% normal estimation error reduction on flat surface (2.47 ± 1.25 degrees for a flat surface) and a 16.9% reduction on uneven surfaces (12.19 ± 5.81 degrees for a mannequin). Regarding system responsiveness, A-SEE has already been successfully applied to human lung ultrasound, demonstrating image quality equivalent to that of manual ultrasound, even in the presence of respiratory motion. A-SEE2.0 demonstrates an 8.99% faster self-normal-positioning response time than A-SEE, confirming its suitability for clinical applications

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

involving minor patient movements. Additionally, A-SEE2.0 is unlikely to suffer from optical occlusion that can encounter for eye-to-hand based RUSS [10], [12]. Lastly, high CNR results with improved consistency (manual: 3.78 ± 1.78 vs. A-SEE2.0: 3.91 ± 1.08) across subjects demonstrate the generalizability of the A-SEE2.0 pipeline for populations with diverse body habitus (i.e., different BMI) [24].

This study is limited to evaluating only the normal-positioning capability and image quality during forearm US experiments on human subjects. Although only normal-positioning experiments have been conducted, the system has strong potential for real-time optimal US probe placement and, potentially, learning-based RUSS approaches. We acknowledge that certain clinical scenarios, such as obstetric ultrasound, may involve unpredictable motion (e.g., fetal movements), where a response time of 3.34 seconds may not be sufficient to maintain the region of interest. Therefore, A-SEE2.0 is currently most suitable for applications with relatively stable targets or minor patient movements, such as limbs, lungs, and liver imaging. As future work, we will adapt and validate A-SEE2.0 for a broader range of clinical scenarios bringing RUSS closer to clinical implementation.

REFERENCES

[1] R. Wright, N. Toussaint, A. Gomez, V. Zimmer, B. Khanal, J. Matthew, et al., "Complete fetal head compounding from multi-view 3D ultrasound," *in Proc. 22nd Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 11767, pp. 384–392, 2019, doi: 10.1007/978-3-030-32248-9_43

[2] K. Li, Y. Xu, and M. Q.-H. Meng, "An Overview of Systems and Techniques for Autonomous Robotic Ultrasound Acquisitions," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 2, pp. 510–524, May 2021, doi: 10.1109/TMRB.2021.3072190

[3] G. Fichtinger, J. P. Fiene, C. W. Kennedy, G. Kronreif, I. Iordachita, D. Y. Song, et al., "Robotic assistance for ultrasound-guided prostate brachytherapy," *Med. Image Anal.*, vol. 12, no. 5, pp. 535–545, Oct. 2008, doi: 10.1016/j.media.2008.06.002

[4] Y. K. Kong, M. U. Cho, C. W. Park, S. Y. Kim, M. J. Kim, J. Moon, et al., "Quantification of physical stress experienced by obstetrics and gynecology sonographers: A comparative study of two ultrasound devices," *Appl. Ergon.*, vol. 100, Art. no. 103665, Apr. 2022, doi: 10.1016/j.apergo.2021.103665

[5] Z. Izadifar, P. Babyn, and D. Chapman, "Mechanical and biological effects of ultrasound: A review of present knowledge," *Ultrasound Med. Biol.*, vol. 43, no. 6, pp. 1085–1104, Jun. 2017

[6] D. Won, J. Walker, R. Horowitz, S. Bharadwaj, E. Carlton, and H. Gabriel, "Sound the Alarm: The Sonographer Shortage Is Echoing Across Healthcare," *J. Ultrasound Med.*, vol. 43, no. 7, pp. 1289–1301, Jul. 2024

[7] T. Wataganara, P. Ruangvutilert, P. Sunsaneevithayakul, A. Sutantawibul, M. Chuchotirot, B. Phattanachindakun, et al., "Minimizing cross transmission of SARS-CoV-2 in obstetric ultrasound during COVID-19 pandemic," *J. Perinat. Med.*, vol. 48, no. 9, pp. 931–942, Nov. 26 2020

[8] Z. Jiang, S. E. Salcudean, and N. Navab, "Robotic ultrasound imaging: State-of-the-art and future perspectives," *Med. Image Anal.*, vol. 89, p. 102878, Oct. 2023, doi: 10.1016/j.media.2023.102878

[9] C. Hennersperger, B. Fuerst, S. Virga, O. Zettinig, B. Frisch, T. Neff, et al., "Towards MRI-Based Autonomous Robotic US Acquisitions: A First Feasibility Study," *IEEE Trans. Med. Imaging*, vol. 36, no. 2, pp. 538–548, Feb. 2017, doi: 10.1109/TMI.2016.2620723

[10] F. Suligoj, C. M. Heunis, J. Sikorski, and S. Misra, "RobUSt–An autonomous robotic ultrasound system for medical imaging," *IEEE Access*, vol. 9, pp. 67456–67465, 2021, doi: 10.1109/ACCESS.2021.3077037

[11] Q. Huang, J. Lan, and X. Li, "Robotic arm based automatic ultrasound scanning for three-dimensional imaging," *IEEE Trans. Industr. Inform.*, vol. 15, no. 2, pp. 1173–1182, Feb. 2019, doi: 10.1109/TII.2018.2871864

[12] C. Yang, M. Jiang, M. Chen, M. Fu, J. Li and Q. Huang, "Automatic 3-D Imaging and Measurement of Human Spines With a Robotic Ultrasound System," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-13, 2021, Art no. 7502713, doi: 10.1109/TIM.2021.3085110

[13] Z. Jiang, M. Grimm, M. Zhou, J. Esteban, W. Simson, G. Zahnd, et al., "Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1342–1349, Apr. 2020, doi: 10.1109/LRA.2020.2967682

[14] L. Hargaš and D. Koniar, Usage of RGB-D Multi-Sensor Imaging System for Medical Applications. IntechOpen, 2023, doi: 10.5772/intechopen.106567

[15] G. Ma, S. R. Oca, Y. Zhu, P. J. Codd, and D. M. Buckland, "A Novel Robotic System for Ultrasound-guided Peripheral Vascular Localization," *in Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 12321-12327, doi: 10.1109/ICRA48506.2021.9561924

[16] B. Zhang, H. Cong, Y. Shen, and M. Sun, "Visual Perception and Convolutional Neural Network-Based Robotic Autonomous Lung Ultrasound Scanning Localization System," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 70, no. 9, pp. 961–974, Sep. 2023, doi: 10.1109/TUFFC.2023.3263514

[17] X. Ma, W. Y. Kuo, K. Yang, A. Rahaman, and H. K. Zhang, "A-see: Active-sensing end-effector enabled probe self-normal-positioning for robotic ultrasound imaging applications," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 12475–12482, Oct. 2022, doi: 10.1109/LRA.2022.3218183

[18] X. Ma, Z. Zhang, and H. K. Zhang, "Autonomous Scanning Target Localization for Robotic Lung Ultrasound Imaging," *in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2021, pp. 9467-9474, doi:10.1109/IROS51168.2021.9635902

[19] F. C. Park and B. J. Martin, "Robot sensor calibration: Solving AX=XB on the Euclidean group," *IEEE Trans. Robot. Autom.*, vol. 10, no. 5, pp. 717–721, Oct. 1994, doi: 10.1109/70.326576

[20] H. Wu, "Camera Calibration Toolbox". Accessed May 2024. https://github.com/hongtaowu67/calibration_toolbox

[21] D. A. Lichtenstein, "BLUE-protocol and FALLS-protocol: Two applications of lung ultrasound in the critically ill," *Chest*, vol. 147, no. 6, pp. 1659–1670, Jun. 2015, doi: 10.1378/chest.14-1313

[22] Z. X. Yan, Y. J. Zhou, Y. X. Zhao, Z. M. Zhou, S. W. Yang, and Z. J. Wang, "Anatomical study of forearm arteries with ultrasound for percutaneous coronary procedures," *Circ. J.*, vol. 74, no. 4, pp. 686–692, Apr. 2010, doi: 10.1253/circj.CJ-09-0577

[23] A. Smereczyński, K. Kołaczyk, and E. Bernatowicz, "Optimization of diagnostic ultrasonography of the gallbladder based on own experience and literature," *J. Ultrason.*, vol. 20, no. 80, pp. e29–e35, 2020, doi: 10.15557/JoU.2020.0006

[24] C. B. Weir and A. Jan, "BMI Classification Percentile and Cut off Points," StatPearls Publishing, 2023. [Online] Available: https://www.ncbi.nlm.nih.gov/books/NBK541070/

[25] K. Su, J. Liu, X. Ren, Y. Huo, G. Du, W. Zhao, et al., "A fully autonomous robotic ultrasound system for thyroid scanning," *Nat. Commun.*, vol. 15, no. 1, p. 4004, May 11 2024, doi: 10.1038/s41467-024-48421-y

[26] X.-X. Lin, M. D. Li, S. M. Ruan, W. P. Ke, H. R. Zhang, H. Huang, et al., "Autonomous robotic ultrasound scanning system: A key to enhancing image analysis reproducibility and observer consistency in ultrasound imaging," *Front. Robot. AI*, vol. 12, Feb. 2025, doi: 10.3389/frobt.2025.1527686

[27] D. Huang, C. Yang, M. Zhou, A. Karlas, N. Navab and Z. Jiang, "Robot-Assisted Deep Venous Thrombosis Ultrasound Examination Using Virtual Fixture," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 381-392, 2025, doi: 10.1109/TASE.2024.3351076

[28] P. Chatelain, A. Krupa and N. Navab, "Optimization of ultrasound image quality via visual servoing," *in Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2015, pp. 5997-6002, doi: 10.1109/ICRA.2015.7140040

[29] G. C. Bolton and D. L. Cox, "Survey of UK sonographers on the prevention of work related muscular-skeletal disorder (WRMSD)," *J. Clin. Ultrasound*, vol. 43, no. 3, pp. 145–152, Mar. 2015, doi: 10.1002/jcu.22216

[30] X. Ma, M. Zeng, J. C. Hill, B. Hoffmann, Z. Zhang and H. K. Zhang, "Guiding the Last Centimeter: Novel Anatomy-Aware Probe Servoing for Standardized Imaging Plane Navigation in Robotic Lung Ultrasound," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 6569-6580, 2025, doi: 10.1109/TASE.2024.3448241

[31] L. Rustler, V. Volprecht and M. Hoffmann, "Empirical Comparison of Four Stereoscopic Depth Sensing Cameras for Robotics Applications," *IEEE Access*, vol. 13, pp. 67564-67577, 2025