ACT-US: End-to-End Visuomotor Policy Learning for Respiratory-Adaptive Robotic Ultrasound

Xingyu Zhang*, Ke Lu*†, Hongjuan Pei*, Qingyuan Liu*, Yihua Shao‡, Xiang Luo*, and Jian Xue*

*School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China

†Peng Cheng Laboratory, Shenzhen 518055, China

‡Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

{zhangxingyu231, liuqingyuan23, luoxiang221}@mails.ucas.ac.cn;

{luk, peihongjuan, xuejian}@ucas.ac.cn; yihuajerry@gmail.com

***Abstract*—Autonomous robotic ultrasound (RUS) aims to standardize diagnostic image acquisition, yet control-based methods often assume a pre-positioned probe and leave the visual-guided approach unresolved. Pure end-to-end learning can plan global trajectories but may not guarantee safety and comfort under physiological motion. We propose ACT-US, a hierarchical framework combining generative imitation learning with respiratory-adaptive control for full-cycle autonomous scanning. ACT-US uses a Conditional Variational Autoencoder (CVAE) with Action Chunking to model multi-modal expert behaviors and generate nominal trajectories across free-space navigation and soft-tissue manipulation. A weakly-supervised quality estimator learns a "comfort-aware" force policy, while a respiratory-adaptive admittance controller uses feedforward visual tracking to compensate for thoracic motion. Experiments on human subjects show that ACT-US achieves expert-level dexterity in unstructured environments and outperforms baselines without physiological adaptation.**

***Index Terms*—Deep Imitation Learning, Medical Robotics, Transformer, Conditional VAE, Multi-modal Fusion.**

I. Introduction

Although robotic ultrasound (RUS) systems have achieved significant progress in force control and path planning, the vast majority operate under a restrictive "Initialization Assumption," presuming the probe is already pre-positioned on the patient's skin. A clinically viable autonomous ultrasound agent must demonstrate end-to-end capabilities: starting from a home position, autonomously localizing the patient, planning a safe trajectory to transition from free space to physical contact, and finally executing a high-quality diagnostic scan.

Achieving this full-cycle autonomy presents substantial challenges, primarily due to the heterogeneous nature of the task phases. The process demands a seamless transition between two distinct dynamic regimes: the Approach Phase, which is vision-dominant and requires avoiding perceptual blind spots as the camera nears the body; and the Contact Phase, which is force-dominant and requires precise management of soft-tissue deformation. Traditional rule-based state machines struggle to handle these dual discontinuities in perception (switching from visual to haptic feedback) and dynamics (switching from free motion to contact-constrained manipulation), often resulting in rigid behaviors that lack the adaptability of human sonographers.

This work was supported by the National Natural Science Foundation of China (62027827), the National Key R&D Program of China (2023YFF0716504), and the R&D Program of Beijing Municipal Education Commission (2019022). (*Corresponding author: Jian Xue*)

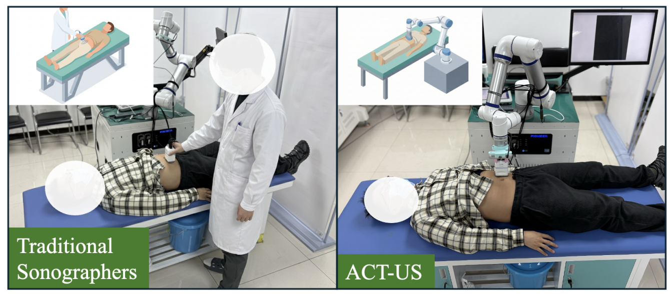


Fig. 1. Comparison between traditional manual sonography and the proposed ACT-US framework. The top panel illustrates a sonographer performing manual scanning, while the bottom panel demonstrates the ACT-US system utilizing a robotic manipulator for autonomous ultrasound acquisition.

To address these hybrid dynamics within a unified framework, we propose **ACT-US** (Action Chunking with Transformers for Ultrasound), adopting a hierarchical control strategy instead of directly outputting motor torques. At the high level, we leverage Imitation Learning (IL) augmented by Action Chunking and a Conditional Variational Autoencoder (CVAE). This allows the system to model the multi-modal distribution of expert behaviors, effectively predicting long-horizon nominal trajectories that bridge the perceptual gap between visual navigation and tissue manipulation, while mitigating the error accumulation typical in sequential decision-making.

Furthermore, to ensure safety and image quality under physiological uncertainties (e.g., respiration), we integrate a Respiratory-Adaptive Admittance Controller at the low level. Unlike standard IL approaches that implicitly learn force profiles, our system explicitly decouples trajectory generation from contact force regulation. We introduce a weakly-supervised quality estimation module, trained on data with deliberately varied contact forces, to learn a "comfort-aware" policy. This module dynamically modulates the target admittance force based on real-time image quality, while a feedforward compensation term synchronizes the end-effector with the patient's thoracic motion.

Our specific contributions are as follows:

- **Full-Cycle Task Formulation:** We redefine the autonomous ultrasound task to emphasize the continuous closed-loop process, explicitly addressing the transition from the visual-guided approach phase to the force-guided scanning phase.
- **Hierarchical ACT-US Architecture:** We propose a novel framework that combines deep imitation learning (ACT) for global trajectory planning with a robust admittance controller for local interaction. This hybrid approach utilizes a unified model to handle multi-modal observations (Vision/Ultrasound/Proprioception) while ensuring operational safety.
- **Physiologically-Adaptive Control:** We introduce a comfort-aware control strategy driven by weak supervision. By modeling the correlation between contact force and image quality, and predicting respiratory motion via depth sensing, our system dynamically optimizes contact pressure to balance diagnostic image clarity with patient comfort.

II. Related Work

This section reviews the state-of-the-art across three key dimensions of robotic ultrasound (RUS): visual navigation, learning-based manipulation [28], and compliant physical interaction.

A. From Teleoperation to Autonomous Navigation

Early research primarily focused on teleoperation to overcome spatial limitations. Pioneers like De Cunha et al. [5] and Salcudean et al. [18] established the feasibility of master-slave control, while recent works have enhanced transparency through haptic feedback [15].

Visual servoing algorithms typically rely on image intensity [13], moments [14], or speckle tracking [11] to maintain scan planes. These methods have been successfully applied to specific anatomical tasks, such as automated breast screening [21], [24] and vascular reconstruction [12]. More recently, multi-modal fusion (e.g., RGB-D and thermal) has been proposed to improve coarse localization [17].

However, a prevalent limitation in these works is the "Initialization Assumption": most systems initiate control only after the probe is in contact. The critical *Approach Phase*, transitioning from free-space to contact, remains under-explored, often requiring manual positioning or simplified assumptions that do not hold in unstructured environments.

B. Imitation Learning and Cognitive Intelligence

To enable higher-level decision-making, researchers are increasingly adopting data-driven approaches. Imitation Learning (IL) and Reinforcement Learning (RL) have enabled robots to perform complex scans. For instance, Huang et al. [8] achieved autonomous carotid scanning via IL, while Su et al. [19] utilized RL for thyroid nodule detection. Beyond motion planning, recent efforts focus on "cognitive" capabilities, such as understanding the "Language of Sonography" [10], semantic segmentation for musculoskeletal structures [20], and handling dynamic targets like the beating heart [22].

Despite these advances, two challenges persist: (1) **Error Accumulation:** Long-horizon tasks in standard IL often suffer from compounding drift; (2) **Safety:** End-to-end models that directly output torques pose safety risks. Our work addresses this by introducing Action Chunking with Transformers (ACT) to mitigate drift, while decoupling the high-level policy from low-level force safety.

C. Force Control and Physiological Adaptation

Stable acoustic coupling is essential for image quality. Classic impedance and admittance controllers [7] and hybrid motion-force strategies [4] are widely used to regulate contact pressure. Some systems have incorporated optimization to maximize vessel visibility [23] or compensate for patient movement [9].

However, physiological motion (e.g., respiration) remains a disturbance that most systems passively reject rather than actively compensate for. Rigid force thresholds can lead to loss of contact or patient discomfort during deep breathing. We advance this by proposing a *Respiratory-Adaptive Admittance Controller* that uses feedforward visual compensation and weakly-supervised image quality feedback to dynamically optimize contact force, ensuring a "comfort-aware" examination.

III. Methodology

We formulate the autonomous ultrasound scanning task as a hierarchical control problem, structured into four interconnected modules: Data Collection, High-Level Policy, Mid-Level Perception, and Low-Level Control.

A. Data Collection

To capture sonographer manipulation skills, we constructed a multi-modal teleoperation dataset. An expert controlled the robot with a haptic device to scan healthy volunteers.

We recorded N demonstration trajectories $D = \{\tau_1, ..., \tau_N\}$, each containing synchronized observations and actions at 30Hz. The observation space includes:

- **Visual Data:** RGB-D streams from a wrist-mounted camera (local view) and a fixed external camera (global view) $\{I^{(i)}_{\text{rgb}}, I^{(i)}_{\text{depth}}\}_{i=1}^{2}$.
- **Ultrasound Data:** Ultrasound images acquired by the transducer mounted on the robotic end-effector I_{us}.
- **Proprioception:** Joint positions $\boldsymbol{q} \in \mathbb{R}^6$ and end-effector pose $\mathbf{x}_{\text{ee}} \in SE(3)$.
- **Force Feedback:** 6-DoF force/torque readings $\mathbf{F}_{\text{ext}} \in \mathbb{R}^6$.

To support weak supervision, we deliberately varied contact force in selected sub-sessions to capture the pressure-image-clarity correlation.

B. High-Level Policy: ACT-US

The high-level policy is the global trajectory generator. We adopt Action Chunking with Transformers (ACT) to map multi-modal observations to future action sequences.

1) Observation and Action Space:

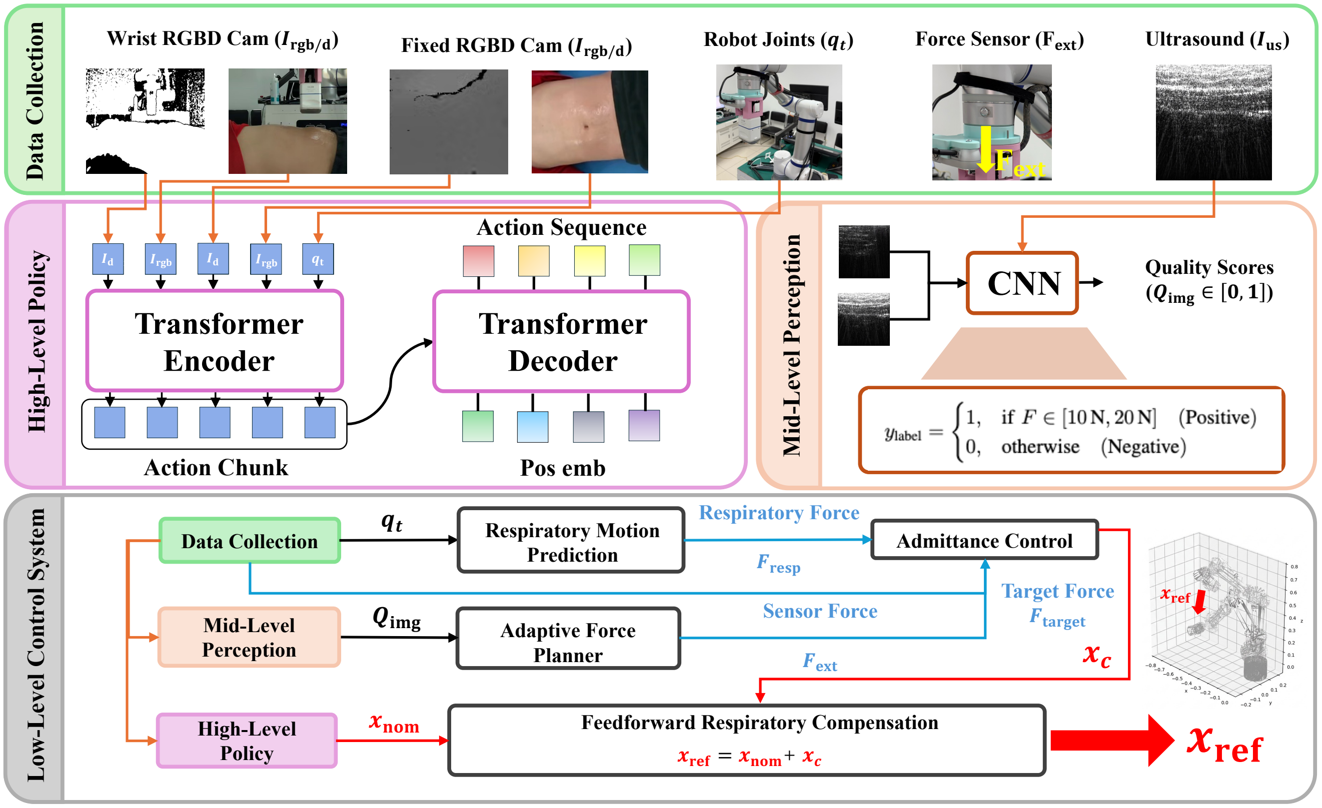


Fig. 2. Overview of ACT-US. Multimodal inputs (RGB-D, force, and ultrasound) are processed by: (1) a high-level Transformer policy generating the nominal trajectory $\mathbf{x}_{nom}$; (2) mid-level perception estimating image quality (Q_{img}) and respiratory motion ($\hat{\mathbf{x}}_{resp}$); and (3) low-level adaptive force planning and respiratory compensation for compliant scanning.

a) Observation Space (O_t): We use multi-modal inputs:

$$O_t = \{I_{\text{us}}, \{I_{\text{rgb}}^{(i)}, I_{\text{depth}}^{(i)}\}_{i=1}^{2}, \boldsymbol{q}_t, \mathbf{F}_{\text{ext}}\}, \tag{1}$$

where I_{us} is the grayscale ultrasound image, $I_{\text{rgb/depth}}^{(i)}$ are RGB-D inputs from wrist and fixed cameras, $\boldsymbol{q}_t \in \mathbb{R}^6$ is the joint configuration, and $\mathbf{F}_{\text{ext}} \in \mathbb{R}^6$ is the force sensor reading.

b) Action Space (A_t): Instead of direct motor torques, the policy predicts the nominal reference trajectory for the downstream admittance controller:

$$A_t = \{\mathbf{p}_t^{\text{ref}}, \boldsymbol{\phi}_t^{\text{ref}}\}, \tag{2}$$

where $\mathbf{p}_t^{\text{ref}} \in \mathbb{R}^3$ and $\boldsymbol{\phi}_t^{\text{ref}} \in \mathbb{R}^3$ are the target end-effector position and orientation. Contact force is *not* dictated by ACT, but is modulated by the comfort-aware controller (Sec. III-C).

2) Action Chunking & CVAE: ACT predicts a future action sequence (a "chunk") of length k. To model expert stochasticity, we train a Conditional Variational Autoencoder (CVAE):

1) **Encoder:** Compresses the ground truth action sequence $\mathbf{a}_{t:t+k}$ and observation O_t into a latent style variable $z \sim q_\phi(z|\mathbf{a}, O_t)$.
2) **Decoder (Policy):** Generates the action sequence conditioned on z and O_t: $\mathbf{a}_{t:t+k} \sim \pi_\theta(\mathbf{a}|O_t, z)$.

The loss function minimizes the ELBO: $\mathcal{L} = \mathcal{L}_{recon} + \beta\mathcal{L}_{KL}$. During inference, Temporal Ensembling is applied to smooth the trajectory.

C. Mid-Level Perception: Weakly-Supervised Quality Estimation

To achieve closed-loop quality optimization without manual labeling, we use weak supervision.

We leverage the correlation between contact force and acoustic coupling by varying contact force from $0\,\text{N}$ to $25\,\text{N}$ and automatically labeling frames:

$$y_{\text{label}} = \begin{cases} 1, & \text{if } F \in [10\,\text{N}, 20\,\text{N}] \quad \text{(Positive)} \\ 0, & \text{otherwise} \quad \text{(Negative)} \end{cases} \tag{3}$$

A lightweight CNN predicts a quality score $Q_{\text{img}} \in [0, 1]$ with Binary Cross-Entropy loss, enabling recognition of "optimal imaging features" even at lower forces.

D. Low-Level Control: Respiratory-Adaptive Admittance

The nominal ACT trajectory is executed through variable impedance control with respiratory compensation.

1) Respiratory Motion Prediction: The measured force F_{ext} contains physiological artifacts, so we isolate the respiratory component F_{resp} with a band-pass filter (0.1–$0.5\,\text{Hz}$). The respiration-induced surface motion is approximated as a quasi-periodic signal and tracked online with a Fourier Series:

$$F_{\text{resp}}(t) \approx A \sin(\omega t + \phi) + F_{\text{bias}}, \tag{4}$$

where A is the amplitude, ω is the angular frequency (measured as $0.25\,\text{Hz} \approx 1.57\,\text{rad/s}$), and ϕ is the phase. These

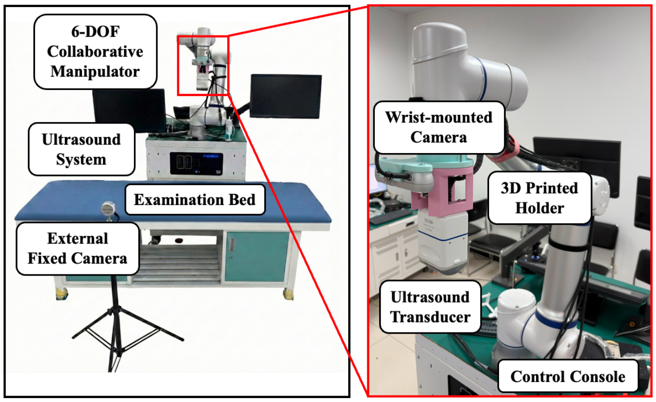


Fig. 3. Overview of the ACT-US platform. Left: 6-DOF manipulator, fixed camera, and examination environment. Right: end-effector with 3D-printed holder, wrist camera, and ultrasound transducer.

parameters estimate $\hat{F}_{\text{resp}}(t + \Delta t)$ and generate feedforward velocity v_{ff} to synchronize the end-effector with thoracic uplift.

2) Image-Quality-Driven Adaptive Force Planning: An outer-loop force planner adjusts the target interaction force F_{target} from real-time ultrasound image quality. Instead of a rigid threshold, it maps the normalized quality metric $Q_{\text{img}} \in [0, 1]$ to the desired contact force:

$$F_{\text{target}}(t) = F_{\min} + \eta \cdot (F_{\max} - F_{\min}) \cdot (1 - Q_{\text{img}}(t)), \quad (5)$$

where $F_{\min} = 10\,\text{N}$ is the baseline force for acoustic coupling, $F_{\max} = 25\,\text{N}$ is the safety limit, and $\eta \in (0, 1]$ is the sensitivity gain. Thus, high image quality drives the controller toward $F_{\min}$, while low quality increases force toward $F_{\max}$ within the safety bound.

3) Respiratory-Compensated Admittance Control: The inner loop tracks F_{target} while compensating for respiratory disturbances. The relation between force error and corrective motion is:

$$\mathbf{M}_d \ddot{\mathbf{x}}_c + \mathbf{D}_d \dot{\mathbf{x}}_c + \mathbf{K}_d \mathbf{x}_c = \mathbf{F}_{\text{ext}} - \mathbf{F}_{\text{target}} + \mathbf{F}_{\text{resp}}, \quad (6)$$

where $\mathbf{x}_c \in \mathbb{R}^6$ is the compliance deviation, $\mathbf{M}_d$, $\mathbf{D}_d$, and $\mathbf{K}_d$ are desired inertia, damping, and stiffness matrices, and $\mathbf{F}_{\text{ext}}$ is the measured external wrench.

a) Feedforward Respiratory Compensation: To mitigate feedback phase lag, we integrate a feedforward term from the estimated respiratory motion. The final reference trajectory is:

$$\mathbf{x}_{\text{ref}}(t) = \mathbf{x}_{\text{nom}}(t) + \mathbf{x}_c(t). \quad (7)$$

Here, $\mathbf{x}_{\text{nom}}$ is the ACT-generated nominal scanning path, and feedforward compensation synchronizes the end-effector with chest motion.

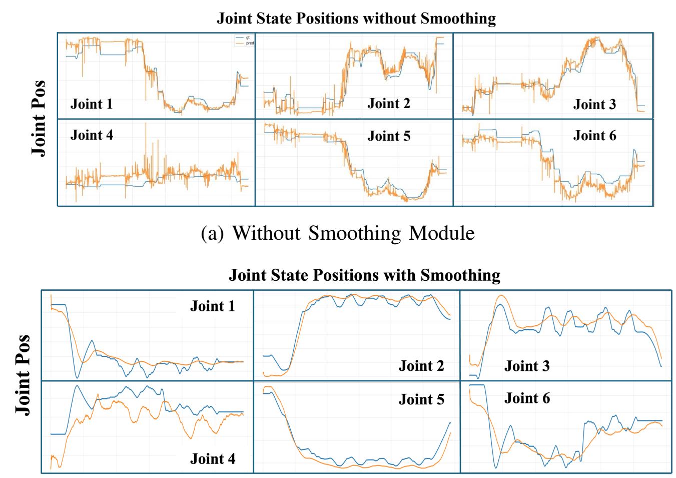
b) Discrete-Time Implementation: For real-time execution at $30\,\text{Hz}$, Eq. (6) is discretized with the Symplectic Euler method, recursively updating $\dot{\mathbf{x}}_c$ and $\mathbf{x}_c$ to compute the corrective offset.

IV. Experiments

A. Experimental Setup

1) Hardware Architecture: As illustrated in Fig. 3, the hardware architecture integrates the robotic manipulator, external fixed camera, examination environment, and the end-effector assembly with the wrist-mounted camera and ultrasound transducer.

2) Teleoperation and Control Interface: Expert demonstrations were collected through a low-latency HID interface running at $500\,\text{Hz}$. It supports *Cartesian Impedance Control* for compliant $SE(3)$ scanning and *Joint Space Control* $(q_1, \ldots, q_6)$ for singularity avoidance, while logging synchronized poses, RGB-D images, and force readings.



(a) Without Smoothing Module

(b) With Smoothing Module (Ours)

Fig. 4. **Impact of Smoothing on Joint Stability.** Without smoothing, joint states exhibit high-frequency jitter; with smoothing, the trajectory remains continuous.

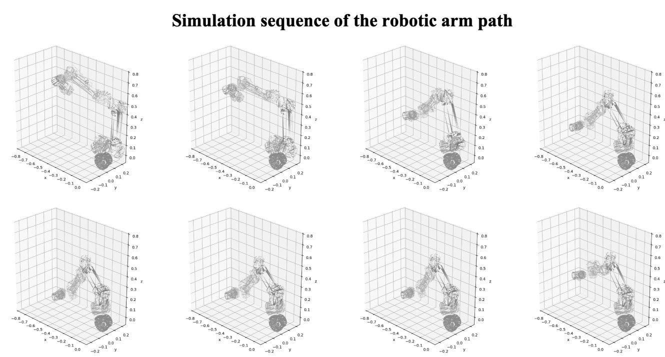


Fig. 5. Illustration of the ACT-based trajectory prediction and ultrasound scanning process.

3) Hardware Design: The core actuation unit is an Elite EC66 collaborative robot with $\pm 0.02\,\text{mm}$ repeatability. A topology-optimized **high-rigidity PLA** end-effector integrates the ultrasound transducer, reduces inertia, and fixes the transformation between the Tool Center Point (TCP) and sensor frame for geometric consistency.

TABLE I
ABLATION STUDY OF CONTROL STRATEGIES

Method	Success Rate (↑)
ACT-US (Ours)	**19/20**
Visual Navigation Only	8/20
ACT-US w/o Smoothing	1/20

TABLE II
QUANTITATIVE EVALUATION OF FORCE STABILITY

Metric	Uncompensated (Original Disturbance)	Compensated (Residual Error)	Improvement (%)
Mean Absolute Error (N)	12.45	2.09	**83.2%** ↓
Root Mean Square (N)	13.48	2.94	**78.2%** ↓
Standard Deviation (σ)	5.15	2.92	**43.3%** ↓
Peak-to-Peak Amplitude (N)	31.96	20.35*	36.3% ↓

* Represents the smoothed prediction range.

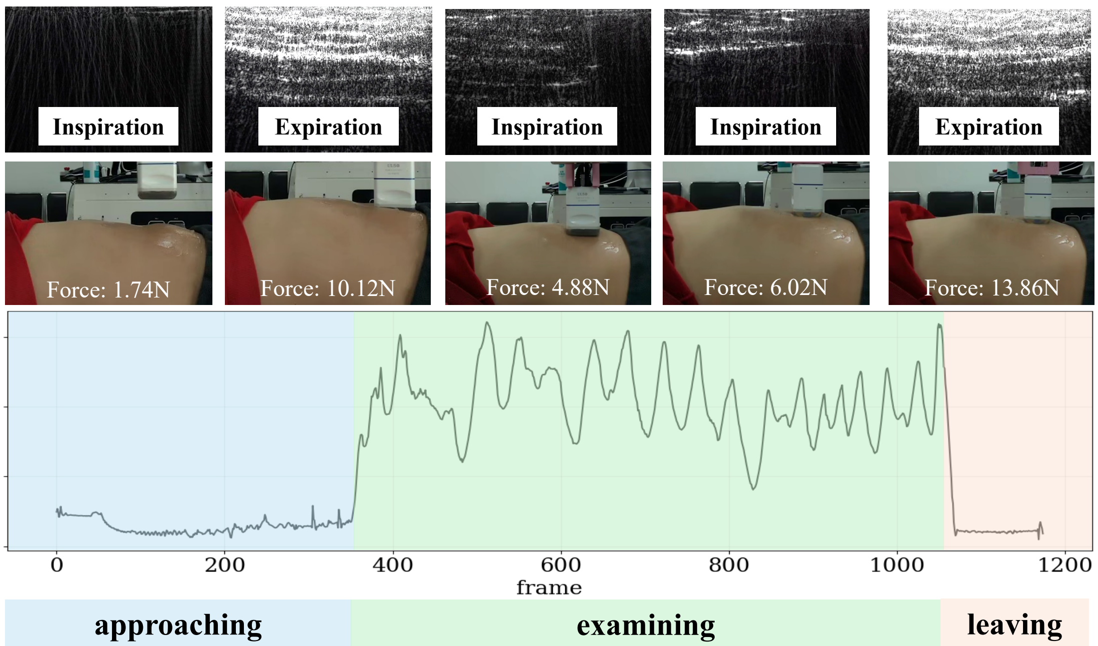


Fig. 6. Scanning under respiratory motion. Ultrasound images, probe-tissue interaction, and force profiles show safe interaction across the physiological cycle.

4) Multi-modal Perception and Computation: We use an Intel RealSense L515 LiDAR for global localization and a D455 stereo camera for local depth on texture-less skin. A wrist-mounted 6-axis Force/Torque sensor captures contact dynamics at 1000 Hz, and all modalities are ROS-aligned and synchronized at 30 Hz on a Threadripper 3970X/RTX 3090 workstation.

B. Ablation Studies

We conducted ablations on trajectory smoothing, respiratory compensation, and multi-modal fusion. Removing smoothing caused severe velocity jitter and emergency stops, reducing success to **1/20** (Fig. 4, Table I). Respiratory compensation reduced force Mean Absolute Error (MAE) by **83.2%** and RMS error by **78.2%** (Table II). The "Visual Navigation Only" baseline reached only **8/20** success because it prioritized force minimization over acoustic coupling.

C. Simulation Validation

Before physical deployment, we validated the trajectory policy in high-fidelity simulation to assess path feasibility and smoothness. As shown in Fig. 5, the manipulator completed scans across diverse abdominal models, generating singularity-free trajectories that respect anatomical constraints.

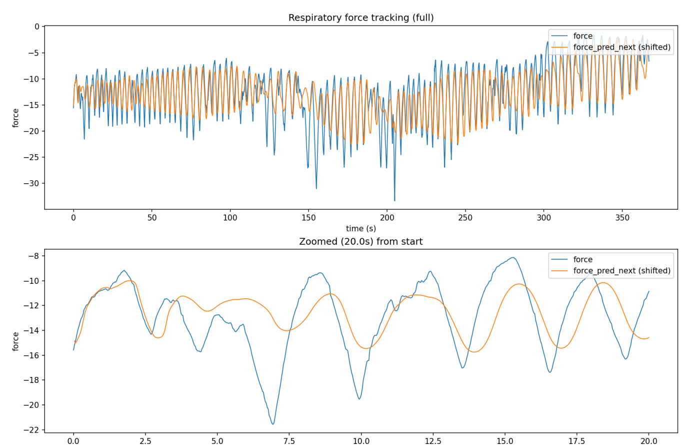


Fig. 7. Comparison of Contact Force Metrics: Uncompensated vs. Compensated Scenarios.

D. Real-World Experiments

We evaluated navigation robustness and force-interaction stability on the physical robotic system.

1) Navigation Robustness: Across 20 consecutive human-subject trials, the vision-based policy achieved a 95% success rate in reaching the abdominal ROI despite subject-position variations.

2) Force Control under Physiological Motion: Contact-force stability was evaluated during active breathing. Fig. 6 shows the relationship among respiratory phase, probe-tissue interaction, and force profile. Quantitative results in Fig. 7 show that the Respiratory Compensation Module suppresses thoracic-motion disturbance; despite minor latency-induced fluctuations, contact force remained within the safety range and image quality stayed consistent across subjects with varying Body Mass Indices (BMI).

E. Analysis of Real-Time Inference Performance

The *ACT-US* framework achieved an average inference latency of $12.9\,\mathrm{ms}$ and throughput of $77.32\,\mathrm{FPS}$, exceeding the $30\,\mathrm{Hz}$ sensor rate and enabling real-time closed-loop control.

V. CONCLUSION

We presented **ACT-US**, a hierarchical autonomy framework that bridges the visual-guided approach phase and force-guided contact phase in robotic ultrasound. Unlike methods that rely on the "Initialization Assumption," ACT-US integrates high-level generative imitation learning with low-level physiological adaptation for full-cycle autonomy.

ACT and a Conditional VAE model multi-modal expert behavior to generate smooth, singularity-free nominal trajectories, while a weakly-supervised respiratory-adaptive admittance controller compensates for thoracic uplift and modulates pressure from real-time acoustic coupling.

Real-world experiments showed that ACT-US outperforms visual-only and uncompensated baselines, achieving a 95% navigation success rate and reducing force variance by over 78% during active respiration.

REFERENCES

[1] P. Abolmaesumi, S. E. Salcudean, and W. H. Zhu, "Visual Servoing for Robot-Assisted Diagnostic Ultrasound," in *Proc. 22nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, USA, 2000, vol. 4, pp. 2532–2535.

[2] A. Antolin, N. Roson, M. Planes, M. Castillo, A. Alberti, and M. Escobar, "Validation of a Tele-Robotic Ultrasound System for Abdomen and Thyroid Gland Explorations: A Comparison with Standard Ultrasound," *The Ultrasound Journal*, vol. 17, no. 1, 2025.

[3] S. Avgousti, A. S. Panayides, A. P. Jossif, E. G. Christoforou, P. Vieyres, C. Novales, S. Voskarides, and C. S. Pattichis, "Cardiac Ultrasonography over 4G Wireless Networks Using a Tele-operated Robot," *Healthcare Technology Letters*, vol. 3, no. 3, pp. 212–217, 2016.

[4] P. Chatelain, A. Krupa, and N. Navab, "Confidence-Driven Control of an Ultrasound Probe," *IEEE Trans. Robot.*, vol. 33, no. 6, pp. 1410–1424, 2017.

[5] D. De Cunha et al., "The MIDSTEP System for Ultrasound Guided Remote Telesurgery," in *Proc. 20th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1998, vol. 3, pp. 1266–1269.

[6] E. Degoulange et al., "HIPPOCRATE: An Intrinsically Safe Robot for Medical Applications," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 1998, vol. 2, pp. 959–964.

[7] R. Finocchi, F. Aalamifar, T. Y. Fang, R. H. Taylor, and E. M. Boctor, "Co-Robotic Ultrasound Imaging: A Cooperative Force Control Approach," in *SPIE Medical Imaging*, 2017.

[8] Y. Huang, W. Xiao, C. Wang, H. Liu, R. Huang, and Z. Sun, "Towards Fully Autonomous Ultrasound Scanning Robot With Imitation Learning Based on Clinical Protocols," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 3671–3678, 2021.

[9] Z. Jiang, N. Danis, Y. Bi, M. Zhou, M. Kroenke, T. Wendler, and N. Navab, "Precise Repositioning of Robotic Ultrasound: Improving Registration-based Motion Compensation Using Ultrasound Confidence Optimization," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022.

[10] Z. Jiang, S. E. Salcudean, and N. Navab, "Robotic Ultrasound Imaging: State-of-the-Art and Future Perspectives," arXiv:2307.05545, 2023.

[11] A. Krupa, G. Fichtinger, and G. D. Hager, "Real-Time Tissue Tracking with B-Mode Ultrasound Using Speckle and Visual Servoing," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007*, Berlin, Heidelberg, 2007, pp. 1–8.

[12] S. Merouche, L. Allard, E. Montagnon, G. Soulez, P. Bigras, and G. Cloutier, "A Robotic Ultrasound Scanner for Automatic Vessel Tracking and Three-Dimensional Reconstruction of B-Mode Images," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 1, pp. 35–46, 2016.

[13] C. Nadeau and A. Krupa, "Intensity-Based Direct Visual Servoing of an Ultrasound Probe," in *2011 IEEE Int. Conf. Robot. Autom.*, 2011, pp. 5677–5682.

[14] C. Nadeau, A. Krupa, J. Petr, and C. Barillot, "Moments-Based Ultrasound Visual Servoing: From a Mono- to Multiplane Approach," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1558–1564, 2016.

[15] Z. Najdovski et al., "HaptiScan: A Haptically-Enabled Robotic Ultrasound System for Remote Medical Diagnostics," *Robotics*, vol. 13, no. 11, 2024.

[16] F. Pierrot et al., "Hippocrate: A Safe Robot Arm for Medical Applications with Force Feedback," *Med. Image Anal.*, vol. 3, no. 3, pp. 285–300, 1999.

[17] M. C. Roshan, M. Isaksson, A. Pranata, and E. M. Hidalgo, "A Sensor Fusion Approach to Autonomous Ultrasound Imaging of the Lumbar Region," *Biomed. Signal Process. Control*, vol. 108, 2025.

[18] S. E. Salcudean et al., "Robot-Assisted Diagnostic Ultrasound – Design and Feasibility Experiments," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI'99*, Berlin, Heidelberg, 1999, pp. 1062–1071.

[19] K. Su et al., "A Fully Autonomous Robotic Ultrasound System for Thyroid Scanning," *Nature Communications*, vol. 15, no. 1, 2024.

[20] D. Sun, A. Cappellari, B. Lan, M. Abayazid, S. Stramigioli, and K. Niu, "Automatic Robotic Ultrasound for 3D Musculoskeletal Reconstruction: A Comprehensive Framework," *Technologies*, vol. 13, no. 2, 2025.

[21] J. Tan et al., "A Flexible and Fully Autonomous Breast Ultrasound Scanning System," *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 3, pp. 1920–1933, 2023.

[22] X. Tang et al., "Autonomous Ultrasound Scanning Robotic System Based on Human Posture Recognition and Image Servo Control," *Front. Robot. AI*, vol. 11, 2024.

[23] S. Virga et al., "Automatic Force-Compliant Robotic Ultrasound Screening of Abdominal Aortic Aneurysms," in *2016 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2016, pp. 508–513.

[24] Z. Wang et al., "Full-Coverage Path Planning and Stable Interaction Control for Automated Robotic Breast Ultrasound Scanning," *IEEE Trans. Ind. Electron.*, vol. 70, no. 7, pp. 7051–7061, 2023.

[25] G.-Z. Yang et al., "Medical Robotics—Regulatory, Ethical, and Legal Considerations for Increasing Levels of Autonomy," *Sci. Robot.*, vol. 2, no. 4, 2017.

[26] R. Ye et al., "Feasibility of a 5G-Based Robot-Assisted Remote Ultrasound System for Cardiopulmonary Assessment of Patients With Coronavirus Disease 2019," *Chest*, vol. 159, no. 1, pp. 270–281, 2021.

[27] Y.-Q. Zhang et al., "A 5G-based Telerobotic Ultrasound System Provides Qualified Abdominal Ultrasound Services for Patients on a Rural Island," *Abdom. Radiol.*, vol. 49, no. 3, pp. 942–957, 2023.

[28] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," *arXiv preprint arXiv:2304.13705*, 2023.