This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Synthetic Data-Assisted Miniature Medical Robot Navigation via Ultrasound Imaging

Chunxiang Wang, Tianlu Wang, *Member, IEEE*, and Metin Sitti, *Fellow, IEEE*

***Abstract*—Wireless miniature robots are promising for minimally invasive biomedical applications. Effective tracking and navigation are essential for their safe deployment, but challenges persist in medical imaging and robot control, especially in localizing the robot in complex imaging scenes. Deep learning, though powerful for object identification, requires large supervised datasets, limiting its clinical applications due to the difficulty and cost of acquiring realistic data. Furthermore, miniature robots frequently exit the field of view of imaging systems, hindering continuous observation. Here, we present a framework for real-time magnetic navigation of wireless miniature robots using ultrasound imaging, leveraging synthetic data generation for deep learning-based detection. First, artificially generated synthetic data is combined with real data from synthetic materials to train a neural network capable of detecting versatile robots in real tissues. Then, a robotic system is developed to automatically track the robot with an ultrasound probe during magnetic actuation in tortuous lumens. With 85% less human-labeled data within synthetic materials, our approach effectively detects versatile robots in ex-vivo tissues, reducing data scarcity, imbalance, and manual labeling burdens. Demonstrations of automatic robot navigation through tortuous lumens in complex ultrasound scenes validate its effectiveness, enhancing the safe applicability of miniature medical robots in complex environments.**

***Index Terms*—Medical robotic systems, synthetic data-assisted neural network training, ultrasound imaging-based control.**

Received 3 December 2024; revised 25 April 2025; accepted 12 June 2025. The work of Chunxiang Wang and Metin Sitti was supported by the Max Planck Queensland Center for the Materials Science of Extracellular Matrices. The work of Chunxiang Wang, Tianlu Wang, and Metin Sitti was supported by the Max Planck Society and European Research Council (ERC) Advanced Grant SoMMoR Project under Grant 834531. The work of Tianlu Wang was supported by the University of Hawai'i at Mānoa. Recommended by Technical Editor A. K. Hoshiar and Senior Editor S. Katsura. *(Corresponding authors: Tianlu Wang; Metin Sitti.)*

Chunxiang Wang is with the Physical Intelligence Department, Max Planck Institute for Intelligent Systems, 70569 Stuttgart, Germany, and also with the Department of Information Technology and Electrical Engineering, ETH Zürich, 8092 Zürich, Switzerland (e-mail: chunxiang@is.mpg.de).

Tianlu Wang is with the Department of Mechanical Engineering, University of Hawaii at Mānoa, Honolulu, HI 96822 USA (e-mail: tianluw@hawaii.edu).

Metin Sitti is with the School of Medicine and College of Engineering, Koç University, Istanbul 34450, Türkey (e-mail: msitti@ku.edu.tr).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TMECH.2025.3583020.

Digital Object Identifier 10.1109/TMECH.2025.3583020

I. Introduction

WIRELESS miniature robots offer the possibility to access enclosed, hard-to-reach, tortuous, and narrow sites within the human body [1], facilitating minimally invasive medical operations, such as drug or cell delivery [2], vascular embolization [3], [4], and biosensing [5]. Effective robotic navigation is crucial for overcoming tissue barriers and reaching targeted locations for conducting these biomedical functions in clinical scenarios [6], [7], [8], [9], [10]. In this process, a medical imaging system continuously monitors the robot's status, providing feedback information for the active control of the robot's behavior under external fields like ultrasound [11] and magnetic fields [12], [13]. However, achieving robust tracking and controllable deployment of these miniature robots under medical imaging remains a significant challenge.

Robot localization in complex medical imaging scenes is the prerequisite for its deployment inside a living body. Various medical imaging modalities, such as X-ray [14], ultrasound [15], and magnetic resonance imaging [16], have been employed for this purpose and further robotic medical interventions. Ultrasound imaging stands out due to its radiation-free nature, portability, deep imaging depth, and high temporal resolution [15], [17]. Conventional B-mode ultrasound imaging has a low signal-to-noise ratio, making it difficult to distinguish millimeter-scale robots from surrounding biological tissues. Although Doppler ultrasound imaging offers better contrast, it requires continuous motion of the robot [18] or acoustically induced bubble oscillations [19] to generate signals, limiting applicability in scenarios like long-term stent angioplasty, where the device movement is disabled [20], [21]. Manual robot identification involves labor-intensive ultrasound probe adjustments and scanning the entire imaging field for the robot, making the process tedious and time consuming. Therefore, developing an automated robot localization system for noisy, low-contrast B-mode ultrasound imaging is essential.

Based on robot localization within the ultrasound imaging plane, continuous robot tracking and its coordination with actuation are necessary for navigating long distances to reach targeted locations. In clinical practice, tracking miniature robots during deployment requires the ultrasound probe to continuously adjust its position to keep the robot within the imaging plane during locomotion. Manual probe adjustments are labor-intensive, highlighting the need for autonomous probe movement. This is often achieved by integrating the probe into flexible motion systems, such as robotic arms or x-y-z stages [22], [23]. Recent studies have explored various techniques for robot tracking [24],

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

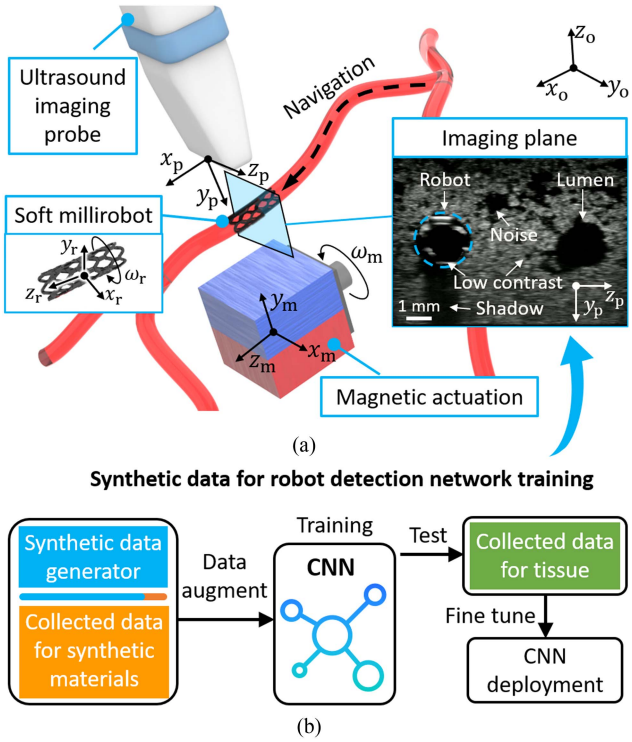


Fig. 1. Framework of ultrasound imaging-guided autonomous navigation of wireless millirobots with deep learning-based robot localization in complex environments. (a) Schematic of the magnetic navigation of a stent-structured robot monitored by ultrasound imaging. The blue window denotes the ultrasound imaging window. Scale bar represents 1 mm. (b) Deep neural network training for robot localization in tissues using synthetic data. Versatile millirobots can be effectively detected in ex-vivo tissues with 85% less human-labeled data within synthetic materials.

[25], including the utilization of pixel value distribution with the sweeping motion of a mobile ultrasound probe [26], tracking via Doppler signals induced by robot rotation [18]. However, these studies typically image robots in simplified environments without realistic features like lumen interference, inhomogeneous pixel distribution, low contrast, and pepper noise. Further investigation is needed to facilitate the robot utility in more complex environments.

Deep neural networks possess strong feature extraction capabilities and have achieved considerable success in both research and practice in recent years [27], [28], [29], [30], [31], [32]. However, these techniques require large amounts of data, and applying them to clinical practice faces the challenge of data scarcity for real animals, due to the difficulty and high cost for data collection [15]. Additionally, biased datasets from specific samples may cause overfitting, hindering model generalization to other medical scenarios [33]. Domain randomization, which randomizes imaging parameters like lighting, object textures, and pose, can mitigate data collection efforts [34]; but its clinical applicability needs further exploration.

To address these challenges, we propose a framework for real-time magnetic navigation of wireless miniature robots using ultrasound imaging, leveraging synthetic data generation for training deep learning-based robot detection algorithms. As illustrated in Fig. 1(a), a magnetic soft millirobot is actuated by a translating and rotating permanent magnet to navigate through tortuous lumens, while an ultrasound imaging probe on a 7 degrees-of-freedom (DOF) robotic arm automatically follows the robot to keep it inside the imaging plane. For continuous robot tracking, we utilize synthetic data to train a deep neural network capable of detecting versatile robots in noisy unstructured ultrasound scenes, as shown in Fig. 1(b). Using 85% less human-labeled data from synthetic materials, with 100 real images and 800 synthetic images instead of 600 real ones, three types of soft millirobots can be effectively detected in ex-vivo tissues, reducing data scarcity, imbalance, and manual labeling burdens. We validate the proposed navigation strategy by controlling and tracking the robot in real time to traverse tortuous lumens inside tissue mimicking phantoms, despite the noisy, unstructured, and low-contrast imaging scenes. The proposed framework, which coordinates medical imaging and magnetic actuation, has great potential to enhance the safe applicability of miniature soft robots for medical purposes in complex environments inside the body.

II. Synthetic Data-Assisted Robot Localization

The workflow of the synthetic data-assisted deep neural network training for robot localization is outlined in Fig. 1(b). The training dataset comprises generated synthetic data that simulate real imaging scenarios and robots within synthetic materials. This approach mitigates the high cost of acquiring extensive robot data within various tissues, the labor-intensive manual labeling, and the risk of overfitting in deep neural networks. Furthermore, compared to conventional data augmentation methods, which involve techniques such as geometric transformations and brightness adjustments [35], data generation offers the following advantages [33], [34].

1) *Addressing Data Scarcity:* Enables the creation of rare objects or scenarios where real-world data is unavailable or expensive to collect.
2) *Customizability:* Provides control over dataset properties, such as object pose, lighting, and background, allowing targeted learning for specific conditions.
3) *Efficient Annotation:* Automatically generated datasets include ground-truth labels, such as bounding boxes and segmentation masks, reducing manual labeling time and errors.
4) *Scalability:* Facilitates the rapid generation of large datasets, supporting the training of data-intensive deep neural networks.
5) *Diverse and Balanced Datasets:* Ensures comprehensive representation of all classes and variations, such as small objects and occluded views, preventing class imbalance.

The dataset is then used to train the object detection neural network, where data augmentation techniques are further applied to enhance the training dataset. For object detection, we selected the convolutional neural network (CNN) YOLO v5s model [36], chosen for its high detection accuracy and suitability for edge devices, effectively balancing precision and real-time processing requirements. Following training, the model's performance is

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

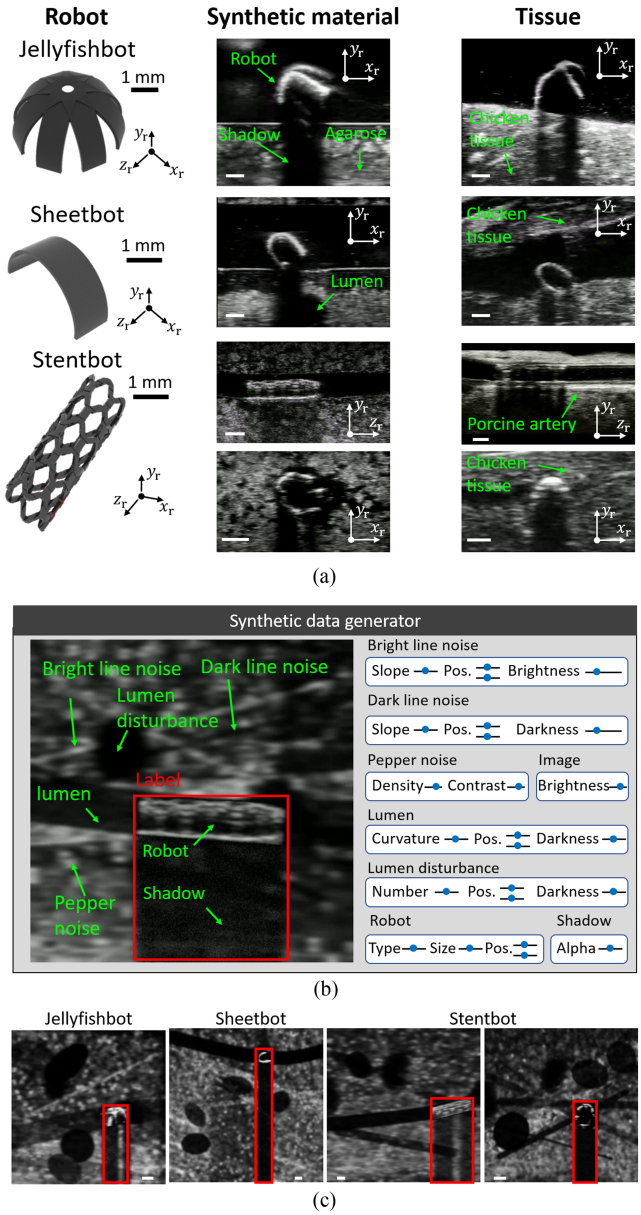


Fig. 2. Dataset for neural network training. (a) Imaging effect of various magnetic soft millirobots. The imaging depth is 15 mm. (b) User interface for synthetic data generation. (c) Generated synthetic data. The red rectangle represents the label for neural network training. In all figures, scale bars represent 1 mm.

evaluated using robot data within tissues, and the model weights yielding the highest precision are selected for deployment. It should be noted that any object detection can be trained using this synthetic data generation-based framework, not limited to a specific model.

We employ this training framework to detect various magnetic soft millirobots, including jellyfish robots (jellyfishbot) [37], sheet-shaped robots (sheetbot) [38], and stent-structured robots (stentbot) [3]. Ultrasound imaging of these robots in tissue-mimicking phantoms and real tissues is illustrated in Fig. 2(a). The complex imaging scenes present challenges, such as inhomogeneous pixel distribution, lumen interference, pepper noise, and low contrast between the robot and its surroundings, especially when embedded in tissues. In B-mode ultrasound imaging, a shadow artifact appears below the robot due to sound waves being reflected to the transducer or absorbed by the tissue [39]. This shadow offers an important feature for localizing the robot, and is included in the labeling to ensure the neural network correlates it with the robot, thereby enhancing model generalization.

To address data scarcity, imbalance, and neural network overfitting, we develop a synthetic data generator using custom Python code to create images simulating robots in complex scenes, as shown in Fig. 2(b). The synthetic image generation process involves the following steps.

1) *Background Preparation:* Random bright and dark lines are introduced to the synthetic material image to simulate background variations, enhancing the diversity of the training dataset.
2) *Lumen Simulation:* Black ellipses are added to represent lumen disturbances, followed by the incorporation of a randomly shaped curve to simulate the actual lumen, contributing to realistic internal structure simulation.
3) *Robot Placement:* The robot is randomly placed within the simulated lumen, randomizing its size, position, shape, and brightness. This step contributes to more robust training and generalization enhancement.
4) *Shadow Artifact Creation:* A shadow artifact is generated beneath the robot with adjustable opacity, simulating the artifacts commonly observed in ultrasound imaging, thus improving the model's ability to handle realistic imaging conditions.
5) *Noise and Blurring:* Pepper noise is applied to introduce pixel-level disturbances, and the image is subsequently blurred to soften the edges, mimicking real-world image imperfections and aiding in model robustness.

The background and robot images are extracted from captured images from synthetic materials. All features are adjustable via a user-friendly interface [Fig. 2(b)], enabling the creation of diverse images for CNN training, as demonstrated in Fig. 2(c). It should be noted that the influence of the phantom size and imaging depth are simulated by adjusting the position and size of the robot and lumen during simulation, and by tuning noise and blur levels to mimic ultrasound intensity variations. This flexibility allows for the incorporation of ultrasound images of various robots and materials, enhancing the diversity of the synthetic data.

III. Ultrasound Imaging-Guided Robotic Navigation

In this section, we propose a supervised autonomous robotic navigation system for tortuous lumens. Different from tethered medical devices whose positions can be estimated from their insertion point, the risk of losing track of wireless robots is higher during actuation. To address this issue, we develop a robot tracking strategy that uses an ultrasound imaging probe to automatically follow the robot, ensuring it remains within the imaging plane over long distances. As shown in Fig. 3(a), there are the following four core modules in this system, including:

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

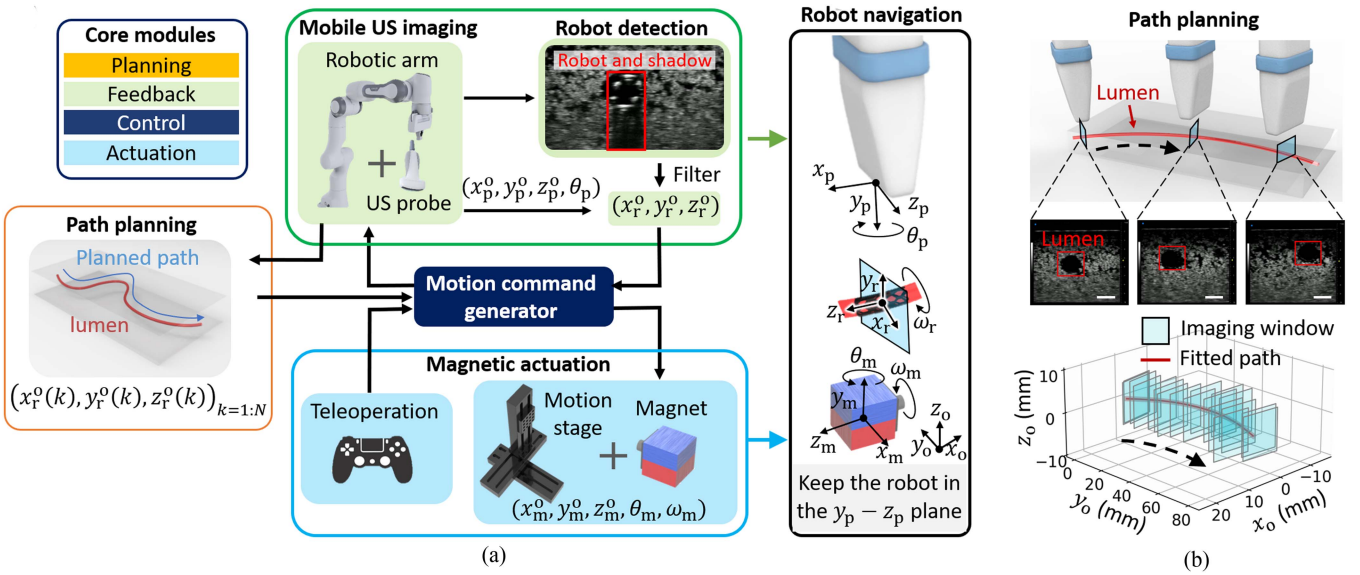


Fig. 3. Schematic of the supervised autonomous robotic navigation system in tortuous lumens. (a) Modular structure of the navigation system, consisting of four core modules: Planning, feedback, control, and actuation. (b) Path planning by ultrasound scanning. Lumens in the ultrasound images are detected and connected to reconstruct their shape for robot deployment. The blue window indicates the ultrasound imaging window, while the red line represents the fitted robot or lumen path. In all figures, scale bars represent 2 mm.

1) *Path Planning:* Extract the shape of the targeted lumen for robot deployment.
2) *Actuation:* Actuate the magnetic millirobot with a translating and rotating permanent magnet.
3) *Feedback:* Detect the robot within the imaging plane using a pretrained CNN.
4) *Control:* Coordinate the movement of the magnet and the ultrasound probe for actuation and tracking.

In this study, we focus on the robotic navigation of the stentbot with a permanent magnet. Four coordinate frames are defined for robot deployment. x_{o}-, y_{o}-, z_{o}-axes denote the global coordinate frame, attached to the initial position of the deployed lumen. x_{m}-, y_{m}-, z_{m}-axes represent the coordinate frame attached to the magnet with the rotating speed ω_{m} around z_{m} axis and the orientation angle θ_{m} around y_{m} axis. x_{p}-, y_{p}-, z_{p}-axes indicate the coordinate frame attached to the ultrasound probe with the orientation angle θ_{p}. x_{r}-, y_{r}-, z_{r}-axes represent the coordinate frame attached to the stenbot with the rotating speed ω_{r} around z_{r}-axis. In addition, the superscript denotes the base frame, while the subscript represents the agent. For example, $x_{\mathrm{r}}^{\mathrm{o}}$ represents the robot position along the x_{o}-axis of the global frame.

A. Path Planning Via Ultrasound Scanning

The first step in robotic navigation is extracting the robot's path through sample scanning. As illustrated in Fig. 3(b), the ultrasound probe scans the sample to locate the lumen, which appears as a black hole in the captured image. During scanning, the scanning probe moves at intervals of 5 mm along the y_{o} -axis and 8mm along the x_{o}-axis. The angle θ_p is adjusted to minimize the area of the black hole at each scanning point. We utilize the data-efficient object detection algorithm in Section II to localize lumens in the image, with the detected lumen center denoted as $(z_{\mathrm{l}}^{\mathrm{p}}, y_{\mathrm{l}}^{\mathrm{p}})$ in the ultrasound probe coordinate frame. The lumen center point coordinate $(x_{\mathrm{l}}^{\mathrm{o}},\ y_{\mathrm{l}}^{\mathrm{o}},\ z_{\mathrm{l}}^{\mathrm{o}})$ in the global frame is calculated with

$$\begin{bmatrix} x_{\mathrm{l}}^{\mathrm{o}} \\ y_{\mathrm{l}}^{\mathrm{o}} \\ z_{\mathrm{l}}^{\mathrm{o}} \end{bmatrix} = \mathbf{R}_{\mathrm{p}}^{\mathrm{o}} \begin{bmatrix} 0 \\ y_{\mathrm{l}}^{\mathrm{p}} \\ z_{\mathrm{l}}^{\mathrm{p}} \end{bmatrix} + \begin{bmatrix} x_{\mathrm{p}}^{\mathrm{o}} \\ y_{\mathrm{p}}^{\mathrm{o}} \\ z_{\mathrm{p}}^{\mathrm{o}} \end{bmatrix} \tag{1}$$

where $\mathbf{R}_{\mathrm{p}}^{\mathrm{o}}$ represents the rotation matrix from the global coordinate frame to the ultrasound probe coordinate frame. Using the lumen center points in each imaging window in Fig. 3(b), A* algorithm [40] is further utilized to create the lumen path. The path consists of a series of discrete points $(x_{\mathrm{l}}^{\mathrm{o}},\ y_{\mathrm{l}}^{\mathrm{o}},\ z_{\mathrm{l}}^{\mathrm{o}})$ in a specific order. Further, the lumen path is interpolated with a three-order spline curve at the interval of 1mm to generate the robot path, which is on the lumen path. The resulting robot path is represented as

$$\mathbf{L}_{\mathrm{r}} = (x_{\mathrm{r}}^{\mathrm{o}}(k), y_{\mathrm{r}}^{\mathrm{o}}(k), z_{\mathrm{r}}^{\mathrm{o}}(k))_{k=1:N} \tag{2}$$

where N is the number of via points. The fitted robot path is denoted as the red line in Fig. 3(b). Based on the planned robot path, the paths of the ultrasound probe are also generated, represented as

$$\mathbf{L}_{\mathrm{p}} = (x_{\mathrm{p}}^{\mathrm{o}}(k), y_{\mathrm{p}}^{\mathrm{o}}(k), z_{\mathrm{p}}^{\mathrm{o}}(k), \theta_{\mathrm{p}}(k))_{k=1:N} \tag{3}$$

where $x_{\mathrm{p}}^{\mathrm{o}}(k) = x_{\mathrm{r}}^{\mathrm{o}}(k)$, and $y_{\mathrm{p}}^{\mathrm{o}}(k) = y_{\mathrm{r}}^{\mathrm{o}}(k)$. To avoid collision with the sample, the ultrasound probe is displaced along the z_{o}-axis at Δz_{p} with $z_{\mathrm{p}}^{\mathrm{o}}(k) = z_{\mathrm{r}}^{\mathrm{o}}(k) + \Delta z_{\mathrm{p}}$. Meanwhile, θ_{p} is adjusted to align the x_{p}-axis parellel to the z_{r}-axis. It should be noted that it is more feasible to detect the robot when the x_{p}-axis parallel to the z_{r}-axis, resulting in a smaller ratio between the robot's area and the entire imaging area. Conversely, aligning the z_{p}-axis with the z_{r}-axis means the robot is visible only when the imaging plane precisely aligns with the y_{r}-z_{r} plane, significantly increasing tracking difficulty due to planning errors.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

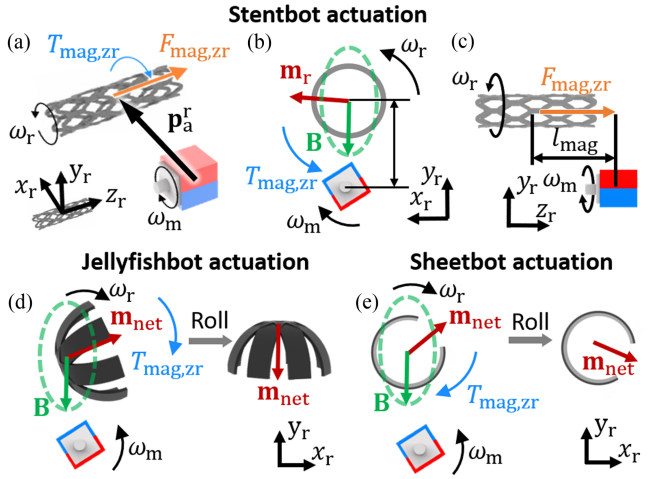


Fig. 4. Schematic of the magnetic actuation. (a) Magnetic torque and force applied to the stentbot along the z_r-axis during the actuation. (b) Magnetic torque on the x_r-y_r plane. (c) Magnetic force on the y_r-z_r plane. (d) Magnetic torque applied to the jellyfishbot on the x_r-y_r plane for rolling. (e) Magnetic torque applied to the sheetbot on the x_r-y_r plane for rolling. The dotted green ellipsis denotes the magnitude of the magnetic field projected on the x_r-y_r plane during the magnet rotation.

B. Magnetic Actuation

We utilize a rotating and translating permanent magnet to actuate the stentbot, leveraging both magnetic torque and force, as illustrated in Fig. 4(a)–(c). Magnetic torque is primarily used to roll the jellyfishbot and sheetbot for data collection, as shown in Fig. 4(d) and (e). The magnetic field $\mathbf{B}$ is calculated with

$$\mathbf{B}\left(\mathbf{p}_{\mathrm{a}}^{\mathrm{r}}\right)=\frac{\mu_0 \mathbf{H}\mathbf{m}_{\mathrm{a}}}{4\pi\|\mathbf{p}_{\mathrm{a}}^{\mathrm{r}}\|^3} \tag{4}$$

where $\mathbf{p}_{\mathrm{a}}^{\mathrm{r}}$ is the vector from the permanent magnet to the robot, $\mathbf{H}=3\hat{\mathbf{p}}_{\mathrm{a}}^{\mathrm{r}}(\hat{\mathbf{p}}_{\mathrm{a}}^{\mathrm{r}})^T-\mathbf{I}$, $\mu_0=4\pi\times10^{-7}\,\mathrm{N}\cdot\mathrm{A}^{-2}$ is the permeability of free space, $\mathbf{m}_{\mathrm{a}}$ is the magnetic moment of the permanent magnet, and $\mathbf{I}\in\mathrm{R}^{3\times3}$ is the identity matrix. As shown in Fig. 4(b), the magnetic field magnitude generated by the permanent magnet fluctuates in an elliptical fashion with the rotation axis $\widehat{\boldsymbol{\omega}}_{\mathrm{b}}=\widehat{\mathbf{H}^{-1}\widehat{\boldsymbol{\omega}}_{\mathrm{m}}}$, and $\widehat{\boldsymbol{\omega}}_{\mathrm{m}}$ is the rotation axis of the magnet [4]. As illustrated in Figs. 3(a), 4(b) and 4(c), the robot rotates in sync with the magnetic field. The stentbot overcomes the resistance to move forward under the magnetic actuation, including the magnetic torque around the z_r-axis, $T_{\mathrm{mag,zr}}$, and magnetic force along the z_r-axis, $F_{\mathrm{mag,zr}}$. The stentbot rotation is initiated by $T_{\mathrm{mag,zr}}$ to overcome the friction and fluid drag, and $F_{\mathrm{mag,zr}}$ further pulls the stentbot forward or backward. $T_{\mathrm{mag,zr}}$ and $F_{\mathrm{mag,zr}}$ are computed with the dipole model as

$$T_{\mathrm{mag,zr}}=\left(\mathbf{m}_{\mathrm{r}}\times\mathbf{B}\left(\mathbf{p}_{\mathrm{a}}^{\mathrm{r}}\right)\right)\cdot\hat{\boldsymbol{\omega}}_{\mathrm{r}}, \tag{5}$$

$$F_{\mathrm{mag,zr}}=\left((\mathbf{m}_{\mathrm{r}}\cdot\nabla)\mathbf{B}\left(\mathbf{p}_{\mathrm{a}}^{\mathrm{r}}\right)\right)\cdot\hat{\boldsymbol{\omega}}_{\mathrm{r}} \tag{6}$$

where $\mathbf{m}_{\mathrm{r}}$ is the magnetic moment of the stentbot, and $\hat{\boldsymbol{\omega}}_r$ is the unit vector along the z_r-axis and the rotation axis of the stentbot. For the actuation of jellyfishbots and sheetbots, their net magnetic moment $\mathbf{m}_{\mathrm{net}}$ is used for calculating $T_{\mathrm{mag,zr}}$.

Algorithm 1: Robotic Actuation and Tracking.

Require: $\mathbf{L}_{\mathrm{r}}$, $\mathbf{L}_{\mathrm{p}}$ ▷ Pre-planned paths
Ensure:
$i \leftarrow i_{\mathrm{init}}$, $k \leftarrow i_{\mathrm{init}}$, send motion command $\mathbf{L}_{\mathrm{p}}(i)$
while $i \leq N$ **do** ▷ N is the number of via-points
  if $0.1\sum_{c=1}^{10} P_{\mathrm{r}}(c) > 0.7$ **then** ▷ Robot is detected
    $k \leftarrow i$
    Send user input u ▷ $u>0$ for moving forward
    Send motion command
    $(x_{\mathrm{m}}^{\mathrm{o}}(k), y_{\mathrm{m}}^{\mathrm{o}}(k), z_{\mathrm{m}}^{\mathrm{o}}(k), \theta_{\mathrm{m}}(k), \omega_{\mathrm{m}})$ with Eqn. 7
  else if $u>0$ **then** ▷ Search forward
    $\omega_{\mathrm{m}} \leftarrow 0$
    if $i \leq (k+10)$ **then**
      $i \leftarrow i+1$
      Send motion command $\mathbf{L}_{\mathrm{p}}(i)$
    else
      $i \leftarrow k-10$
    end if
  else if $u<0$ **then** ▷ Search backward
    $\omega_{\mathrm{m}} \leftarrow 0$
    if $i \geq (k-10)$ **then**
      $i \leftarrow i-1$
      Send motion command $\mathbf{L}_{\mathrm{p}}(i)$
    else
      $i \leftarrow k+10$
    end if
  end if
end while

The magnet needs to lead the stentbot to ensure $|F_{\mathrm{mag,zr}}|>0$ for actuation. Based on $\mathbf{L}_{\mathrm{r}}$ in Section III-A, the magnet position is generated with

$$\begin{bmatrix} x_{\mathrm{m}}^{\mathrm{o}}(k) \\ y_{\mathrm{m}}^{\mathrm{o}}(k) \\ z_{\mathrm{m}}^{\mathrm{o}}(k) \end{bmatrix} = \mathbf{R}_{\mathrm{m}}^{\mathrm{o}} \begin{bmatrix} 0 \\ 0 \\ l_{\mathrm{mag}} \end{bmatrix} + \begin{bmatrix} x_{\mathrm{r}}^{\mathrm{o}}(k) \\ y_{\mathrm{r}}^{\mathrm{o}}(k) \\ z_{\mathrm{r}}^{\mathrm{o}}(k) - \Delta z_{\mathrm{m}} \end{bmatrix} \tag{7}$$

where $\mathbf{R}_{\mathrm{p}}^{\mathrm{m}}$ represent the rotation matrix from the global coordinate frame to the magnet coordinate frame, and l_{mag} is the displacement between the magnet and the stentbot along the z_{m}-axis, as shown in Fig. 4(c). $l_{\mathrm{mag}}>0$ when the stentbot moves forward along the z_{m}-axis, and $l_{\mathrm{mag}}<0$ for moving backward. To prevent collision with the sample, the magnet is displaced along the z_{o}-axis at Δz_{m} with $z_{\mathrm{m}}^{\mathrm{o}}(k)=z_{\mathrm{r}}^{\mathrm{o}}(k)-\Delta z_{\mathrm{m}}$. Further, θ_{m} is adjusted to ensure z_{m}-axis parellel to z_{r}-axis.

C. Real-Time Robotic Navigation and Tracking

The goal of robotic navigation is to maintain continuous visibility of the miniature robot within the ultrasound imaging plane throughout its journey to the destination. Given the dynamic environment and nonlinear behavior of the robot in response to the magnetic field, accurately predicting its position based on magnet location and the dynamic model is challenging, increasing the risk of losing track of the robot. Therefore, we introduce a generic tracking-via-searching method to address these challenges.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Based on the preplanned paths, navigation involves two iterative steps: 1) actuation; and 2) tracking, corresponding to the movement of the magnetic actuator and the ultrasound imaging probe, respectively. Algorithm 1 outlines the process: Initially, the ultrasound probe moves to capture the initial robot at the via-point $\mathbf{L}_{\mathrm{p}}(i_{\mathrm{init}})$. Upon receiving the user motion command u dictating the robot moving direction, the magnet adjusts to the actuation position calculated using (7) and rotates at frequency ω_{m}. During actuation, the CNN analyzes the captured image to localize the robot, along with a filtering step to remove false positive detections based on detection confidence and bounding box shape, and the average detection confidence P_{r} of ten frames is utilized for robot tracking, minimizing the influence of false positives. The robot keeps moving until it moves outside the imaging plane, followed by the tracking step. In the tracking process, the magnet rotation is disabled as the ultrasound probe scans the lumen in the direction of robot movement. If the robot cannot be detected within 10 via points (10 mm), the probe reverses by 10 mm to search. This span is sufficient to re-track the robot since the magnet keeps rotating at a fixed position. Upon detection, ultrasound scanning ceases, and the magnet resumes rotating to continue actuation.

In Algorithm 1, we focus on real-time robot tracking. For detailed robot 3D position, the robot is further scanned by the ultrasound probe. As shown in Fig. 5, the ultrasound probe scans robot with the x_{p}-axis or z_{p}-axis parallel to z_{r}-axis, respectively. The robot's position in the image can be calculated with

$$\begin{bmatrix} x_{\mathrm{r}}^{\mathrm{o}} \\ y_{\mathrm{r}}^{\mathrm{o}} \\ z_{\mathrm{r}}^{\mathrm{o}} \end{bmatrix} = \mathbf{R}_{\mathrm{p}}^{\mathrm{o}} \begin{bmatrix} 0 \\ y_{\mathrm{r}}^{\mathrm{p}} \\ z_{\mathrm{r}}^{\mathrm{p}} \end{bmatrix} + \begin{bmatrix} x_{\mathrm{p}}^{\mathrm{o}} \\ y_{\mathrm{p}}^{\mathrm{o}} \\ z_{\mathrm{p}}^{\mathrm{o}} \end{bmatrix} \tag{8}$$

where $\mathbf{R}_{\mathrm{p}}^{\mathrm{o}}$ represents the rotation matrix from the global coordinate frame to the ultrasound probe coordinate frame. Further, the robot 3D position $\mathbf{p}_{\mathrm{robot}}$ can be estimated with

$$\mathbf{p}_{\mathrm{robot}} = \frac{1}{M} \sum_{i=1}^{M} \begin{bmatrix} x_{\mathrm{r}}^{\mathrm{o}}(i) \\ y_{\mathrm{r}}^{\mathrm{o}}(i) \\ z_{\mathrm{r}}^{\mathrm{o}}(i) \end{bmatrix} \tag{9}$$

where M is the number of captured robot images.

Navigation safety is ensured via two key strategies. First, the robot is engineered for shape adaptation to varying lumen diameters, self-anchoring to maintain position during external actuation loss, traversal through curved paths and bifurcations, and mobility against blood flow for retrieval or correction [3], [4]. Second, the path planning system enables retracking in case of displacement, with the ultrasound probe scanning the entire route to locate the robot.

IV. Experiments

A. System Setup

The system included the mobile ultrasound imaging component and the magnetic actuation component controlled by a host computer, as depicted in Fig. 6. The preclinical ultrasound

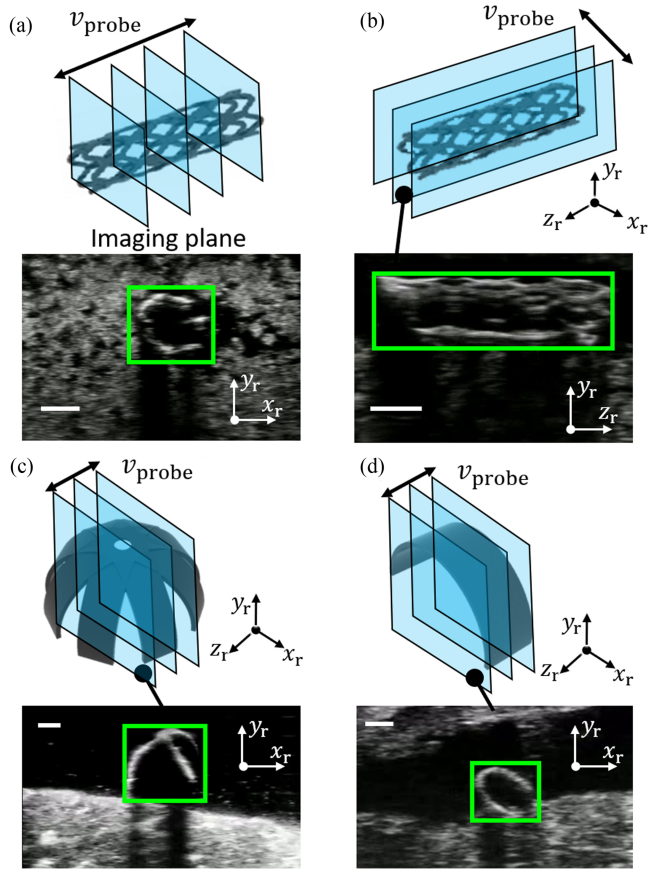


Fig. 5. Estimation of the robot's 3-D position via ultrasound probe scanning. (a) Estimating the stentbot position with x_{p}-axis parallel to z_{r}-axis. (b) Estimating the stentbot position with z_{p}-axis parallel to z_{r}-axis. (c) Estimating the jellyfishbot position. (d) Estimating the sheetbot position with x_{p}-axis parallel to z_{r} -axis. The green rectangles are the bounding boxes of robots. In all figures, scale bars represent 1mm.

imaging device (Vevo 3100 and MX550, FUJIFILM VisualSonics, Inc.) was mounted on a 7-DOF robotic arm (Franka Emika GmbH) with the positionsing precision of 1 mm. The 5-DOF magnetic actuation component included an X-Y-Z motion stage (LTS300/M, Thorlabs Inc.), a stepper motor (535-0372, RS Components GmbH) and a servo motor (SKU 900-00360, Parallax Inc.), and a 20 mm cubic magnet (N45, IMPLOTEX GmbH). The robotic arm was connected to a low-level PC with the Linux real-time kernel, and the host computer serves as the central control unit for all system components.

The communication framework was built using the Robot Operating System (ROS) Noetic, organized into four primary nodes: A motion command generator, a magnetic controller, an arm controller, and an image processing node. Commands from a joystick triggered the motion command generator node to publish instructions to the magnetic controller, directing the rotating permanent magnet to adjust its position and orientation accordingly. Subscribing to the image processing node, the motion command generator node delivered the motion command to the arm controller node to control the ultrasound probe for robot tracking.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

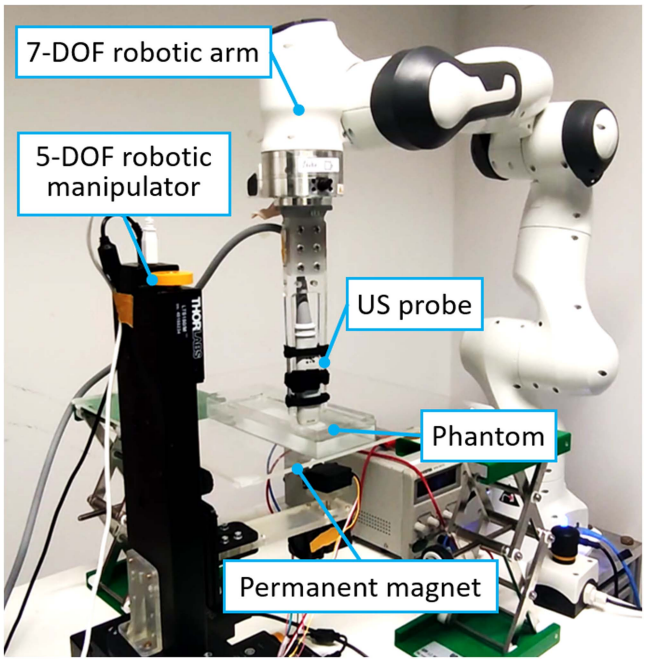


Fig. 6. System hardware setup. The system includes a mobile ultrasound imaging component and a magnetic actuation component, which are controlled by a slave computer and a host computer.

B. Preparation of Synthetic Materials

The phantoms were fabricated using agarose gels containing internal lumens. Positive molds, shaped and sized using a 3-D printer (Form 3, Formlabs Inc.), were employed. Agarose powder (BioReagents A9539, Sigma-Aldrich Co.) was added to deionized water (1.0 wt %) at 90 °C under continuous stirring until complete dissolution. Subsequently, the solution was boiled for 5 minutes to ensure complete dissolution and cooled down to 60°C. Additional agarose powder (1.0wt %) was added to the solution and stirred up to make it evenly distributed, which was poured into the petri dish containing the positive molds. After cooling to room temperature (approximately 24°C) for 30 min, the positive mold was extracted. Porcine hearts and chickens were sourced as animal by-products with proper permits and registrations issued by authorities overseeing food control, consumer protection, and veterinary services in Stuttgart. Before testing, tissues were cleaned with Phosphate Buffered Saline.

C. Performance of Robot Localization

To evaluate the synthetic data-assisted robot detection neural network, we collected the ultrasound imaging data from synthetic materials and ex-vivo tissues, including 600 labeled images of synthetic materials and 200 labeled images of ex-vivo tissues. The dataset includes four robot types: jellyfishbot, sheetbot, and two orientations of stentbot (in the x_r-y_r plane and y_r-z_r plane), each type accounts for 25% of the dataset. The training dataset includes images of synthetic materials and synthetic images, while the testing dataset involves the images of ex-vivo tissues. This comprehensive dataset facilitates the evaluation of the neural network's performance across diverse robotic scenarios and imaging conditions.

TABLE I
PERFORMANCE OF THE DETECTION MODEL (METRICS: MAP@0.5)

n_syn \ n_real	40	100	600
0	0.697	0.762	0.963
400	0.966	0.936	0.978
800	0.931	0.978	0.977
1600	0.943	0.967	0.963
3200	0.952	0.972	0.978

TABLE II
PERFORMANCE OF THE DETECTION MODEL (METRICS: MAP@0.5:0.95)

n_syn \ n_real	40	100	600
0	0.365	0.487	0.718
400	0.696	0.625	0.731
800	0.645	0.714	0.753
1600	0.673	0.721	0.709
3200	0.627	0.718	0.734

We conducted multiple experiments modifying the number of synthetic data (n_syn) and real data of synthetic materials (n_real), evaluating the model performance on tissue data. Training images were randomly selected from the collected dataset. Two metrics were adopted: mAP@0.5 and mAP@0.5:0.95, and mAP represents the mean average precision (AP) of all classes. mAP@0.5 considers the prediction a true positive if the intersection over union (IoU) is over 0.5, while mAP@0.5:0.95 averages the AP values ranging from 0.5 to 0.95. Both metrics yield a maximum score of 1, with higher values indicating superior performance, as detailed in [41]. These metrics provide comprehensive insights into the neural network's accuracy across different detection thresholds and are crucial for assessing its robustness in varied experimental conditions.

The evaluation results are presented in Tables I and II. The model was trained for 40 epochs using pre-trained weights from the COCO dataset [42], with variations in n_real-n_syn ratio. It should be noted that all data are augmented during the training process. For small-sized datasets, there is a significant enhancement of the tracking accuracy after incorporating synthetic data. For instance, with $n_\mathrm{real} = 40$ and $n_\mathrm{syn} = 400$ (10 real images per robot type), mAP@0.5 and mAP@0.5:0.95 witnessed an increase of 38.7% and 90.3%, respectively. Moreover, models trained with synthetic data can achieve comparable performance to those trained with a larger dataset ($n_\mathrm{real} : n_\mathrm{syn} = 100 : 1600$ versus $n_\mathrm{real} : n_\mathrm{syn} = 600 : 0$), Additionally, the incorporation of synthetic data can improve mAP@0.5:0.95 by 4.9% for $n_\mathrm{real} = 600$. These findings underscore the efficacy of synthetic data in improving detection accuracy across varying dataset sizes.

Integrating synthetic data also enhances model generalization to new imaging scenes and robot apperances. The CNN was trained using real synthetic material images and artificial synthetic images, and tested on ex-vivo tissues to assess deployment feasibility in clinical settings. As demonstrated in Fig. 7 and Supplementary Video 1, the CNN trained with a smaller

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

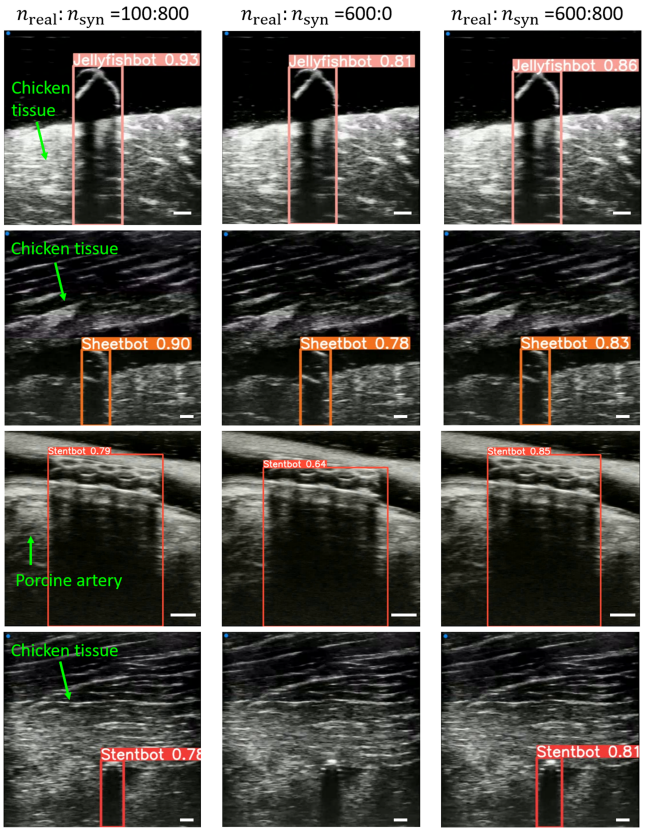


Fig. 7. Robot localization in ex-vivo tissues utilizing the model trained with synthetic data and real data of the synthetic materials. Green arrows indicate the tested tissues, including the chicken breast and porcine heart arteries. Rectangles denote the localization bounding box with robot type and detection confidence above it. In all figures, the imaging depth is 15 mm, and scale bars represent 1 mm.

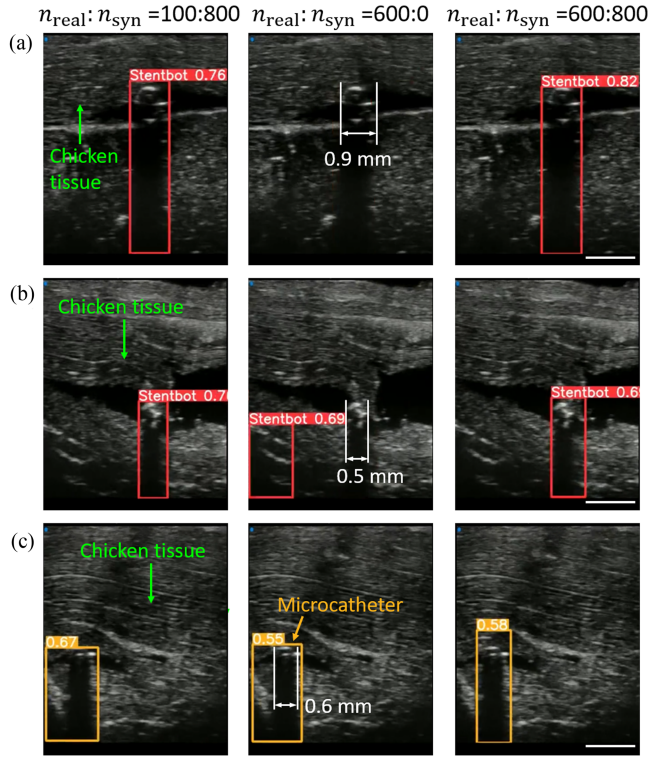


Fig. 8. Evaluation of model generalization trained with the synthetic data framework. (a) Detection of stentbots with 0.9 mm outer diameter. (b) Detection of stentbots with 0.5 mm outer diameter. (c) Detection of microcatheters with 0.6 mm outer diameter. Green arrows indicate the tested tissues, and white lines present the device size. Rectangles denote the localization bounding box with robot type and detection confidence above it. In all figures, the imaging depth is 8 mm, and scale bars represent 1 mm.

dataset ($n_{\text{real}} : n_{\text{syn}} = 100 : 800$) achieved superior localization compared to one trained with a larger dataset of pure real data ($n_{\text{real}} = 600$). This improvement includes higher object detection confidence and more accurate bounding boxes, particularly evident in challenging scenarios like stentbot detection within chicken tissue with partial occlusion. In such cases, the CNN trained solely on real data failed to identify the robot where the integrated synthetic data model ($n_{\text{real}} : n_{\text{syn}} = 100 : 800$) succeeds with high confidence scores. After integrating synthetic data, the generalization of the model trained on pure real data notably improved ($n_{\text{real}} : n_{\text{syn}} = 600 : 800$), as depicted in the third column of Fig. 7. The results also demonstrate that synthetic data leads to improved detection performance compared to conventional data augmentation methods.

To evaluate the generalization capability across variations in robot appearance, we employed submillimeter-scale stentbots and microcatheters, while the model was exclusively trained on a stentbot with a 1.5 mm outer diameter. As shown in Fig. 8 and Supplementary Video 2, lower-resolution and noisier imaging conditions increased the incidence of missed detections and false negatives. For stentbots with 0.9 mm and 0.5 mm outer diameters, models trained on synthetic data successfully detected the robots, whereas models trained solely on real data were more prone to false negatives. Additionally, for microcatheters that are absent from the training dataset, models incorporating both real and synthetic images achieved more accurate localization. This framework also enables the rapid training of learning-based state-of-the-art object detection models for tracking robots with novel designs and markedly different shapes under ultrasound imaging, minimizing the need for extensive data collection and labeling.

D. Real-Time Robotic Navigation in Tortuous Lumens

We validated the proposed robotic navigation and tracking framework on the stentbot within tortuous lumens, illustrated in Fig. 9 and Supplementary Video 3. Utilizing preplanned robot paths derived from ultrasound scans, the mobile ultrasound probe tracked the robot during magnetic actuation. In the ultrasound tracking process, the robot is detected utilizing the pretrained CNN followed by a filtering step to remove any false positive detections based on detection confidence and bounding box shape. The detection is performed in real time with an average processing speed of 51.5 frames per second (fps) using an ordinary GPU (Nvidia Quadro T1000). Despite challenges such as lumen interference and noisy, unstructured,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

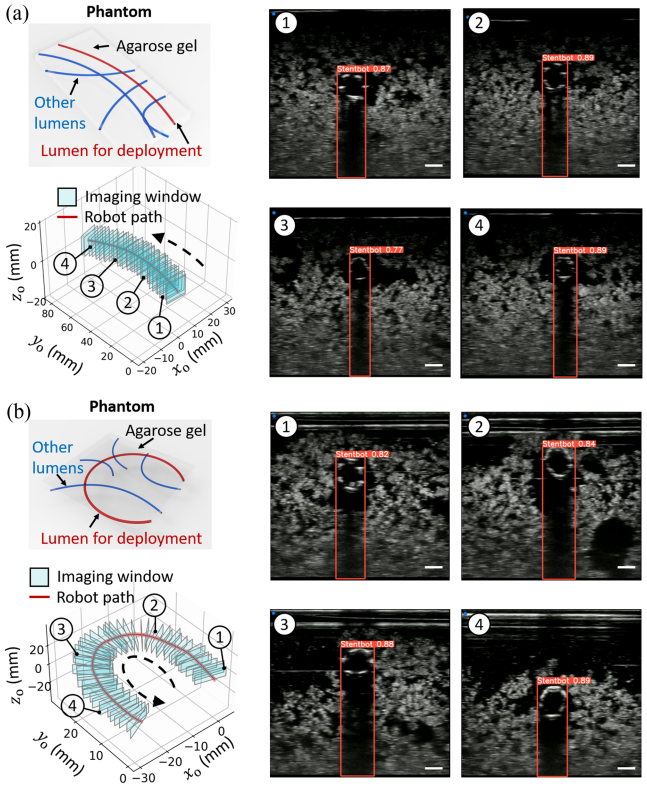


Fig. 9. Real-time robotic navigation in tortuous lumens. (a) Robotic navigation and tracking inside a phantom with interlaced lumens. The red rectangles denote the localization bounding box with robot type and detection confidence above it. (b) Robotic navigation and tracking inside a lumen with a large curvature. In all figures, the imaging depth is 10 mm, and scale bars represent 1 mm.

low-contrast imaging environments, the system effectively localized the robot.

In Fig. 9(a), real-time control and tracking of the robot were successfully achieved within a phantom of agarose gels featuring interlaced lumens. Upon receiving the user motion command operated by the joystick, which dictated the robot movement, the magnet relocated to the actuation position calculated by (7), rotating at 1 Hz. During magnetic actuation, the robot kept moving until it was outside the imaging plane, and then the tracking process was initiated. Algorithm 1 facilitated automatic ultrasound probe scanning of the lumen to reestablish robot tracking. Ultrasound scanning ceased once the robot was detected, prompting magnet relocation and restart of rotational actuation. This iterative process continued until the robot reached its destination. Furthermore, the framework's efficacy was validated in a highly curved lumen [Fig. 9(b)], characterized by noisy and unstructured imaging conditions. Experimental results indicated a robot average speed of 0.7 mm/s, a parameter that can be potentially enhanced from various aspects, such as increasing command publish frequency and upgrading computational hardware. These findings underscore the framework's robustness in navigating complex lumen geometries under challenging imaging scenarios.

V. Conclusion

Tracking various forms of soft medical millirobots in noisy, low-contrast B-mode ultrasound imaging scenes over long distances during actuation presents challenges for robotic navigation within biological environments. To tackle this challenge, we have developed a framework for real-time magnetic navigation of wireless miniature robots. Leveraging synthetic data for deep neural network training, we have achieved effective localization of several robots in ex-vivo tissues, with 85% less human-labeled data from synthetic materials. This approach addresses critical issues such as data scarcity, imbalance, and the labor-intensive nature of manual labeling, enabling state-of-the-art object detection neural networks to be trained with minimal data collection effort. Integrating our robot detection algorithm with a robotic navigation system, we have successfully controlled soft magnetic millirobots to traverse complex lumens within tissue-mimicking phantoms, where an ultrasound imaging probe, mounted on a 7-DOF robotic arm, automatically tracks the robot to maintain its position within the imaging plane. This work provides a promising solution to real-time navigation of wireless miniature robots under medical imaging feedback, underscores the potential of miniature robots for safe and effective medical applications in complex environments, enhancing their utility and reliability in clinical settings.

Compared with existing ultrasound-based robot localization methods, such as Doppler signal-based [18], [19], filter-based [43], learning-based [44], external signal source-based [45], and motion-dependent approaches [46], we propose a synthetic data-assisted framework that trains neural networks to detect wireless robots by leveraging shadow features in B-mode imaging. While filter-based methods struggle in complex environments and motion-dependent methods are limited in scenarios where device movement is restricted (long-term device implantation disables movement and affects bubble stability [19]), our approach provides a solution to localize both static and dynamic robots in complex imaging scenes. In addition, ultrasound signal emission from catheters limits its use to tethered systems rather than wireless miniature robots [45]. Furthermore, unlike conventional learning-based methods that require extensive labeled real data, our framework reduces data collection burden and enhances model generalization, outperforming standard data augmentation techniques.

Several aspects are worth further investigation in future works for applications in real clinical scenarios. First, investigating robot tracking and navigation within dynamic living body environments, such as those influenced by blood flow and respiratory movements, is crucial, which could involve integrating motion-aware control [47], [48] and adaptive compliance strategies [49], [50]. Second, the safe interaction between the ultrasound probe and tissues on a three-dimensional surface needs consideration. Strategies may include integrating force control into ultrasound scanning with external force sensors [51], [52], [53], [54] or intrinsic torque sensors of the robotic arm [55], and adopting rigid-flexible coupled integrated structures to achieve passive safety [56]. Third, to enhance micrometer-scale localization

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

accuracy, various strategies can be employed, including utilizing multifeature fusion (shadow artifacts and Doppler signals), high-resolution ultrasound probes like Fujifilm UHF71x (up to 71 MHz), and advanced image processing techniques such as super-resolution [57] and denoising neural networks [58]. In addition, multirobot navigation under ultrasound imaging can be further developed to enhance the efficacy of biomedical functions. Lastly, rigorous evaluation under real biologically relevant conditions *in vivo* is essential to validate system performance and safety in clinical settings.

REFERENCES

[1] M. Sitti, *Mobile Microrobotics*. Cambridge, MA, USA: MIT Press, 2017.

[2] Y. Wu, X. Dong, J.-k. Kim, C. Wang, and M. Sitti, "Wireless soft millirobots for climbing three-dimensional surfaces in confined spaces," *Sci. Adv.*, vol. 8, no. 21, 2022, Art. no. eabn3431.

[3] T. Wanget al., "Adaptive wireless millirobotic locomotion into distal vasculature," *Nature Commun.*, vol. 13, no. 1, 2022, Art. no. 4465.

[4] C. Wang, T. Wang, M. Li, R. Zhang, H. Ugurlu, and M. Sitti, "Heterogeneous multiple soft millirobots in three-dimensional lumens," *Sci. Adv.*, vol. 10, no. 45, 2024, Art. no. eadq1951.

[5] C. Wang, Y. Wu, X. Dong, M. Armacki, and M. Sitti, "In situ sensing physiological properties of biological tissues using wireless miniature soft robots," *Sci. Adv.*, vol. 9, no. 23, 2023, Art. no. eadg3988.

[6] T. Wang, Y. Wu, E. Yildiz, S. Kanyas, and M. Sitti, "Clinical translation of wireless soft robotic medical devices," *Nature Rev. Bioeng.*, vol. 2, pp. 470–485, 2024.

[7] M. Sitti, "Miniature soft robots—road to the clinic," *Nature Rev. Mater.*, vol. 3, no. 6, pp. 74–75, 2018.

[8] S. Zhonget al., "Double-modal locomotion of a hydrogel ultra-soft magnetic miniature robot with switchable forms," *Cyborg Bionic Syst.*, vol. 5, 2024, Art. no. 0077.

[9] S. Zhou, Y. Li, Q. Wang, and Z. Lyu, "Integrated actuation and sensing: Toward intelligent soft robots," *Cyborg Bionic Syst.*, vol. 5, 2024, Art. no. 0105.

[10] Z. Li, D. Weng, L. Chen, Y. Ma, Z. Wang, and J. Wang, "Enhanced digital light processing-based one-step 3-dimensional printing of multifunctional magnetic soft robot," *Cyborg Bionic Syst.*, vol. 6, 2025, Art. no. 0215.

[11] A. Aghakhani, A. Pena-Francesch, U. Bozuyuk, H. Cetin, P. Wrede, and M. Sitti, "High shear rate propulsion of acoustic microrobots in complex biological fluids," *Sci. Adv.*, vol. 8, no. 10, 2022, Art. no. eabm5126.

[12] H. Daguerreet al., "A localization method for untethered small-scale robots using electrical impedance tomography," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 5, pp. 3506–3516, Oct. 2022.

[13] D. Son, S. Yim, and M. Sitti, "A 5-D localization method for a magnetically manipulated untethered robot using a 2-D array of Hall-effect sensors," *IEEE/ASME Trans. Mechatron.*, vol. 21, no. 2, pp. 708–716, Apr. 2016.

[14] D. Mery, "Inspection of complex objects using multiple-X-ray views," *IEEE/ASME Trans. Mechatron.*, vol. 20, no. 1, pp. 338–347, Feb. 2015.

[15] Q. Huang, J. Zhou, and Z. Li, "Review of robot-assisted medical ultrasound imaging systems: Technology and clinical applications," *Neurocomputing*, vol. 559, 2023, Art. no. 126790.

[16] M. E. Tiryaki and M. Sitti, "Magnetic resonance imaging-based tracking and navigation of submillimeter-scale wireless magnetic robots," *Adv. Intell. Syst.*, vol. 4, no. 4, 2022, Art. no. 2100178.

[17] Y. Bi, Z. Jiang, F. Duelmer, D. Huang, and N. Navab, "Machine learning in robotic ultrasound imaging: Challenges and perspectives," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 7, pp. 335–357, 2024.

[18] Q. Wang, X. Du, D. Jin, and L. Zhang, "Real-time ultrasound doppler tracking and autonomous navigation of a miniature Helical robot for accelerating thrombolysis in dynamic blood flow," *ACS nano*, vol. 16, no. 1, pp. 604–616, 2022.

[19] L. De Arcangelis, E. Lippiello, M. Pica Ciamarra, and A. Sarracino, "Induced and endogenous acoustic oscillations in granular faults," *Philos. Trans. Roy. Soc. A*, vol. 377, no. 2136, 2019, Art. no. 2100178.

[20] M. R. Bennett and M. O'Sullivan, "Mechanisms of angioplasty and stent restenosis: Implications for design of rational therapy," *Pharmacol. Therapeutics*, vol. 91, no. 2, pp. 149–166, 2001.

[21] J. R. Stevens, A. Zamani, J. I. A. Osborne, R. Zamani, and M. Akrami, "Critical evaluation of stents in coronary angioplasty: A systematic review," *Biomed. Eng. Online*, vol. 20, no. 1, 2021, Art. no. 46.

[22] Z. Yang, L. Yang, M. Zhang, C. Zhang, S. C. H. Yu, and L. Zhang, "Ultrasound-guided catheterization using a driller-tipped guidewire with combined magnetic navigation and drilling motion," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 5, pp. 2829–2840, Oct. 2022.

[23] J. Xiong, C. Xu, K. Ibrahim, H. Deng, and Z. Xia, "A mechanism-image fusion approach to calibration of an ultrasound-guided dual-arm robotic brachytherapy system," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 6, pp. 3211–3220, Dec. 2021.

[24] S. Pane, G. Faoro, E. Sinibaldi, V. Iacovacci, and A. Menciassi, "Ultrasound acoustic phase analysis enables robotic visual-servoing of magnetic microrobots," *IEEE Trans. Robot.*, vol. 38, no. 3, pp. 1571–1582, Jun. 2022.

[25] Z. Jiang, S. E. Salcudean, and N. Navab, "Robotic ultrasound imaging: State-of-the-art and future perspectives," *Med. Image Anal.*, vol. 89, 2023, Art. no. 102878.

[26] T. Wang, W. Hu, Z. Ren, and M. Sitti, "Ultrasound-guided wireless tubular robotic anchoring system," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 4859–4866, Jul. 2020.

[27] M. E. Tiryaki, S. O. Demir, and M. Sitti, "Deep learning-based 3 D magnetic microrobot tracking using 2 D MR images," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6982–6989, Jul. 2022.

[28] S. Zhuanget al., "Microinjection in biomedical applications: An effortless autonomous omnidirectional microinjection system," *IEEE Ind. Electron. Mag.*, vol. 18, no. 4, pp. 55–68, Dec. 2024.

[29] C. Wang, M. Tong, L. Zhao, X. Yu, S. Zhuang, and H. Gao, "DanioSense: Automated high-throughput quantification of zebrafish larvae group movement," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 2, pp. 1058–1069, Apr. 2022.

[30] S. Zhuanget al., "The physics of micro-and nano-manipulation: Fundamentals and applications," *Phys. Rep.*, vol. 1090, pp. 1–70, 2024.

[31] L. Zhao, K. Gatsis, and A. Papachristodoulou, "Stable and safe reinforcement learning via a barrier-Lyapunov actor-critic approach," in *Proc. IEEE 62nd Conf. Decis. Control*, IEEE, 2023, pp. 1320–1325.

[32] G. Zhanget al., "Visual-based contact detection for automated zebrafish larva heart microinjection," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 4, pp. 1803–1813, Oct. 2021.

[33] A. Paproki, O. Salvado, and C. Fookes, "Synthetic data for deep learning in computer vision & medical imaging: A means to reduce data bias," *ACM Comput. Surv.*, vol. 56, pp. 1–37, 2024.

[34] J. Tremblayet al., "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 969–977.

[35] P. Kaur, B. S. Khehra, and E. B. S. Mavi, "Data augmentation for object detection: A review," in *Proc. 2021 IEEE Int. Midwest Symp. Circuits Syst.*, IEEE, 2021, pp. 537–543.

[36] G. Jocher, "YOLOv5 by ultralytics," May 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[37] Z. Ren, W. Hu, X. Dong, and M. Sitti, "Multi-functional soft-bodied jellyfish-like swimming," *Nature Commun.*, vol. 10, no. 1, 2019, Art. no. 2703.

[38] W. Hu, G. Z. Lum, M. Mastrangeli, and M. Sitti, "Small-scale soft-bodied robot with multimodal locomotion," *Nature*, vol. 554, no. 7690, pp. 81–85, 2018.

[39] F. W. Kremkau and K. Taylor, "Artifacts in ultrasound imaging," *J. Ultrasound Med.*, vol. 5, no. 4, pp. 227–237, 1986.

[40] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Trans. Syst. Man. Cybern.*, vol. 4, no. 2, pp. 100–107, Jul. 1968.

[41] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[42] T.-Y. Linet al., "Microsoft coco: Common objects in context," in *Proc. Comput. Vis.–ECCV 13th Eur. Conf.*, Springer, 2014, pp. 740–755.

[43] M. Kaya, E. Senel, A. Ahmad, O. Orhan, and O. Bebek, "Real-time needle tip localization in 2 d ultrasound images for robotic biopsies," in *Proc. 2015 Int. Conf. Adv. Robot.*, IEEE, 2015, pp. 47–52.

[44] Y. Wang, L. Chen, H. Han, D. Zhang, H. Liao, and F. Chen, "Localization and angle estimation of capsule robots in ultrasound images using spatially adaptive Gaussian distribution," in *Proc. 2024 IEEE Int. Conf. Bioinform. Biomed.*, IEEE, 2024, pp. 2613–2620.

[45] F. Juet al., "A flexible surgical robot with hemispherical magnet array steering and embedded piezoelectric beacon for ultrasonic position sensing," *Adv. Intell. Syst.*, vol. 6, no. 3, 2024, Art. no. 2300641.

[46] Z. Jianget al., "Needle segmentation using GAN: Restoring thin instrument visibility in robotic ultrasound," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 4012211.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[47] X. Yanet al., "A unified interaction control framework for safe robotic ultrasound scanning with human-intention-aware compliance," in *Proc. 2024 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, IEEE, 2024, pp. 14004–14011.
[48] Z. Jianget al., "Motion-aware robotic 3 d ultrasound," in *Proc. 2021 IEEE Int. Conf. Robot. Autom.*, IEEE, 2021, pp. 12494–12500.
[49] J. Jiang, J. Luo, H. Wang, X. Tang, F. Nian, and L. Qi, "Force tracking control method for robotic ultrasound scanning system under soft uncertain environment," *Actuators*, vol. 13, no. 2, 2024, Art. no. 62.
[50] X. Yan, Y. Jiang, G. Wu, C. Chen, G. Huang, and X. Li, "Multi-modal interaction control of ultrasound scanning robots with safe human guidance and contact recovery," 2023, *arXiv:2302.05685*.
[51] X. Bao, S. Wang, L. Zheng, R. J. Housden, J. V. Hajnal, and K. Rhode, "A novel ultrasound robot with force/torque measurement and control for safe and efficient scanning," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 4002012.
[52] T.-Y. Fang, H. K. Zhang, R. Finocchi, R. H. Taylor, and E. M. Boctor, "Force-assisted ultrasound imaging system through dual force sensing and admittance robot control," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, pp. 983–991, 2017.
[53] H. Huanget al., "Robust passivity-based dynamical systems for compliant motion adaptation," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 6, pp. 4819–4828, Dec. 2022.
[54] N. Marahrens, D. Jones, N. Murasovs, C. S. Biyani, and P. Valdastri, "An ultrasound-guided system for autonomous marking of tumor boundaries during robot-assisted surgery," *IEEE Trans. Med. Robot. Bionics.*, vol. 6, no. 4, pp. 1699–1712, Nov. 2024.
[55] M. Iskandar, A. Albu-Schäffer, and A. Dietrich, "Intrinsic sense of touch for intuitive physical human-robot interaction," *Sci. Robot.*, vol. 9, no. 93, 2024, Art. no. eadn4008.
[56] Y. Xieet al., "Robot-assisted trans-esophageal ultrasound and the virtual admittance-based master-slave control method thereof," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 5, pp. 2505–2516, Oct. 2023.
[57] D. Qiu, Y. Cheng, and X. Wang, "Medical image super-resolution reconstruction algorithms based on deep learning: A survey," *Comput. Methods Programs Biomed.*, vol. 238, 2023, Art. no. 107590.
[58] A. Kazerouniet al., "Diffusion models in medical imaging: A comprehensive survey," *Med. Image Anal.*, vol. 88, 2023, Art. no. 102846.



Chunxiang Wang received the B.E. and M.S. degrees in automation from the Harbin Institute of Technology, Harbin, China, in 2019 and 2021, respectively.

Since 2021, he has been a Ph.D. student with the Department of Physical Intelligence, Max Planck Institute for Intelligent Systems, Stuttgart, Germany, and the Department of Information Technology and Electrical Engineering, ETH Zürich, Zürich, Switzerland. His research interests include soft robotics, image processing, and robotic control.



Tianlu Wang (Member, IEEE) received the B.Eng. degree in mechanical engineering from the Harbin Institute of Technology, Harbin, China, in 2015, and the M.Sc. degree in robotics, systems and control and the Dr.Sc. degree in information technology and electrical engineering from ETH Zürich, Zürich, Switzerland, in 2018 and 2023, respectively.

From 2018 to 2023, he was a Ph.D. student and Postdoctoral Researcher with the Department of Physical Intelligence„ Max Planck Institute for Intelligent Systems, Stuttgart, Germany. In 2024, he joined the Department of Mechanical Engineering, University of Hawaii at Mānoa, Honolulu, HI, USA, as an Assistant Professor. His research interests include miniature soft robotics and its applications in real-world underwater systems, bioinspired robots, and medical devices.



Metin Sitti (Fellow, IEEE) received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Bogaziçi University, Istanbul, Turkey, in 1992 and 1994, respectively, and the Ph.D. degree in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1999.

From 1999 to 2002, he was a Research Scientist with the University of California at Berkeley, Berkeley, CA,USA, and from 2002 to 2016, a Professor of mechanical engineering with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. From 2014 to 2023, he was the Director of the Department of Physical Intelligence, Max Planck Institute for Intelligent Systems, Stuttgart, Germany, and from 2020 to 2024, a Professor with the Department of Information Technology and Electrical Engineering, ETH Zürich, Zürich, Switzerland. Since 2023, he has been the President of Koç University, Istanbul, Turkey. His research interests include small-scale mobile robotics, bioinspiration, wireless medical robots/devices, and physical intelligence.