VesNet-RL: Simulation-Based Reinforcement Learning for Real-World US Probe Navigation

Yuan Bi, Zhongliang Jiang, *Graduate Student Member, IEEE*, Yuan Gao, Thomas Wendler, Angelos Karlas, and Nassir Navab, *Fellow, IEEE*

***Abstract*—Ultrasound (US) is one of the most common medical imaging modalities since it is radiation-free, low-cost, and real-time. In freehand US examinations, sonographers often navigate a US probe to visualize standard examination planes with rich diagnostic information. However, reproducibility and stability of the resulting images often suffer from intra- and inter-operator variation. Reinforcement learning (RL), as an interaction-based learning method, has demonstrated its effectiveness in visual navigating tasks; however, RL is limited in terms of generalization. To address this challenge, we propose a simulation-based RL framework for real-world navigation of US probes towards the standard longitudinal views of vessels. A UNet is used to provide binary masks from US images; thereby, the RL agent trained on simulated binary vessel images can be applied in real scenarios without further training. To accurately characterize actual states, a multi-modality state representation structure is introduced to facilitate the understanding of environments. Moreover, considering the characteristics of vessels, a novel standard view recognition approach based on the minimum bounding rectangle is proposed to terminate the searching process. To evaluate the effectiveness of the proposed method, the trained policy is validated virtually on 3D volumes of a volunteer's in-vivo carotid artery, and physically on custom-designed gel phantoms using robotic US. The results demonstrate that proposed approach can effectively and accurately navigate the probe towards the longitudinal view of vessels.**

***Index Terms*—Robotic ultrasound, reinforcement learning, medical robotics, standard plane identification.**

Manuscript received February 24, 2022; accepted April 18, 2022. Date of publication May 23, 2022; date of current version May 27, 2022. This letter was recommended for publication by Associate Editor A. Schlaefer and Editor J. Burgner-Kahrs upon evaluation of the reviewers' comments. *(Yuan Bi and Zhongliang Jiang contributed equally to this work.) (Corresponding authors: Zhongliang Jiang; Yuan Bi.)*

Yuan Bi, Zhongliang Jiang, and Thomas Wendler are with the Chair for Computer-Aided Medical Procedures and Augmented Reality, Technical University of Munich, 85748 Garching bei München, Germany (e-mail: yuan.bi@tum.de; zl.jiang@tum.de; wendler@tum.de).

Yuan Gao is with the Chair for Computer-Aided Medical Procedures and Augmented Reality, Technical University of Munich, 85748 Garching bei München, Germany, and also with the Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: gao.yuan@sjtu.edu.cn).

Angelos Karlas is with the Institute of Biological and Medical Imaging, Helmholtz Zentrum München, 85764 München, Germany, and also with the Department for Vascular and Endovascular Surgery, Rechts Der Isar University Hospital, Technical University of München, 81675 München, Germany (e-mail: angelos.karlas@tum.de).

Nassir Navab is with the Chair for Computer-Aided Medical Procedures and Augmented Reality, Technical University of Munich, 85748 Garching bei München, Germany, and also with the Laboratory for Computer-Aided Medical Procedures, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: navab@cs.tum.edu).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2022.3176112, provided by the authors.

Digital Object Identifier 10.1109/LRA.2022.3176112

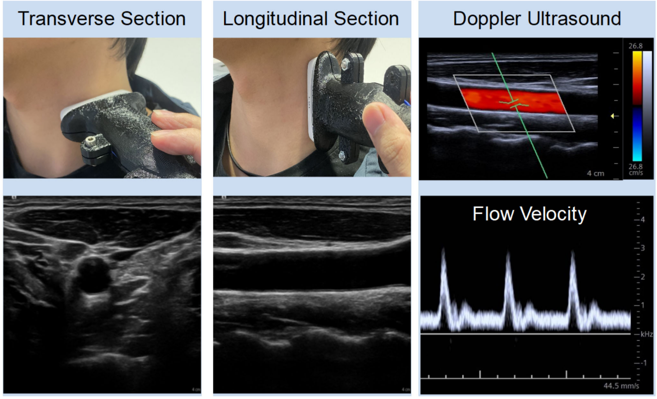

Fig. 1. Illustration of standard planes navigation task, from transverse section to longitudinal section on a representative carotid artery where the flow velocity of the blood can be measured by doppler imaging.

I. Introduction

In THE field of medical imaging, ultrasound (US) is one of the most popular diagnostic tools for medical examinations of internal organs. Compared to computed tomography (CT) and magnetic resonance imaging (MRI) examinations, US is real-time, low cost and radiation free [1]. For vascular medicine, in particular, US plays a critical role in everyday practice, namely, for the diagnostics, image-guided interventions and therapy assessment of diseases. In carotid ultrasonography, the optimal acquisition of the longitudinal view of the carotid artery (see Fig. 1) is required for evaluation of the intima-medial thickness (IMT) [2], the plaque morphology [3], or the peak systolic velocity of the blood over plaques [4]. As for real-time US-guided femoral arterial access, longitudinal views of the target vessel provide a clear visualization of the needle path and the real-time guidance of the guidewire in the vessel of interest [5].

Such planes are often defined as standard planes in US examinations. To properly display the standard planes, sonographers often need to be trained for a few years to gain the necessary anatomical and clinical knowledge. However, since the quality of US imaging highly depends on the level of the operator's experience, the conventional freehand US often suffers from low reproducibility (both intra- and inter-operator) [6]. Furthermore, force-induced deformation also degrades the imaging quality by [7].

A. Robotic US

Due to the superior performance in accuracy, and repeatability, robotic technologies have been employed to develop a robotic US system (RUSS) to overcome the limitation of operator-variation and further improve the clinical acceptance of US modality. Since US imaging quality is highly related to the contact force, Pierrot *et al.* employed a 6-DoF robotic manipulator with a compliant controller to maintain a constant force between the patient skin and the US probe [8]. Besides, Hennersperger *et al.* proposed a workflow to realize autonomous US scans based on imaging registrations [9]. To optimize probe orientation, Jiang *et al.* proposed a method to estimate the normal direction of the contact surface based on the force measured at the tip of US probe [10], [11]. Yet, the aforementioned work is not aimed at determining an optimal view based on the live feed of the US.

Benefited from the development of machine learning, some learning-based approaches have been introduced to address complex recognition and exploring tasks for surgical robotics [12], [13] or autonomous driving [14]. Specific to the task of US standard views recognition, Baumgartner *et al.* proposed SonoNet to assist clinicians in identifying the fetal standard planes in real-time during mid-pregnancy US examinations [15]. In order to provide guidance to sonographers for standard planes navigation, Droste *et al.* used an imitation learning-based system to predict the next action and the final position of the standard view [16]. However, due to the nature of imitation learning, the demonstrations cannot include all the state space. Hence, it is necessary to allow the agent interact with the environment and update its learned policy based on the feedback [17]. Reinforcement learning (RL), on the other hand, provides a unique and alternative solution, since the foundation of RL is based on interaction with the environment.

B. Reinforcement Learning for RUSS

RL has been proved to be reliable to solve complex decision making and exploration problems and has achieved human-level performance in various scenarios, including virtual environments like Atari games [18], real-life applications such as robotic grasping, and indoor navigation tasks [19], [20]. To exploit the potential of RL in the medical field, Alansary *et al.* implemented a deep Q-learning based RL framework to locate the standard planes in brain and cardiac MRI volumes [21]. Regarding RL applications on RUSS, Hase *et al.* trained a DQL agent based on the US images recorded from volunteers' spines to guide a US probe to visualize sacrum. The effectiveness of their proposed method was demonstrated in a virtual environment using unseen data [22]. Nonetheless, only 2-DoFs translational movements were considered, implying that a good orientation initialization is required for its success. Li *et al.* then took a step forward by proposing a DQL framework that accounts for all 6-DoFs while constraining the movements of the agent to the patient surface [23]. Similar to [22], they trained and tested their agent in a virtual environment built by real spine US images. The relatively unsatisfactory performance on the unseen dataset limits its applicability in real-world scenarios. Unlike [22], [23], who used US images as state representations to navigate the US probe, Guo *et al.* attempted to infer the information of US images and force from scene images, and used the scene as state representation to train an RL agent. They used proximal policy optimization algorithms [24] and demonstrate the performance on both phantom and humans.

Due to the nature of RL, it requires a large number of training episodes, which hinders the possibility to train an RL agent directly with a robot in a real-world scenario [25]. Data collection is a difficult task and the generalizability of the trained agent is hampered by the limited and biased training data set. Thus, bridging the simulation-reality gap remains a challenge for the community.

C. Proposed Solution

To address these issues, we proposed an RL-based framework, namely VesNet-RL, to perform US standard plane searching for vascular anatomies. We present a method with high generalization ability by first applying a UNet to segment US images and then running RL on the segmentation results. To eliminate the ambiguity caused by the symmetry of vessel, multi-modality information is involved to create a comprehensive state representation. Since the longitudinal view of the vessel appears as a rectangle across the whole US frame, this view can be easily identified by using the minimum area rectangle of the vessel area in US image. The following are the paper's main contributions:

- An advantage actor critic (A2C) deep RL agent is trained based on the real-time observations to navigate a US probe to the longitudinal view of a vessel. Considering the whole procedure can be interpreted as a partially observable Markov decision process (POMDP), a long short-term memory (LSTM) cell is implemented to exploit the useful information from sequential data.
- In order to make the trained model transferable from a simulation environment to real scenarios and even to other similar vascular applications, we used a UNet to segment the vessel area from the US images before using it as a state representation. To create a comprehensive representation of the probe state, a multi-modality state representation is proposed, including a sequence of consecutive segmented US images, the action history, and sequential changes of the segmented area.
- A novel standard plane recognition method is introduced based on the minimum area bounding rectangle of the segmented area to estimate the real-time vascular diameter and identify the longitudinal view of vessels.

Finally, the proposed VesNet-RL is validated both virtually on a volunteer's carotid and physically on a phantom.[1,2]

II. Method

The proposed VesNet-RL (see Fig. 2) is based on the standard structure of actor-critic RL agents. The actor-network generates the action based on the observation of the current timestamp,

[1] The code: https://github.com/yuan-12138/VesNet-RL

[2] The video: https://www.youtube.com/watch?v=bzCO07Hquj8

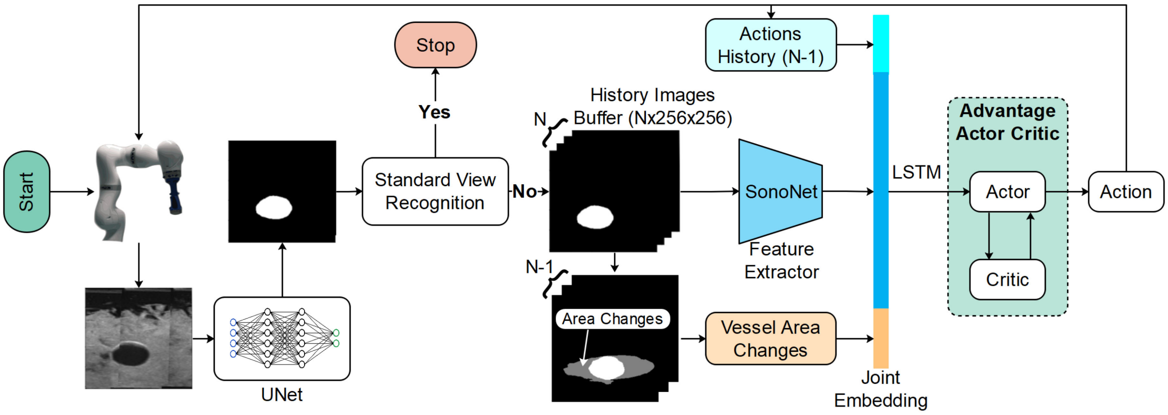

Fig. 2. Schematic overview of VesNet-RL.

whereas the critic-network estimates the preference of the current state. Based on the feedback reward, the network is updated using policy gradient methods. To improve the generalization of the model, the US images are first segmented using a UNet so that the irrelevant background information is erased and the network can focus on the meaningful elements, i.e., vessels. Thereby, the training can be done on the simulated binary images (see Fig. 5(a)), where real B-mode images are not required. This speeds up the training process and expands the diversity of the training set. An RL agent's performance is determined by its ability to accurately estimate its relative position to the goal based on current observations. For such purpose, the features extracted from the history images are concatenated with the previous actions and sequential changes of the segmented vessel area to create a comprehensive representation of the state. Area changes of the vessel denote the size differences of the segmented vessels in US frames between each step. In addition, an LSTM cell is applied to extract potential crucial information from historical data. The searching process will be stopped when the minimum area bounding rectangle condition is triggered in the standard view recognition module.

A. RL-Based US Standard Plane Acquisition

A Markov decision process (MDP) is a standard RL architecture that includes a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, a transition dynamics $\mathcal{T}(s_{t+1}|s_t, a_t)$, a reward function $\mathcal{R}(s_t)$, and a discount factor $\gamma \in [0, 1]$. The actual state of the US probe is not directly observable in the US standard plane acquisition task, resulting in a partially observable MDP (POMDP).

1) Action Space: In our system, all the translational and rotational movements are performed in the end-effector coordinate frame (CF) allowing the trained agent to be used in a variety of standard plane acquisition setups. In comparison, the performance of an RL agent using an action space in base CF is dependent on the initial layout of the target object. Translational movements along the x- and y-axis of the end-effector CF with 5 mm step size and the rotational movements around the z-axis of the end-effector CF with 10° step size make up our action space, which is associated with three DoFs of the probe. The probe's movements are eventually restricted to a plane (object surface), defined as an operation surface (OS), where the searching task takes place.

2) State and Observations: The actual state of our agent is defined as the relative position between the probe and the target. Because the real state cannot be directly measured, observations such as US images together with actions history and segmented area changes are used as observations (o_t) to estimate the actual state.

3) Reward: Since the goal of RL is to train an agent that can execute optimal policy to maximize the expected accumulated rewards, the reward function actually provides guidance to the agent and determines the objective of the learned policy. In our case, the reward should motivate the agent to locate the vessels's largest longitudinal section. The size of the segmented vessel area is also considered in the reward design, rather than just the distance to the goal. The translational distance to the standard view position can be defined as:

$$d_t = \frac{\|(p_t - p_{l_1}) \times (p_t - p_{l_2})\|_2}{\|p_{l_2} - p_{l_1}\|_2} \tag{1}$$

where p_t is the current position of the probe in the base CF, p_{l_1} and p_{l_2} are two points on the projected vessel centerline in the OS. Because the vessel's centerline in our setup can be approximated as a straight line, d_t actually measures the distance between the probe and the projected vessel centerline in the OS.

Afterward, the score of the current state in relation to the distance to the goal is given by:

$$\nu_{dis,t} = 1 - \frac{d_t}{d_{max}} \tag{2}$$

where $\nu_{dis,t} \in [0, 1]$ and d_{max} is the maximum distance between the probe and the projected vessel centerline in the current virtual environment. This is determined by the location of the vessel and the size of the virtual environment.

The score of the current state is also related to the size of the segmented vessel area in the current US frame:

$$\nu_{ves,t} = \frac{D_t}{D_{max}} \tag{3}$$

where D_t is the size of the segmented vessel of the US images (size: 256×256) in pixel, and D_{max} denotes the largest segmented area in this simulation environment, $\nu_{ves,t} \in [0,1]$.

The overall score ν_t of the current state is defined as:

$$\nu_t = \mu_{dis}\nu_{dis,t} + \mu_{ves}\nu_{ves,t} \tag{4}$$

where μ_{dis} and μ_{ves} are the weights of $\nu_{dis,t}$ and $\nu_{ves,t}$, $\mu_{dis} + \mu_{ves} = 1$, and $\nu_t \in [0,1]$. Here the two weights are set to $\mu_{dis} = 0.2$ and $\mu_{ves} = 0.8$. Then, the reward function is given by:

$$r_t = \begin{cases} -0.2, & if\ D_t < D_{th}; \\ 1, & if\ near\ the\ goal \\ 5, & if\ reaching\ the\ goal \\ \nu_t - \nu_{t-1}, & otherwise. \end{cases} \tag{5}$$

where D_{th} is the confirmation threshold for the vessel's existence. If the size of the segmented area is smaller than D_{th}, which, we assume, means that there is no vessel in the US image, the agent will be punished by -0.2. A positive $(+1)$ reward is assigned to the agent if it is near the goal ($\nu_t > 0.9$), while a higher reward $(+5)$ is gained by the agent when it reaches the goal ($\nu_t > 0.95$). Otherwise, the reward is given by the change of the score ($\nu_t - \nu_{t-1} \in [-1,1]$).

4) Advantage Actor Critic: The RL is basically intended to find an optimal policy $\pi^*(a_t|s_t)$, that maximizes future reward at each step [26]. By estimating the policy π with an actor-network, parameterized by θ, the goal can be refactored as maximizing the objective function $J(\theta)$ representing the sum of the reward of the trajectories τ selected by π_θ. The optimisation of the actor network is then done by gradient ascent $\theta \leftarrow \theta + \eta\nabla_\theta J(\theta)$. The gradient of the objective function takes the form:

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau\sim\pi_\theta}[\nabla_\theta \log \pi_\theta(s_t, a_t)\psi_t] \tag{6}$$

where s_t and a_t are state and action in the trajectory τ. ψ_t can have different designs, which distinguishes different policy gradient approaches. For A2C, it is defined by:

$$\psi_t = R_t - V_\omega = \sum_{k=0}^{n-1} \gamma^k r_{t+k+1} + \gamma^n V_\omega(s_{t+n+1}) - V_\omega(s_t) \tag{7}$$

where V_ω is the critic network parameterized by ω, which is a function estimator of the value function. The critic network is updated to minimize the mean square error between the estimated and real values.

As previously mentioned, the entire process in our problem setting is a POMDP. To deal with the uncertainty introduced by POMDP, an LSTM cell is implemented to make full use of the sequential information [27]. The LSTM cell tries to infer the useful information from all previous observations ($o_{0\ldots t}$) and outputs the hidden state (h_t) at the current timestamp as a comprehensive state representation (s_t) for the actor- and critic-network (see Fig. 3).

B. Multi-Modality State Representation

1) State Embedding From Segmentation: The implementation of a UNet to segment the US image and using it as part of the state representation is motivated by the characteristics of RL. The success of RL is based on a large amount of experience, which necessitates not only a large amount, but also a diverse set of data [25]. To implement a trained model into real scenarios, the data ought to be collected from the real scene. In the task of finding the largest longitudinal sections along the vessel, the US data must then be gathered from realistic vascular phantoms, or ideally from patients and volunteers. Taking into account the distinction between phantoms and humans tissue and even individual differences between humans, it is hard to transfer the trained model to similar applications without retraining, which is time-consuming [28]. However, by using a UNet as a preprocessing step, the vascular US images in different applications will have similar geometries, allowing the learned model to be easily transferred to other similar applications. It is sufficient to retrain the UNet rather than the entire RL agent because the training time is much shorter. There are already a plethora of mature segmentation techniques [29], [30].

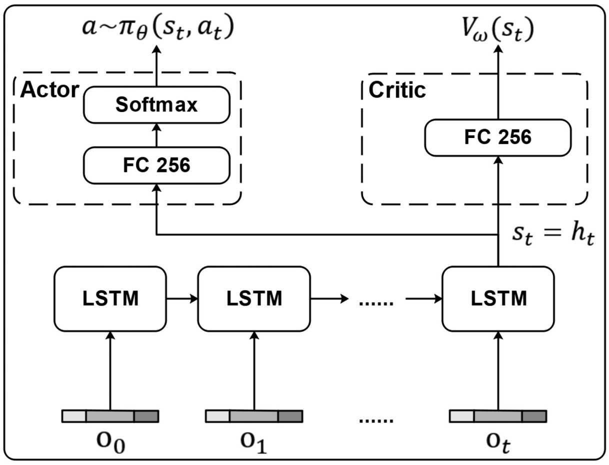

Fig. 3. Illustrations of actor critic architecture with LSTM cell.

Due to the use of raw images, the size of the features extracted from the images is all larger than 256 in [22]–[24]. This creates an ample state space, making it difficult to train an applicable RL agent. Since the RL algorithms are an experience-driven learning procedure, it is evident that using low-dimensional representations is preferable to using high-dimensional ones when both can contain the necessary information [26]. By segmenting the US images, it is possible to represent all information from a history image buffer of size 4 with a feature size of 20, greatly reducing the complexity of the state space while ensuring good network convergence.

2) Multi-Modality State Concatenation: The feature extractor is modified from SonoNet-16 [15] by deleting the last softmax layer. The UNet structure is identical to that of [31]. Inspired by [32], the history of the actions is also involved in the state representation. Asides from the segmented images and actions history, the state representation also includes information about the area changes of the segmented vessel. When the agent is given information about the area difference between each timestamp, the agent can acquire a better understanding of the

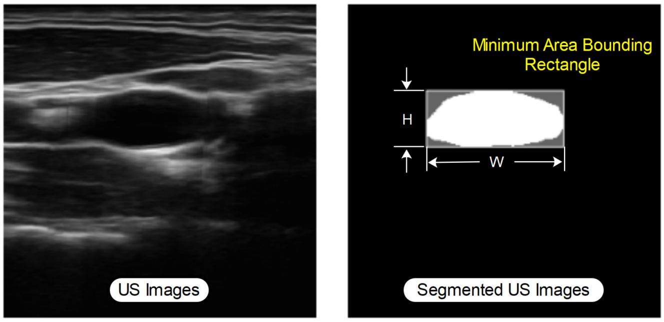

Fig. 4. Illustration of the minimum area bounding rectangle based standard view recognition method.

environment because the reward is also related to the size of the segmented area. The effect of providing this extra information will be further validated in the experiments.

C. Standard View Recognition

To terminate the searching process when the largest longitudinal section along the vessel is found, a standard view recognition method is proposed. Rather than using a network [22], [23], we applied an approach based on the minimum area bounding rectangle of the vessel in the US image. Compared to a network structure, it is more straightforward and time-efficient. The segmented vessel's minimum bounding rectangle is calculated as shown in Fig. 4. In each step t, the diameter of the vessel d_v is estimated by:

$$d_{v,t} = \frac{1}{t}\sum_{k=0}^{t} H_k \tag{8}$$

where H_k is the height of the minimum area rectangle in step k. Then, a termination ratio R_{ter} is calculated to measure the similarity between the segmented area and the rectangle:

$$R_{ter} = \frac{H_t \times W_t - D_t}{H_t \times W_t} \tag{9}$$

where H_t and W_t are the height and width of the minimum area rectangle respectively and D_t is the size of the segmented vessel. $H_t \times W_t - D_t$ basically represents the gray area in Fig. 4. Since the standard view in our use case is approximately a rectangle, the ratio should be as small as possible.

The whole searching process will be terminated when the following conditions are fulfilled:

$$\begin{aligned} &Condition1: && R_{ter} < 0.1 \\ &Condition2: && d_{v,t} - H_t < Th \\ &Condition3: && W_t > \alpha W_i \end{aligned} \tag{10}$$

where Th is a small threshold set to $10\,pixels$ in practice, W_i is the width of the US image in pixel, and α is a discount factor with a value of 0.99. $Condition1$ is used to ensure that the segmented vessel resembles a rectangle as closely as possible. $Condition2$ ensures that the height of the bounding rectangle is roughly equal to the vessel's estimated diameter, implying that the US image plane intersects with the centerline. $Condition3$ makes sure that the bounding box's width is equal to the width of the US image. It is given by the characteristic features of the standard view, where the vessel appears as a rectangle across the entire US image.

III. Experiments and Results

A. Hardware Setup

The proposed US standard plane acquisition system is built by two parts: a robotic arm (KUKA LBR iiwa 7 R800, KUKA Roboter GmbH, Augsburg, Germany) controlled using a Robot Operating System (ROS) interface [9] and two different types of US imaging systems. A first (Cephasonics, California, USA) with a linear US probe (CPLA12875, Cephasonics, California, USA) is used to acquire US images from vascular phantoms. A second (ACUSON Juniper Ultrasound System, Siemens AG, Erlangen, Germany), also equiped with a linear US probe (12L3, Siemens AG, Erlangen, Germany), is applied to take US images of human carotids, since it provides qualitatively better human tissue images. The US probes are mounted to the end-effector of the robot by 3D-printed holders. The US settings are mainly adopted from the build-in files from the manufacturers for vascular imaging. The US images from Cephasonics are accessed by a USB interface provided by the manufacturer, while the B-mode images from the Siemens system were captured by a frame grabber (DVI2USB 3.0, Epiphan Video, Ottawa, Canada). The pose of the robot arm is synchronized with the US images in real-time in a software platform (ImFusion Suite, ImFusion GmbH, Munich, Germany) to reconstruct 3D US volumes of the region of interest and build a 3D virtual environment for the US acquisitions.

To validate the performance of VesNet-RL in different vascular standard plane searching tasks, three custom-made blood vessel phantoms ($Vessel_1$, $Vessel_2$, and $Vessel_3$) were employed. They were made of gelatin powder (175 g/L), paper pulp (3–5 g/L), and liquid disinfectant mixed with water, where the paper pulp is used to mimic the human tissue, and liquid disinfectant is adopted to extend its preservation time. To mimic the structure of vessels, after the solidification of the gel, a round tube was used to create holes in different depth of the phantoms.

B. Training Details

1) UNet Training: The UNet for the vascular phantoms was trained using 4,421 US images acquired from $Vessel_1$ with various poses of the US probe relative to the vessel. The US images, which only display backgrounds, are included in the training dataset to teach the network how to recognize the presence of vessels. If background images are excluded from the training dataset, the performance of the UNet is very unstable when there is no vessel in the images.

The training data for the human carotid UNet consists of 1,041 US images of a volunteer. The acquisition was performed within the Institutional Review Board Approval by the Ethical Commission of the Technical University of Munich (reference number 244/19 S), having the volunteer signed an informed consent. Considering the carotid pulse during the US sweep, the vascular wall exhibits a wave appearance in the longitudinal

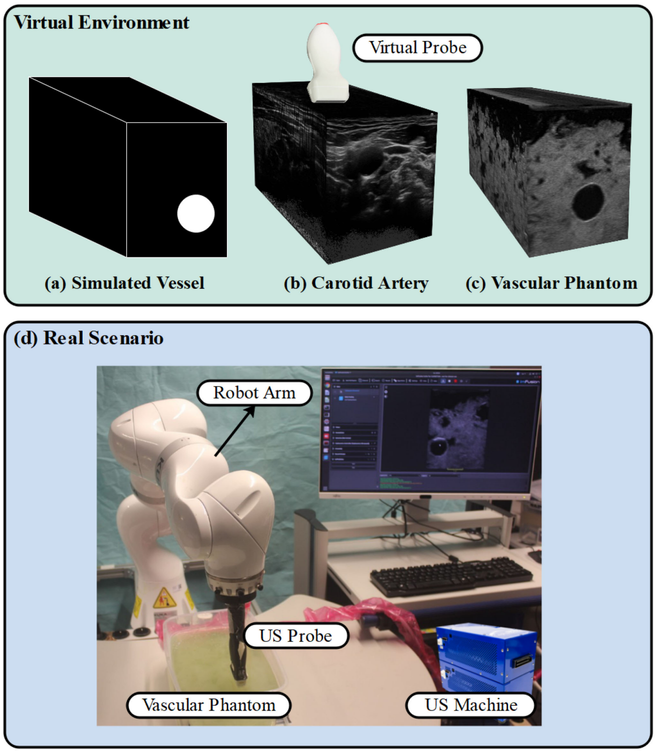

Fig. 5. Different virtual environments, including (a) simulated binary vessel, (b) carotid artery, and (c) vascular phantom, and (d) the experiment setup in real scenario on a vascular phantom.

view of the vessel in the reconstructed US volumes. As a result, these data had to be included in the training set. A sweep along the carotid was performed to reconstruct the US volume of the artery. The probe was attached to the robotic arm, positioned approximately orthogonal to the carotid centerline, and manually moved along the vessel, with the robot only assisting in the pose acquisition. The sweep frames were manually labeled after the carotid reconstruction, and another 3D volume of the same size was built using the labeled images. By taking images in the same position in these two compounding volumes, a pair of training data can be collected. The training set included 1,266 images from a 3D compounding volume. In total, 2,307 images served as the training data for the carotid UNet.

2) RL Agent Training: For training of the RL agent, a virtual environment is built on simulated vessels, with the vessels appearing as a white tube on a black background (see Fig. 5(a)). The depth and size of the vessels are generated at random. Ten binary vessels are created for the training of the RL agent. There is no need to apply a UNet for the training because the simulated vessel images already have the same characteristics as the segmented US frames. The history image buffer size is set to 4. The size of the hidden state of the LSTM cell is set to 256. A total of 3,000 training episodes were executed. At the beginning of each episode, a vascular environment is randomly selected, and the RL agent is randomly initialized, where the vessel is at least partially observable. The network was trained every 20 interaction steps using Adam optimize, with a maximum step size of 500 in each episode. The learning rate for the first 500 episodes is 5×10^{-4}, then drops to 3×10^{-4} for the next 1,000 episodes, and finally declines to 1×10^{-4} for the remaining 1,500 training episodes.

C. Experiments in Virtual Environments

In order to demonstrate the generalization ability of VesNet-RL, the RL agent, which was purely trained on simulated vessels, was tested in virtual environments built by the 3D images of vascular phantoms (see Fig. 5(b) and (c)). In each testing episode, the agent is randomly initialized in each testing environment where the vessel is at least partially observable. The agent is considered successful if it can complete its search in less than 50 steps using the proposed termination criteria. Furthermore, the efficiency of the trained model is evaluated by calculating the average steps of all successful test episodes. The translational error is defined as the distance between the probe tip and the projected centerline of the vessel on OS (equivalent to the upper surface of the US volume). The rotational error is then calculated by the angular difference between the probe and the projected vessel centerline on OS.

1) Evaluation of the Standard View Recognition Module: The ground truth position of the vessel centerline is used to calculate the position and orientation error in the fifth and sixth columns of Table I. To obtain the vessel centerline of the vascular phantom in the corresponding 3D images, a sweep along the vessel in the virtual environment is executed, and the centers of the segmented area are fitted into a linear regression model to form a line [33]. As shown in the Table I, our standard view recognition method is able to terminate the searching process with extremely high precision in position and orientation.

2) Evaluation of the Architecture Design: To demonstrate the efficacy of our framework design, we compared it to various leave-one-out models (ablation study), in which one of the modules of our original design is removed while the rest remains unchanged. The training details for all the different architectures are the same as in Section III-B2, except the one without segmentation. Because of the absence of the UNet, the simulated vessels cannot be used as training data. Instead, three 3D images of the vascular phantom ($Vessel_1$) are used as the training dataset, and the trained model is then tested on $Vessel_{2\&3}$. The test dataset for the other architectures includes all three vascular phantom models.

When the LSTM cell is removed from the original design, the success rate drops dramatically, demonstrating that only considering the previous information in the state representation is insufficient. LSTM exhibits superior performance in revealing the underlying persistence in sequential data. When no segmentation network is employed, the performance of the trained agent shows a weak generalization ability in analogous application environments. When the area changes information is excluded from the state representation, the success rate drops by 20%, and the average number of steps to the goal nearly doubles when compared to VesNet-RL because the area changes information can tell the agent whether it is moving or rotating in the right direction.

TABLE I
PERFORMANCES OF DIFFERENT ARCHITECTURES IN VIRTUAL ENVIRONMENT

Method	Test Environment	Success rate	Average number of steps	Position error (mm)	Orientation error (°)	Number of samples
VesNet-RL	Vascular phantom	92.3%	15	2.06 ± 1.37	4.08 ± 3.20	300
	Carotid	91.5%	13	1.29 ± 0.83	1.32 ± 2.55	400
	Carotid*	56.5%	18	0.92 ± 0.64	2.25 ± 2.75	400
LSTM♠	Vascular phantom	52.0%	25	2.06 ± 1.22	3.51 ± 3.34	300
	Carotid	41.3%	27	1.18 ± 0.83	1.77 ± 2.85	400
Segmentation♠	Vascular phantom ($Vessel_{2\&3}$)	18.7%	34	1.75 ± 1.27	3.33 ± 3.26	200
Area changes♠	Vascular phantom	73.0%	28	1.67 ± 1.19	3.22 ± 3.11	300
	Carotid	64.3%	27	1.38 ± 0.94	2.72 ± 3.15	400
Historical information♠	Vascular phantom	52.3%	32	1.75 ± 1.07	3.26 ± 3.15	300
	Carotid	48.5%	32	1.40 ± 0.87	1.71 ± 2.82	400
VesNet-RL (image buffer size: 8)	Vascular phantom	24.0%	25	1.37 ± 1.21	3.62 ± 3.15	300
	Carotid	10.3%	40	1.62 ± 0.94	1.72 ± 2.83	400

[♠] means the corresponding module is removed from the proposed VesNet-RL framework.

We trained a model that only takes the current observations as state representation to showcase that multiple consecutive images are still required even after the LSTM cell is implemented. By comparing the result to the original model, whose state consists of 4 consecutive images, actions, and area changes, we can conclude that including previous information in the state representation using an LSTM cell still improves the model. Because for symmetrical structures like vessels, the same image can be acquired in the same position; but with different orientations, a sequence of consecutive images along with actions history allows the network to gain a better understanding of the surrounding and eliminate the ambiguity, resulting in more accurate state descriptions and faster training. However, when the history images buffer size is set to 8 and the corresponding images buffer feature size is set to 40, the trained model performs poorly compared to others, showing that expanding the state space can sometimes prevent models from learning a delicate policy.

3) Performance Comparison Between Phantom and In-Vivo Human Data: To test our model in a more realistic scenario, four 3D models of human carotid were built as described in Section III-B1 (see Fig. 5). It is worth noting that the US machine used for human data acquisition differs from the one used for vascular phantoms., Carotid* in Table I indicates that the images from the 3D reconstructed volume are not included in the UNet training set for carotid as described in Section III-B1. If the UNet fails to segment the vessel area properly, then the success rate of the RL model is much lower. On the contrary, when images from the 3D reconstructed volume of the carotid are included in the UNet training data, VesNet-RL achieves a 91.5% success rate in locating the longitudinal section of the carotid. For the other architectures, the success rates drop slightly. Except for the one without segmentation, if there is no retraining, the trained network is not able to be transferred to carotid applications.

D. Experiments in Real Scenarios

To showcase the performance of VesNet-RL in real scenarios, we tested our trained model on the real vascular phantoms with a robot arm (see Fig. 5). The trained model is the same as in Section III-B2. The OS is then defined as the upper surface of the gel phantom, while the actions are identical to Section II-A1. The phantom is immerged into water so there is no need to use US gel. 60 tests were carried out on a custom-made vascular phantom. The robot executed the learned policy with a maximum of 50 steps. At the beginning of each test, the probe is randomly initialized orthogonal to the upper surface of the phantom, where the vessel is at least partially observable. The defination of success rate is identical to that of Section III-C. Table II shows that VesNet-RL has a high success rate (80%) and high accuracy in navigating the US probe to the standard view of a vessel. In the vast majority of cases, the failure was due to incorrect vessel segmentation, which resulted in a misestimation of the actual state.

TABLE II
PERFORMANCE OF VESNET-RL IN REAL SCENARIO

Success rate	Average number of steps	Position error (mm)	Orientation error (°)
80.0%	17	0.79 ± 0.55	2.08 ± 3.05

E. Discussion

Besides the anatomy of interest, the background of US images also contains certain information. For the tasks like searching for specific anatomies, e.g., kidney, the background can also help clinicians quickly locate the anatomy of interest, particularly when the searching process starts from a random position. However, for tasks like locating the standard planes of arteries (e.g., longitudinal view), the displayed view of the objects is more important to accurately navigate the probe. Since the background of B-mode images is sensitive to practical factors like contact force, amount of gel, and orientation, which will hinder the convergence of the trained model and affect the generalizability of the trained model for unseen patients.

IV. CONCLUSION

In this work, we present a simulation-based RL approach for automatically navigating a US probe to a vascular standard plane (i.e., the largest longitudinal view). Segmented binary images are used as part of a multimodality state representation to bridge the gap between the simulation training environment and the real scenario, as well as to address the challenge of low generalization ability Thanks to an explicit segmentation of the US frames, the RL agent, trained with a wide variety of simulated binary vessels, can be used to guide the US probe in actual practice, such as the carotid standard view acquisition. Experiments were conducted in both virtual and real scenarios to demonstrate the efficacy of VesNet-RL. The proposed model was compared with various network structures on 3D models of vascular phantoms and a human carotid in virtual environments. With the highest success rate (92.3% for vascular phantoms and 91.5% for the carotid artery) and the minimum average number of steps (15 for vascular phantoms and 13 for the carotid artery), the proposed framework outperforms the competition. The novel standard view recognition method for vascular use-case also achieves excellent results in the tests for tubular phantom and carotid artery in the virtual environments (2.06 ± 1.37 mm and 1.29 ± 0.83 mm in terms of position error and $4.08 \pm 3.20^\circ$ and $1.32 \pm 2.55^\circ$ in terms of orientation error, respectively). We also demonstrate that the model trained in a simulation environment can be directly applied in the real scenario on a vascular phantom without extra training or retraining, achieving a success rate of 80%, 0.79 ± 0.55 mm position error, and $2.08 \pm 3.05^\circ$ orientation error. In the future, we will further consider the contact force and the deformation of the human tissue [7] to further pave the way to real clinical applications.

References

[1] P. R. Hoskins, K. Martin, and A. Thrush, *Diagnostic Ultrasound: Physics and Equipment*, Boca Raton, FL, USA: CRC Press, 2019.

[2] T. Nezu, N. Hosomi, S. Aoki, and M. Matsumoto, “Carotid intima-media thickness for atherosclerosis,” *J. Atherosclerosis Thromb.*, vol. 23, no. 1, 2015, Art. no. 31989.

[3] J. D. Spence, M. Eliasziw, M. DiCicco, D. G. Hackam, R. Galil, and T. Lohmann, “Carotid plaque area: A tool for targeting and evaluating vascular preventive therapy,” *Stroke*, vol. 33, no. 12, pp. 2916–2922, 2002.

[4] D. Inzitari *et al.*, “The causes and risk of stroke in patients with asymptomatic internal-carotid-artery stenosis,” *New England J. Med.*, vol. 342, no. 23, pp. 1693–1701, 2000.

[5] A. H. Seto *et al.*, “Real-time ultrasound guidance facilitates femoral arterial access and reduces vascular complications: FAUST (femoral arterial access with ultrasound trial),” *JACC: Cardiovasc. Interv.*, vol. 3, no. 7, pp. 751–758, 2010.

[6] S. D. Kanters, A. Algra, M. S. Van Leeuwen, and J. D. Banga, “Reproducibility of in vivo carotid intima-media thickness measurements: A review,” *Stroke*, vol. 28, no. 3, pp. 665–671, Mar. 1997.

[7] Z. Jiang, Y. Zhou, Y. Bi, M. Zhou, T. Wendler, and N. Navab, “Deformation-aware robotic 3D ultrasound,” *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 7675–7682, 2021.

[8] 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.

[9] C. Hennersperger *et al.*, “Towards MRI-based autonomous robotic us acquisitions: A first feasibility study,” *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 538–548, Feb. 2017.

[10] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab, “Automatic force-based probe positioning for precise robotic ultrasound acquisition,” *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11200–11211, Nov. 2021.

[11] Z. Jiang *et al.*, “Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement,” *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1342–1349, Apr. 2020.

[12] B. Lu *et al.*, “A learning-driven framework with spatial optimization for surgical suture thread reconstruction and autonomous grasping under multiple topologies and environmental noises,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 3075–3082.

[13] B. Lu *et al.*, “Toward image-guided automated suture grasping under complex environments: A learning-enabled and optimization-based holistic framework,” *IEEE Trans. Automat. Sci. Eng.*, to be published, doi: 10.1109/TASE.2021.3136185.

[14] T. Zou *et al.*, “KAM-Net: Keypoint-aware and keypoint-matching network for vehicle detection from 2-D point cloud,” *IEEE Trans. Artif. Intell.*, vol. 3, no. 2, pp. 207–217, Apr. 2021.

[15] C. F. Baumgartner *et al.*, “SonoNet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound,” *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2204–2215, Nov. 2017.

[16] R. Droste, L. Drukker, A. T. Papageorghiou, and J. A. Noble, “Automatic probe movement guidance for freehand obstetric ultrasound,” in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2020, pp. 583–592.

[17] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, “Imitation learning: A survey of learning methods,” *ACM Comput. Surv.*, vol. 50, no. 2, pp. 1–35, 2017.

[18] V. Mnih *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[19] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, “Learning synergies between pushing and grasping with self-supervised deep reinforcement learning,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 4238–4245.

[20] Y. Zhu *et al.*, “Target-driven visual navigation in indoor scenes using deep reinforcement learning,” in *Proc. IEEE Int. Conf. Robot. Automat.*, pp. 3357–3364.

[21] A. Alansary *et al.*, “Evaluating reinforcement learning agents for anatomical landmark detection,” *Med. Image Anal.*, vol. 53, pp. 156–164, 2019.

[22] H. Hase *et al.*, “Ultrasound-guided robotic navigation with deep reinforcement learning,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 5534–5541.

[23] K. Li *et al.*, “Autonomous navigation of an ultrasound probe towards standard scan planes with deep reinforcement learning,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 8302–8308.

[24] G. Ning, X. Zhang, and H. Liao, “Autonomic robotic ultrasound imaging system based on reinforcement learning,” *IEEE Trans. Biomed. Eng.*, vol. 68, no. 9, pp. 2787–2797, Sep. 2021.

[25] W. Zhao, J. P. Queralta, and T. Westerlund, “Sim-to-real transfer in deep reinforcement learning for robotics: A survey,” in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2020, pp. 737–744.

[26] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep reinforcement learning: A brief survey,” *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.

[27] V. Mnih *et al.*, “Asynchronous methods for deep reinforcement learning,” *in Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937.

[28] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, “Building machines that learn and think like people,” *Behav. Brain Sci.*, vol. 40, 2017, Art. no. E253.

[29] Y. Weng, T. Zhou, Y. Li, and X. Qiu, “Nas-unet: Neural architecture search for medical image segmentation,” *IEEE Access*, vol. 7, pp. 44247–44257, 2019.

[30] L. Christodoulou, C. P. Loizou, C. Spyrou, T. Kasparis, and M. Pantziaris, “Full-automated system for the segmentation of the common carotid artery in ultrasound images,” in *Proc. 5th Int. Symp. Commun. Control Signal Process.*, 2012, pp. 1–6.

[31] Z. Jiang *et al.*, “Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback,” *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7064–7075, Jul. 2021.

[32] S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Young Choi, “Action-decision networks for visual tracking with deep reinforcement learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2711–2720.

[33] Y. Liu *et al.*, “Globally optimal linear model fitting with unit-norm constraint,” *Int. J. Comput. Vis.*, vol. 130, no. 4, pp. 933–946, 2022.