IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2024, Yokohama, Japan.
Cite as T-ASE paper.

DopUS-Net: Quality-Aware Robotic Ultrasound Imaging based on Doppler Signal

Zhongliang Jiang*, Felix Duelmer*, and Nassir Navab, *Fellow, IEEE*

***Abstract*—Medical ultrasound (US) is widely used to evaluate and stage vascular diseases, in particular for the preliminary screening program, due to the advantage of being radiation-free. However, automatic segmentation of small tubular structures (e.g., the ulnar artery) from cross-sectional US images is still challenging. To address this challenge, this paper proposes the DopUS-Net and a vessel re-identification module that leverage the Doppler effect to enhance the final segmentation result. Firstly, the DopUS-Net combines the Doppler images with B-mode images to increase the segmentation accuracy and robustness of small blood vessels. It incorporates two encoders to exploit the maximum potential of the Doppler signal and recurrent neural network modules to preserve sequential information. Input to the first encoder is a two-channel duplex image representing the combination of the grey-scale Doppler and B-mode images to ensure anatomical spatial correctness. The second encoder operates on the pure Doppler images to provide a region proposal. Secondly, benefiting from the Doppler signal, this work first introduces an online artery re-identification module to qualitatively evaluate the real-time segmentation results and automatically optimize the probe pose for enhanced Doppler images. This quality-aware module enables the closed-loop control of robotic screening to further improve the confidence and robustness of image segmentation. The experimental results demonstrate that the proposed approach with the re-identification process can significantly improve the accuracy and robustness of the segmentation results (dice score: from 0.54 to 0.86; intersection over union: from 0.47 to 0.78). The Code[1] and Video[2] are publicly accessible.**

***Note to Practitioners*—The Doppler signal is important for the diagnosis of vascular disease, e.g., peripheral arterial disease, in clinical practices, nevertheless it is not of similar significance for state-of-the-art robotic ultrasound (US) examination systems yet. This paper explores various neural network structures to effectively extract the blood vessels from US images by incorporating the Doppler signal into the segmentation process. The final DopUS structure with two encoders extracting differentiated information from two different inputs and fusing the latent feature representations in the bottleneck layer can also inspire other tasks like multi-senor fusion. In addition, this work developed a Doppler-based tracker to assess the quality of the segmentation results in real-time. The assessment is subsequently used for a quality-aware module that enables closed-loop control of the robotic screening. Preliminary physical experiments suggest that the quality-aware robotic screening system can improve the confidence and robustness of autonomous US examination results. In the future, the Doppler signal could also be used to support clinical diagnosis. We believe the proposed quality-aware autonomous screening system is important for the development of large-scale robotic US screening programs. It will not only benefit the examination of limb arteries but also other vascular structures, e.g., carotid or aorta.**

***Index Terms*—Robotic ultrasound, vessel segmentation, Ultrasound segmentation, 3D visualization**

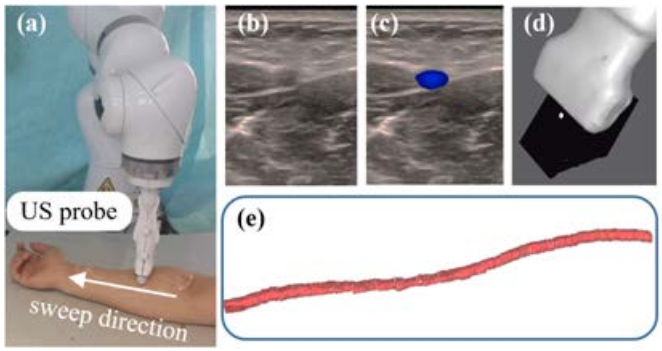

Fig. 1. (a) A scene of robotic US scanning of a volunteer's arm; (b) and (c) are B-mode and corresponding color Doppler images, respectively; (d) is the 3D view of tracked binary segmentation results in real-time; (e) is the reconstructed 3D vessel of interest.

I. Introduction

PERIPHERAL arterial disease (PAD) refers to the pathological process causing obstruction to blood flow in the arteries. It is a chronic disease of the aortic, iliac, and limb arteries (see Fig. 1) [1]. Stenosis is one of the most common PAD, which narrows the blood vessel due to the plaque building up inside arteries; thereby, restricting the supply of blood. This can result in stroke and amputation in the worst cases. PAD affects approximately 20% of the population older than 55 years [2]. However, the patient awareness of PAD diagnosis is low, particularly for the cases with mild symptoms, e.g., numbness, muscle shrinking, and ulcers [2]. Thereby, a regular screening program for PAD will benefit early detection, which can further lead to the improvement of long-term life quality and a decrease in systemic cardiovascular risk.

Preliminary screening for PAD is often performed in primary care practices using the ankle-brachial index (ABI). The ABI examines the blood pressure in the extremities but may result in an underdiagnosis of PAD [2]. Although the ABI method is cheap, it cannot provide the location of the stenosis and is highly user-dependent. Conversely, computed tomography angiography (CTA), or magnetic resonance angiography (MRA) [3] can provide accurate clinical

* Authors with equal contributions.
Z. Jiang, F. Duelmer, and N. Navab are with the Chair for Computer Aided Medical Procedures and Augmented Reality (CAMP), Technical University of Munich (TUM), 85748 Garching, Germany. (zl.jiang@tum.de)

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by Institutional Review Board, No. 2022-87-S-KK, Declaration of Helsinki.

[1]Code: https://github.com/Felixduelmer/DopUs

[2]Video: https://www.youtube.com/watch?v=ZH7K63GngdA

information. However, MRA is costly and CTA has ionizing radiation. Therefore, both of them are not suitable for extensive preliminary screening programs, where a large part of the participants are healthy. Considering the aforementioned constraints, ultrasound (US) imaging has been seen as a promising alternative for examining the extremity artery tree due to the absence of contraindications [4]. Furthermore, US has been extensively utilized for visualizing internal lesions and organ abnormalities due to its ability to provide rapid and straightforward diagnosis [5]. Regarding PAD, US imaging can provide information on the degree of calcification [4]. Nevertheless, nonnegligible inter- and intra-operator variations limit the clinical acceptance of the traditional US examination.

To tackle this shortage, robotic US systems (RUSS) have been seen as a promising solution to provide accurate, stable, and clinician-independent diagnosis results [6]–[10]. Specific to the RUSS developed for screening tubular structures, Janvier *et al.* employed a 6-DoF industrial robotic arm to hold a US probe and evaluated the performance of their proposed RUSS on a lower-limb mimicking phantom [11]. The experimental results demonstrated that the RUSS could be of value for the clinical evaluation of lower limb vessels over long and tortuous segments below the knee by accurately identifying the stenosis section. Compared to the traditional 2D images, 3D images are more intuitive for clinicians to identify lesion locations based on the reconstructed 3D anatomy. To characterize the geometry of the vessel, Merouche *et al.* segmented vessel lumen using a fast marching method based on gradients [12]. However, the efficiency of this approach is limited as the process takes 15 minutes to scan $156\ mm$ of a femoral artery. Since the aforementioned work was validated using a phantom without any blood flow, no color Doppler images are present. Yet, Doppler imaging is widely used in real scenarios for vascular examination and provides valuable information to facilitate both blood vessel detection and clinical analytical diagnosis.

To bridge this gap, we propose a novel RUSS combining Doppler with B-mode images to facilitate the real-time and accurate segmentation of small vessels like the radial artery (mean$\pm$SD diameter: 2.4 ± 0.4 mm [13]). To fully take advantage of the Doppler effect, a deep neural network (DopUS-Net) consisting of two encoders and one decoder is proposed. The Doppler image is first concatenated to the synchronized B-mode image to form a two-channel image to be used as one of the inputs to the DopUS-Net. The other input is the individual Doppler image, which is further used to provide coarse vessel locations and therefore acts as a region proposal module to increase the overall robustness and reliability. The main contributions of this work are summarized as follows:

- We explore the optimal way to take advantage of the Doppler signal beside B-mode images to facilitate the accurate and robust segmentation of small limb arteries. The DopUS-Net with two encoders is proposed to effectively fuse the Doppler and B-mode images. The two encoders with two-channel duplex images (Doppler and B-mode) and pure Doppler images are designed to accurately extract the artery boundary and provide the coarse region proposal, respectively.
- We first present an online segmentation quality-aware module based on the Doppler signal. Such a module enables close monitoring of imaging quality of robotic screening to improve the confidence and robustness of image segmentation, which can effectively improve the accuracy and completeness of the reconstructed 3D vessel.
- We visualize the tubular structure of interest in a 3D view [see Fig. 1 (e)] to facilitate the intuitive assessment of clinicians. Due to the short segmentation time ($9\ ms$), the 3D visualization process can be seen as real-time.

The experiments are performed on seven healthy volunteers and the experimental results demonstrate that the proposed DopUS-Net can outperform existing methods [14], [15] in terms of dice score. The online evaluation of the Doppler signals' stability and quality helps to improve the robustness and quality of the resulting 3D compounding image.

The rest of this paper is organized as follows. Section II presents related work. The dataset preparation and the implementation details of the DopUS-Net are presented in Section III. Section IV describes the details of robotic scanning and the online artery re-identification approach. The experimental results on seven volunteers are provided in Section V. Finally, the summary of this study is presented in Section VI.

II. Related Work

A. Feature-based Ultrasound Vessel Segmentation

Compared to other popular medical imaging modalities, e.g., CT or MRI, US often suffers from inconsistent quality, speckle, and artifacts due to the intrinsic physics of wave propagation, like interference and scattering effects [16]. To optimize the acoustic coupling performance, sonographers need to carefully adjust the pressure and probe orientation to avoid acoustic shadow and improve the imaging contrast. These factors contribute to making US one of the most challenging modalities for robust and accurate segmentation. Regarding the segmentation of blood vessels from cross-sectional US images, the Frangi filter [17] was developed based on the Hessian matrix in the early studies. Such an approach allows real-time segmentation, but presents limited segmentation accuracy.

Since the shape of vascular structures on cross-sectional images are close to the ellipse, Smistad *et al.* proposed a template-based approach to automatically segment the object of interest [18]. Yet, this approach is limited with regard to the segmentation accuracy because the segmented shapes are forced to be ellipse. To further enable accurate detection of vessel boundaries, Karami *et al.* proposed an adaptive polar active contours method to segment the jugular vein [19]. They introduced a set of energy functions to consider local information, which enables robust segmentation of vessels even when the imaging quality is poor. However, this method requires a manual initialization of the object on the first frame. In addition, Abolmaesumi *et al.* compared the performance of five popular feature tracking approaches (the cross-correlation,

the sequential similarity detection, the Star algorithm, the Star-Kalman algorithm, and the discrete snake algorithm) on carotid arteries [20]. The experimental results demonstrated that the Similarity Detection method and the Star-Kalman algorithm can achieve noted tracking performance, while the Correlation and Star algorithms result in poorer performance with higher computational cost.

B. Learning-based Ultrasound Vessel Segmentation

Compared with traditional feature-based approaches, learning-based approaches have been seen as a promising alternative for the accurate segmentation of US images in recent years. Thanks to the inventions related to the convolutional neural network (CNN), learning-based approaches have achieved phenomenal success on various computer vision tasks as well as medical imaging segmentation [21]. Specific to biomedical image segmentation, U-Net and its variants have been widely used [14], [22], [23]. Such U-shape networks are based on fully convolutional neural networks (FCN). Multiple skip connections are used to pass the information from the encoder to the decoder assisting in accurate boundary extraction. To achieve automatic and accurate segmentation of US images, Mishra *et al.* proposed a FCN trained to learn structural boundary definitions to extract vessels from US images [16]. Considering that the biomedical anatomy is continuous, Chen *et al.* employed a recurrent unit to preserve the historical information to assist the real-time segmentation [24]. To improve generality, the salient image features were extracted from each spatial resolution level within the encoder-decoder structure. Yet, it is still challenging to accurately segment small vessels from B-mode images. To tackle this problem, B. Jiang *et al.* used both B-mode and color Doppler images to train the VesNetSCT++ for small vessels, e.g., femoral and tibial artery [15]. To effectively leverage the spatiotemporal context in the image sequences to improve the segmentation of small-scale arteries, VesNetSCT+ incorporates temporal, spatial, and feature-aware contextual embedding from two-channel images consisting of B-mode and Color Doppler images. Compared with classic approaches, learning-based approaches demonstrated advantages in time efficiency and segmentation accuracy.

C. Robotic US Screening System

The traditional free-hand US examination suffers from inter- and intra-operator variations, which significantly impair the clinical acceptance of the US modality. Benefiting from the controllable robotic mechanism, accurate and repeatable US images can be achieved by properly tuning the acquisition parameters. The characteristic of reproducibility is crucial for clinical applications requiring long-term care, e.g., monitoring the changes of vascular plaque or internal tumors [25]. In addition, the development of RUSS has the potential to relieve sonographers from tedious and burdensome workloads, thereby reducing work-related musculoskeletal disorders.

To achieve reproducible US images, similar acquisition parameters (i.e., contact force, and probe orientation) are necessary across multiple US sweeps, yet, this is challenging even for experienced sonographers. The advantage of multiple sensing sources and precise adjustment of servomotors allows robotic manipulators to accurately repeat the US acquisitions. To obtain high-quality US images, Gilbertson *et al.* presented a compliant controller for a one degree of freedom (DoF) mechanism to stabilize images during examinations [6]. Pierrot *et al.* proposed a hybrid force/position controller for a 7-DoF RUSS based on an external force/torque sensor [7]. A low-level embedded joint controller incorporating a PID algorithm was employed. The outer control loop, driven by the PID algorithm, was executed on an external workstation and utilized either force or position as the reference variable. This significantly improved the adaptability to other robotic manipulators. Regarding the optimization of probe orientation, Z. Jiang *et al.* quantitatively measured the effects of probe orientation on the resulting images and proposed a mechanical, model-based, approach to automatically identify the normal direction of an unknown constrained surface [8]. Due to the intrinsic physics of wave signal propagation, the orthogonal orientation of the probe can lead to US images with higher contrast because more signals can be reflected back to the US probe rather than scattering away [26]. To provide stable image quality during scans, particularly for soft tissue like the breast, Tan *et al.* proposed a flexible RUSS and an online force adjustment approach based on real-time image feedback [9].

To quantify limb arterial stenosis, Janvier *et al.* used a 6-DoF robotic manipulator to control and standardize the 3D US acquisition process for large scanning distances [11]. The scanning path is generated in manual teaching mode for individual patients. To identify the stenosis location, the inner diameters of the vascular phantom were computed. Janvier *et al.* further proved clinical feasibility on volunteers by comparing 3D vascular volumes computed from B-mode and Doppler images, respectively, to the pre-scanned CTA [27]. To realize autonomous scanning, Merouche *et al.* moved the probe in a given direction step by step and computed the in-plane movement based on the segmented lumen to centralize the vessel of interest [12]. Based on both in-vitro and in-vivo validations, the feasibility of applying RUSS to reconstruct limb vessels in a clinical context was investigated.

To further eliminate the requirements of manual selection of the scanning start and end point, Virga *et al.* employed a surface registration to transfer a generic scanning path planned on a preoperative MRI to the current environment for autonomous aortic screening [28]. To guarantee the overall imaging quality, they optimized the contact force by maximizing the overall confidence value of the resulting images. The US confidence value is often used to approximate the strength of US signal at each pixel, which can be computed based on [29]. Since the screening trajectory was fixed during the scanning, the resulting image quality decayed significantly if the objects moved during the scan. Regarding the examination of objects that extend over a large distance, i.e., limb arteries, sonographers even need to actively adjust patients' limbs to fully visualize the complete arterial tree. To tackle this practical challenge, Z. Jiang *et al.* proposed a motion-aware system based on an RGB-D camera and passive markers to

monitor and compensate for potential movements of the patient during the scans [30].

Recently, Z. Jiang *et al.* presented an end-to-end RUSS to automatically scan limb arteries based on the real-time image feedback [22]. The probe orientation adjustment was calculated, estimating the local vascular diameters and consequently solving a set of modeled optimization equations. The results demonstrated that their approach can effectively improve the accuracy and stability of scans compared with the free-hand US manner. Huang *et al.* imitated clinical protocols to automatically move the probe along the longitudinal direction of the carotid artery [23]. To automatically search for the standard longitudinal plane of tubular structures, Bi *et al.* proposed a reinforcement learning network [31]. The segmented binary masks generated by a classic U-Net [14] were used as the state representation. This can bridge the gap between the simulated training environment and the real scenario, thereby achieving good generalization ability.

III. Vessel Segmentation

Accurate segmentation of tubular structures from cross-sectional B-mode images is crucial for achieving accurate geometry of the vessel of interest; thereby, realizing accurate diagnosis and evaluation of PAD. Due to the difficulties implied by US artifacts such as speckle and occlusion, it is challenging to obtain a robust, reliable, and repeatable segmentation process [16]. This limits the development of the autonomous screening program for PAD. To improve the segmentation performance, we propose a DopUS network using two encoders and one decoder to fully take advantage of the two modalities present in the duplex US images. Finally, the proposed DopUS Network is compared to the standard U-Net and other reference structures.

A. Dataset

1) US Data Recording: In this work, all US images were recorded from an ACUSON Juniper US machine (Siemens Healthineers, Germany) using a linear probe 12L3 (Siemens Healthineers, Germany, acquisition width: $51.3\ mm$). To access the duplex images, a frame grabber (Epiphan Video, Canada) was used to connect the US machine and the main workstation via a USB interface. Due to the generation of color Doppler images, a small recording frequency ($10\ fps$) was used. Since the main limb arteries are located close to the skin surface, the image depth for the transducer was set to $45\ mm$. The other acquisition parameters were set as follows: Tissue Harmonic Imaging (THI): $8.4\ MHz$, Dynamic Range (DynR): $75\ dB$, US imaging focus: $20\ mm$. The focus defines the region with the highest imaging quality on the resulting images. The THI is the method often used to increase contrast resolution, and the DynR represents the difference between the largest and smallest signals, which governs the images' gray scale levels.

The US images were recorded from seven volunteers including two females and five males. For sweep recording, the volunteers were asked to sit comfortably and put their left arm on a flat table. Every patient was scanned two times from the start of the brachial artery at the inner side of the shoulder joint towards the wrist. Both scans include the section from the shoulder to the bifurcation of the brachial artery into the radial and ulnar arteries. Thereafter, one scan followed the ulnar artery, while the other focused on the radial artery. Thus, the major arteries of the upper limb are present in the dataset.

The scans were carefully evaluated and manually annotated under the supervision of an experienced physician. As a consequence two of them were discarded due to improper image quality. A total of 12 sweeps with $400 - 500$ images per sweep were used for the dataset. Due to the fact that a recurrent structure is used in the network, sequential data is required. The sweeps were therefore split into 279 sequences with a fixed length of 20 images/sequence. All in all 5580 labeled images, with corresponding color Doppler images and B-mode images, were used for training.

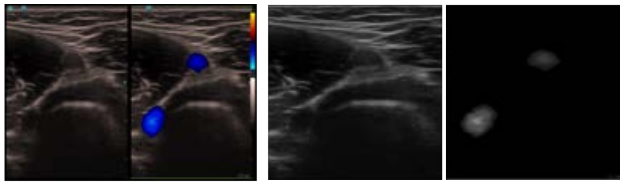

Fig. 2. Illustration of the pre-processing pipeline.

2) Duplex Data Pre-Processing: The frame grabber can only provide access to the same images displayed on the US machine screen Fig. 2 (a). In order to segment the US images and further visualize the segmented binary masks in 3D in real-time, online pre-processing was necessary. The received images [Fig. 2 (a)] were processed to generate B-mode images [Fig. 2 (b)] and Doppler images [Fig. 2 (c)], respectively. To extract Doppler features, the recorded images were converted to the HSV color space. Thereby, the colored Doppler features (blue or red in Fig. 2 (a)) can be easily extracted by setting a threshold in terms of saturation (≥ 100) and value (≥ 20). To reduce the number of input parameters, both duplex images were compressed to 320×320 pixels for the segmentation network. The original size of the images was 497×733 pixels. It should be noted that this downsampling procedure may introduce some degree of distortion and possible information loss. However, we deem it acceptable given that the blood vessels in question occupy a contiguous area that remains discernible in the lower-resolution representation. The implementation of the pre-processing pipeline was done using the open-source library OpenCV [32].

B. Network Architecture

Benefiting from the development of deep learning, neural networks have been widely employed to solve real-time image segmentation tasks. In the field of medical image segmentation, U-Net [14] is one of the most commonly used and successful networks. The U-Net and its variants have also been used for US vessel segmentation tasks [22], [24].

In this work, the proposed DopUS-Net also uses the classic U-shape backbone for the encoder and decoder (see Fig. 3).

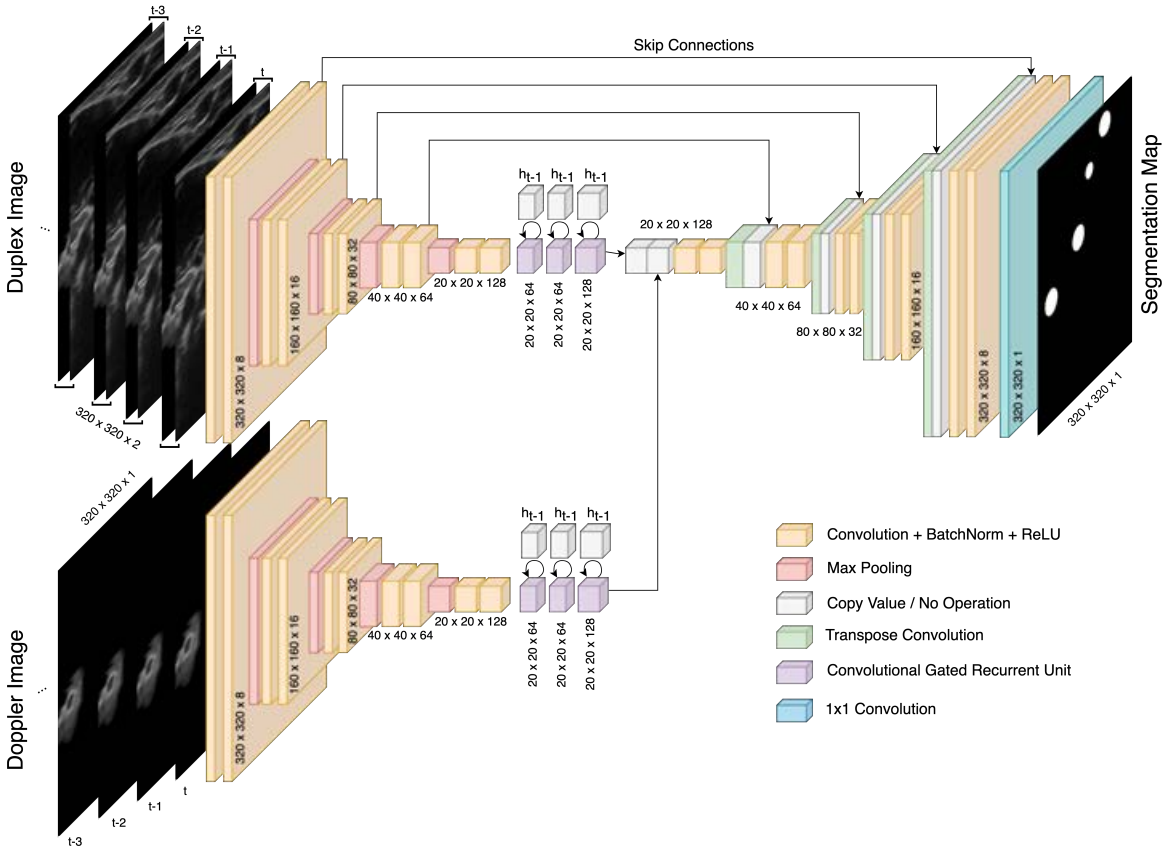

Fig. 3. The architecture of the proposed DopUS-Net.

Contrary to other networks which also take the Doppler inputs into account [15], [24], the DopUS-Net uses two separate encoders. The paired duplex image is considered as a two-channel input of the top encoder, while the bottom encoder only uses the pre-processed color Doppler images. Since only Doppler images are used in the bottom encoder, the network is forced to filter out the noise and focus on the main artifacts. This structure can limit the negative impact of the unstable Doppler signal (e.g., spatial accuracy and consistency), but still preserves the functionality of a region proposal signal. The skip connections between the top encoder and decoder are used to ensure the spatial correctness of the B-mode image meanwhile also including the region proposal functionality of the Doppler images, respectively. This extensive use of the Doppler signal in both encoders proved to be beneficial in terms of performance, which also indicates its usefulness. Besides, DopUS-Net employs batch normalization instead of dropout between the layers as suggested by [33], where it is claimed that using dropout can have detrimental effects on CNN training.

Due to the continuity of vascular tissues, recurrent neural network (RNN) structures are employed to facilitate the segmentation task by taking advantage of sequential historic information. Due to the weight update rule, gradient exploding or vanishing can happen when working with long input sequences [34]. To tackle this problem, long short-term memory (LSTM) [35] and gated recurrent unit (GRU) [36] networks have been developed by using multiple gates internally to keep track of important information meanwhile discarding irrelevant data. Chung *et al.* tested both RNNs and claimed that GRU networks are less complex while maintaining comparable performance to an LSTM [37]. Thereby, GRUs are used in the proposed network structure, which leads to fewer computations and therefore better inference time. To further reduce the number of parameters needed, Ballas *et al.* introduced the convolutional gated recurrent unit (convGRU) network [38]. The hidden state $\mathbf{h}_t$ extracted from sequential inputs is computed as follows:

$$\begin{aligned} \mathbf{z}_t &= \sigma(\mathbf{W}_z[\mathbf{x}_t, \mathbf{h}_{t-1}]^T) \\ \mathbf{r}_t &= \sigma(\mathbf{W}_r[\mathbf{x}_t, \mathbf{h}_{t-1}]^T) \\ \tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}[\mathbf{x}_t, (\mathbf{r}_t \odot \mathbf{h}_{t-1})]^T) \\ \mathbf{h}_t &= (1 - \mathbf{z}_t)\mathbf{h}_{t-1} + \mathbf{z}_t\tilde{\mathbf{h}}_t \end{aligned} \tag{1}$$

where $\odot$ is an element-wise multiplication, $\mathbf{x}_t$ is the sequential input, $\mathbf{h}_{t-1}$ is the hidden state computed in the last iteration, $\mathbf{z}_t$ is an update gate that dominates the degree to which the unit refreshes its content, and $\mathbf{r}_t$ represents the reset gate. $\mathbf{W}_z$ and $\mathbf{W}_r$ are the weights for the update gate and reset gate, respectively. $\tilde{\mathbf{h}}_t$ further operates on combining the input with an updated hidden state of the previous time step. The final hidden state $\mathbf{h}_t$ is computed based on the result of the update gate.

The convGRU is used at the bottleneck in both encoders

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2024, Yokohama, Japan.

to enable state tracking functionality. The bottom encoder is, therefore, able to filter out noise and speckles from the Doppler effect meanwhile maintaining a representative view of the recurring pulsation of the Doppler artifacts in the arteries. The top encoder can use its historic information to emphasize previously found vessel locations in the current segmentation. The fusion of the two encoders and their respective recurrent unit is done via a concatenation. Afterward, a convolution ties the input together and reduces the feature dimensionality. Consequently, the result of the combined bottleneck is fed into the decoder where a 1×1 convolution is used at the initial image resolution to generate the binary output mask (see Fig. 3).

C. Training

1) Data Augmentation: In this work, data augmentation is applied during the training process to improve robustness and reduce potential overfitting [39]. To maintain a constant augmentation pattern for the complete sequence, the same parameters were applied to all images in the same sequence. The spatial augmentations parameters are listed in TABLE I. During the training, a random number is chosen among the given ranges. We empirically determined that applying horizontal and vertical shifts of only $\pm 1\%$ was optimal for our model's performance. Larger shifts, in combination with the other parameters, decreased the model's efficacy on real US data, although further optimization is possible in future work.

TABLE I
APPLIED AUGMENTATIONS TO THE IMAGE SEQUENCE

Augmentation Method	Range / Value
Horizontal Shift	[-1% , 1%]
Vertical Shift	[-1% , 1%]
Rotation	[-15°, 15°]
Scale	[80%, 120%]
Horizontal Flip Probability	50%

2) Loss Function: In medical binary segmentation problems, binary cross entropy (BCE) is the most commonly used loss function [40]. The pixel-wise approach of class probabilities leads to a smooth gradient curve. However, BCE is sensitive to imbalanced classes. Particularly in our case, the limb arteries are very small in comparison to the background. The area recorded by the US probe is around 2300 mm^2 (transducer length: 52 mm, depth: 45 mm), while the radial or the ulnar artery only covers approx $5-10$ mm^2 (diameter: $2-2.5$ mm [13]) of this region. This leads to a possible class imbalance of $1:250$. To tackle this problem, the soft dice loss function [41] is used instead in this work. The soft dice loss for all classes is then averaged to obtain a final score. Besides, Bertels *et al.* also reported that the soft dice loss is recommended when using the dice score as a performance index [40]. The computation of the soft dice loss L is presented in Eq. (2).

$$L = 1 - \frac{2|\tilde{Y} \cap Y| + 1}{|\tilde{Y}| + |Y| + 1} = 1 - \frac{2 * \sum_{i=1}^{N} \tilde{y} * y + 1}{\sum_{i=1}^{N} \tilde{y}^2 + \sum_{i=1}^{N} y^2 + 1} \quad (2)$$

where N represents the total amount of pixels, y and $\tilde{y}$ symbolize the target value (ground truth) and predicted label value, respectively.

3) Training Details: To optimize the weights of DopUS network, the Adam [42] optimizer is used in this work. Additionally, mixed-precision training [43] is enabled to speed up the training process. This form of auto casting weight precision reduces the overall memory requirements and therefore increases the training speed significantly. There are 20 images in each sequence data set. Since the robot moved at a pace of 10 mm/s and the duplex images are recorded in 10 fps, the length of each sequence is around 20 mm. The parameters for the training epoch and the batch size are set to 1000 and 16, respectively. The initial learning rate is 1e-4 and halved every 250 iterations. The use of a small learning rate at the end of the training can benefit the accurate convergence. To prevent unnecessary training epochs, an early stopping mechanism was implemented. If the validation loss is not decreasing within 15 epochs, the algorithm is assumed to have converged and no further improvements are expected. Since recurrent structures are present in the network, truncated backpropagation through time is applied. Every 4 steps the weights of the network are updated. This value represented a good compromise between performance and training time.

IV. ROBOTIC US SCANNING AND ARTERY RE-IDENTIFICATION

To guarantee the patient's safety and image quality, the robotic scans are performed using compliant control [22]. The scan trajectory is manually defined by selecting the start and end position on the patient's skin. To achieve a highly accurate 3D visual representation of the artery tree, Doppler images and B-mode images are used jointly to extract the vessel from the cross-sectional images. A stable and continuous Doppler signal is beneficial for the performance of the proposed DopUS-Net. To account for unstable Doppler signals and problems introduced by the blood pulsation, a tracking algorithm is developed in Section IV-B. Based on this tracker, an online image check is activated to guarantee sufficient quality of the Doppler signal. Once this check identifies poor Doppler signal, which is often accompanied by unstable segmentation results, a re-identification procedure is performed to relocate the Doppler signals (see Section IV-C).

A. Compliant Control Architecture

The impedance controller is often used to maintain the contact force between the probe and the contact surface [22], [44]. Due to the use of built-in joint torque sensors in all seven joints, the impedance control law can be defined as follows:

$$\tau = \mathbf{J}^T[\mathbf{F}_d + \mathbf{K}_m e + \mathbf{D}\dot{e} + \mathbf{M}\ddot{e}] \quad (3)$$

where τ is the computed torque, J^T is the transposed Jacobian matrix, $e = (x_d - x_c)$ is the pose error (position and orientation) between the current pose x_c and the target pose x_d in Cartesian space, $\mathbf{F}_d$ is the supposed exerted force/torque at end-effector, $\mathbf{K}_m$, $\mathbf{D}$ and $\mathbf{M}$ represent the matrices of stiffness, damping and inertia terms, respectively. According to [44], the stiffness in the direction of the probe centerline is usually set in the range $[125, 500]$ N/m for human tissues.

B. Doppler Signal Tracker

Due to the heart pulsation, the Doppler images are not as stable as B-mode images. Regarding the scanning of limb arteries, both the accuracy and strength of Doppler signals are influenced by factors like the contact condition and the probe orientation. This may impair the performance of the segmentation. However, the ability to automatically identify the flow location in the US image is important, particularly for developing autonomous scanning programs. The inclusion of the Doppler signal enables quality-aware robotic screening, which can assess the segmentation results online and leads to improved 3D reconstruction results across the scans.

To assess the quality of real-time Doppler signals, we monitor the centers of detected flow areas [see Fig. 2 (c)]. Due to the continuity of the objects, the corresponding centers should be close to each other on consecutive frames. To compute the centers, all pre-processed Doppler images are examined for contours using the minimum enclosing circle algorithm, which is a standard algorithm wrapped in the open-source library OpenCV [32]. Considering the inevitable Doppler noise, only the contours above a certain empirical threshold (radius > 1.2 mm) are kept. Afterward, the centers are assigned to tracking objects capturing the historical information of the Doppler signal. To further process the potential case that multiple vessels are displayed on a single image, the centers need to be assigned to an existing tracking object $\mathbb{O}^{t-1} \in R^{N\times 1}$ or to be considered as a new object, where N is the number of the tracked objects, namely blood vessels. To this end, the distance between the centers extracted from the images recorded at time $t-1$, $\mathbf{C}^{t-1} \in R^{N\times 2}$, and the centers computed on the current images at time t, $\mathbf{C}^{t} \in R^{N'\times 2}$, is computed as follows:

$$\mathbf{D}_i^t = \text{CalDis}(\mathbf{C}_i^{t-1},\ \mathbf{C}_{all}^t) \tag{4}$$

where CalDis() represents the operation to compute Euclidean distances between one point and each element in a point set. $i = 1, 2, ..., N$ is the iterator referring to the centers of tracked objects, $\mathbf{C}_i^{t-1} \in R^{1\times 2}$ is the i-th center saved at $\mathbb{O}^{t-1}$, $\mathbf{C}_{all}^t \in R^{N'\times 2}$ is the set of centers computed based on the Doppler image obtained at time t, N' is the filtered number of contours present at time t, and $\mathbf{D}_i^t \in R^{N'\times 1}$ is the distances between $\mathbf{C}_i^{t-1}$ and all elements of $\mathbf{C}_{all}^t$. The minimum distance d_{min} is stored with its associated index j and used to update the tracking object $\mathbb{O}^{t-1}$ as described in Algorithm 1.

The tracking object $\mathbb{O}$ is updated based on the distance between the contours' centers calculated in the last frame $t-1$ and the current frame t [Eq. (4)]. The center point of the current frame with the smallest distance to a respective tracking object's center point, $\mathbf{C}_{all}^t(j)$, will become its new center point. To avoid arbitrary center point assignment, a distance limit T_d of maximal 30 pixels is empirically set, which means 4.2 mm maximum deviation from the previous center point. For all center points that could not be matched to an existing tracking object, a new tracking object is created. If, on the other hand, no Doppler signal could be assigned to a tracking object, the center point of the previous frame is stored. Additionally, this value will be labeled so that a distinction between a tracked value and a copied value can be made.

Algorithm 1: Doppler Signal Tracker

Input: previous tracking object $\mathbb{O}^{t-1}$, extracted vessel centers on previous and current images $\mathbf{C}_{all}^{t-1}$ and $\mathbf{C}_{all}^t$

Output: current tracking object $\mathbb{O}^t$

$\mathbb{O}^t \leftarrow \mathbb{O}^{t-1}$;
for $i = 1$; $i \leq$ *len*$(\mathbb{O}^{t-1})$; $i++$ **do**
  $\mathbf{D}_i^t \leftarrow$ CalDis($\mathbf{C}_i^{t-1}$, $\mathbf{C}_{all}^t$) [Eq. (4)];
  $d_{min} \leftarrow \min(\mathbf{D}_i^t) = \mathbf{D}_i^t(j)$;
  if $d_{min} \leq T_d$ **then**
    $\mathbb{O}^t(i) \leftarrow \mathbf{C}_{all}^t(j)$;
  else
    $\mathbb{O}^t \leftarrow [\mathbb{O}^t, \mathbf{C}_{all}^t(j)]$;
  end
end

C. Artery Location Re-Identification

Compared with existing vessel segmentation and tracking approaches, the use of the Doppler signal enables quality-aware robotic scanning, leading to robust and accurate 3D visualization. Although some learning-based approaches have been proposed to achieve accurate segmentation, there is no approach that can evaluate the segmentation results online. Regarding pure segmentation tasks from images, a large segmentation error on a few slices would not affect the overall performance. However, such errors may lead to incorrect robotic motion when the segmentation is further used to control a robotic manipulator. Thereby, it is necessary to re-identify the artery location once the segmentation quality is inadequate. Here we consider imaging quality to be insufficient when the real-time segmentation results are not consistent with the Doppler images. To ensure a consistently high segmentation performance, a re-identification procedure is developed (see Fig. 4).

To evaluate the segmentation performance, Doppler images and real-time segmentation results are jointly used as inputs. The imaging quality is determined by comparing the last center point of the tracking object $\mathbb{O}$ to the segmentation mask. A tracking object is only considered if at least 20% of its center points within the last 15 frames represent new values (not a copy from the last one). The threshold is determined based on empirical studies and the assumption that a flow is visible in

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2024, Yokohama, Japan.

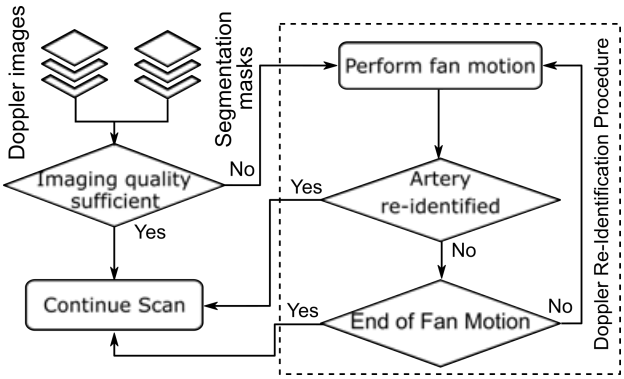

Fig. 4. The flow chart of the artery re-identification procedure.

the last 15 frames. Considering that the images are recorded in 10 fps, 15 frames correspond to a time window of 1.5 s Note that the interval is based on the assumption that at least one heartbeat can occur within that time frame and may need to be adjusted for patients with slower heart rates.

Furthermore, if at least one of the tracked center points of the valid tracking objects is within the areas of the predicted vessels, the quality is assumed to be sufficient. Otherwise, we consider that the tracker and real-time segmentation results are in conflict and thereby automatically start the re-identification algorithm. In this case, the current probe pose is referred to as the original orientation for this re-identification procedure (see Fig. 5). Due to the physical principles of the Doppler signal, its quality gets better when the beam of the US probe is aligned with the direction of the vessel flow. Better Doppler imaging quality can further facilitate the segmentation performance of the proposed DopUS network. Thereby, once the system is aware that the imaging quality is not good enough, an out-of-plane fan motion is performed to obtain better imaging quality (see Fig. 5). However, the contrast of the B-mode images decreases when tilting the probe away from the optimal perpendicular pose. In order to keep both the US and Doppler image quality as high as possible, the fan motion should only rotate as far as necessary.

Taking that into account, the probe is rotated in the out-of-plane direction in 5° steps every time. In total, a range of $\pm 10^\circ$ from the current probe orientation can be covered. After every step, the robot stops for two seconds to be notified if the Doppler quality check reports a good signal. A maximum of five different orientations (out-of-plane rotation deviation from the current pose: $[-10^\circ, -5^\circ, 0, 5^\circ, 10^\circ]$) are visited, assuming that no sufficient quality of the Doppler signal is detected. In this case, the robot returns to its original orientation and continues the sweep. In order to move on to a potentially better location for the Doppler signal the re-identification process is blocked for at least three seconds. If on the other hand the Doppler signal relocalizes the vessel during the re-identification process, the current orientation is used to update the trajectory orientation and the sweep is continued. Finally, to prevent the robot from moving too close to the human arm, a security threshold referring to the original direction is implemented. So the out-of-plane rotation is stopped if it deviates more than 20° compared to the original orientation.

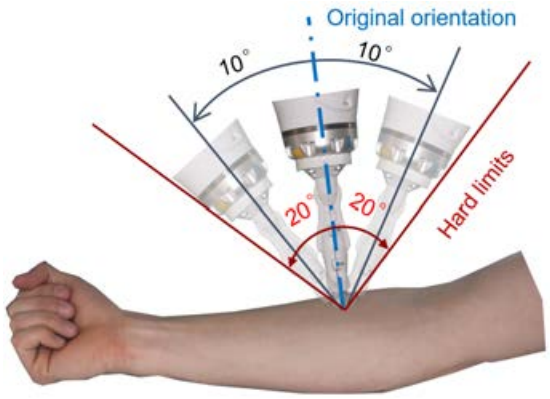

Fig. 5. The illustration of the out-of-plane rotation for the re-identification process.

V. Results

A. Experimental Setup

The overall setup is visualized in Fig. 6. A linear probe is rigidly attached to the robot manipulator (LBR iiwa 14 R820, KUKA GmbH, Germany). The connector between the transducer and the end-effector of the robotic arm is a custom-designed probe holder. To guarantee safety and imaging quality, the robotic arm is controlled using impedance control during scans (see Section IV-A). The desired force in the probe centerline is set to 1 N, and the stiffness is 200 N/m. The main objective of using a non-zero force is to ensure firm contact between the probe and patients during scans, while force could be varied from 1 N. However, during the re-identification process, the control mode automatically switches to a position control strategy to maintain the same position when restarting.

Regarding the image processing part, the duplex images are recorded and fed to the DopUS-Net for real-time vessel segmentation. The inference time of DopUS-Net, including pre-processing and the online imaging quality evaluation process, is around 20 ms on average, which is quick enough to process the real-time US images captured at 10 fps (100 ms). In addition, the segmentation masks are stacked in 3D based on the robotic tracking information and visualized in real-time on a visualization platform (ImFusion Suite, ImFusion GmbH, Germany).

B. Segmentation Performance on Volunteers

The network was trained on a workstation (GPU: GeForce GTX Titan X, CPU: Intel i7-4820K). To measure the similarity between the manually annotated ground truth data Y (see Section III-A) and the binary segmentation result $\tilde{Y}$, the Dice-Score C_{dice} is applied (Eq. 5).

$$C_{dice} = \frac{2|\tilde{Y} \cap Y|}{|\tilde{Y}| + |Y|} \tag{5}$$

TABLE II
COMPARISON OF THE RESULTS

Network	Top Encoder	Bottom Encoder	# Parameters	Patient 0	1	2	3	4	5	6	Dice Score Mean (SD)
U-Net	B	-	0.6 M	0.61	0.51	0.45	0.40	0.45	0.29	0.38	0.44 (0.09)
U-Net	BD	-	0.6 M	0.83	0.59	0.54	0.52	0.49	0.39	0.45	0.55 (0.13)
DopUS-Net$^{(0)}$	BD	D	1.3 M	0.83	0.61	0.58	0.60	0.54	0.43	0.47	0.58 (0.12)
VesNet	BD-RNN	-	2.6 M	0.80	0.62	0.60	0.43	0.33	0.42	0.43	0.52 (0.15)
VesNet+	BD-RNN	-	6.3 M	0.80	0.62	0.68	0.58	0.58	0.47	0.53	0.61 (0.10)
U-Net	BD-RNN	-	3.0 M	0.86	0.61	0.71	0.66	0.68	0.60	0.50	0.66 (0.10)
U-Net+	BD-RNN	-	6.5 M	0.78	0.58	0.37	0.40	0.36	0.32	0.41	0.46 (0.15)
DopUS-Net$^{(1)}$	B-RNN	D-RNN$^{\clubsuit}$	6.2 M	0.78	0.54	0.45	0.56	0.55	0.27	0.18	0.48 (0.19)
DopUS-Net$^{(2)}$	BD	D-RNN	3.7 M	0.87	0.69	0.76	0.76	0.72	0.56	0.61	0.71 (0.10)
DopUS-Net$^{(3)}$	B-RNN	D-RNN	6.1 M	0.88	**0.71**	0.79	0.76	0.75	0.57	**0.61**	0.72 (0.10)
DopUS-Net$^{(4)}$	BD-RNN	D-RNN	6.1 M	**0.88**	0.69	**0.79**	**0.78**	**0.76**	**0.62**	0.60	**0.73 (0.09)**

**Nomenclature:* B:B-Mode, D: Doppler, RNN: convGRU module, +: increased parameters, DopUS-Net$^{(x)}$: specific DopUS-Net version, $^{\clubsuit}$: additional skip connections from the bottom encoder to the decoder.

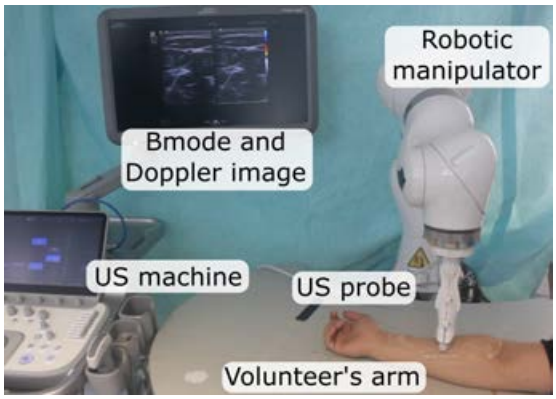

Fig. 6. The experimental setup: robot manipulator with attached linear US probe and the used US machine.

We employed the leave-one-out-cross-validation (LOOCV) method [44] to optimize the performance of the DopUS network with a limited training dataset. LOOCV entails training the model on data from six patients and validating it on the remaining one, iteratively for each patient. This reduces the potential bias of the model compared with conventional approaches that split the data into fixed training and test sets. LOOCV ensures that the model is evaluated on data from all seven patients, albeit at a higher computational cost. It yields a more robust and reliable estimate of the network performance.

To compare the performance with existing networks and further explore the most effective way to incorporate the Doppler signal, the performances of different network architectures have been investigated. All presented networks are implemented using PyTorch[3], and they are trained from scratch based on the same data set in each case. The results have been summarised in TABLE II. To ensure a valid comparison throughout the study, all the models are trained using the same configuration on the same dataset. The baseline and comparison networks represented are U-Net [14] and VesNetSCT++ [15]. **It must be noted**, that the VesNetSCT++ architecture is a custom re-implementation as the source code was not publicly available. Results should be treated with care. The self-developed VesNetSCT++ is carefully implemented by following the architecture presented in [15]. The final version achieved comparable dice scores on small blood vessels (diameter is 2-3 mm) with respect to the original paper [15]. To maintain simplicity, the VesNetSCT++ [15] is referred to as VesNet in the remaining part of this work.

As all involved network architectures use the U-shape backbone, descriptions of their top and bottom encoders are used to facilitate the understanding of the key differentiation aspects. The top encoder represents the encoder in the U-Net/VesNet architecture. Since the proposed DopUS-Net follows the approach of two separate encoders, the second one is described as the bottom encoder (see Fig. 3). The corresponding inputs of these two encoders used in different networks are specified in TABLE II. The letter "B" and "D" are B-mode and Doppler images, respectively. BD represents a two-channel tensor consisting of the combined duplex images. RNN corresponds to the convGRU module, which is used before the bottleneck at the highest feature dimensionality space.

1) Effectiveness of Doppler Signal on Vessel Segmentation: To validate whether the Doppler signal can enhance the segmentation accuracy from US images, the classic U-Net [14] is used as a baseline in this study. In the first two rows of TABLE II, we compare the effect of pure B-Mode image input to a two-channel Doppler and B-Mode tensor input. The average segmentation results are improved from 0.44 to 0.55 in terms of dice score. Such improvement is consistent with the results reported by B. Jiang *et al.* using VesNet [15]. Then, the additional bottom encoder using the Doppler images as input is tested. The double encoder approach further improved the segmentation results to 0.58. It is noteworthy that DopUS-Net^0 consists of a classic U-Net and an additional encoder with the same architecture as the upper encoder. Thereby, we consider the use of Doppler signals can significantly improve

[3]https://pytorch.org/

the segmentation performance.

2) Effectiveness of RNN on Vessel Segmentation: Due to the continuity of blood vessels, RNNs are used to take advantage of historical information in this study. To evaluate the impact of adding an RNN, we compare the performance of the U-Net with and without the RNN module. Based on the experimental results in TABLE II, the performance is significantly improved from 0.55 (2-nd row) to 0.66 (6-th row) when the RNN module is used. Similarly, the best performing version of the DopUS-Net$^{(4)}$ with RNN for both encoders increases its performance by roughly 15%. Thereby, we consider the RNN to be beneficial for the achievement of accurate vessel segmentation.

To account for parameter differences introduced by the different structures, the network parameters of the comparison networks have been increased by increasing the number of filters in each layer. The version with increased parameters of U-Net and VesNet are marked as U-Net+ and VesNet+, respectively. The parameter numbers of U-Net+ ($6.3\ M$) and VesNet+ ($6.5\ M$) roughly match the size of the final DopUS-Net ($6.1\ M$). Based on the experimental results, the VesNet+ achieved a better result (0.61), while the U-Net+ performed significantly worse than the baseline (only 0.46). We consider this to be caused by the overfitting when the number of parameters (U-Net) significantly increased. To address this issue, necessary fine-tuning and regularization would be required, e.g., dropout, which in turn would result in a parameter reduction. Therefore, the better performing version with fewer parameters is used for further comparison.

3) Effectiveness of Network Structures on Vessel Segmentation: To further explore the most effective way of incorporating the Doppler signals to facilitate segmentation accuracy, four different versions of DopUS-Net have been investigated. DopUS-Net$^{(1)}$ and DopUS-Net$^{(3)}$ only use the Doppler signal in the bottom encoder, while DopUS-Net$^{(2)}$ and DopUS-Net$^{(4)}$ use Doppler signals in both encoders. First, we quantitatively evaluated the effect of the skip connections. DopUS-Net$^{(1)}$ incorporates additional skip connections from the bottom encoder to the decoder. Compared to DopUS-Net$^{(3)}$, which only has the skip connections from the top encoder to the decoder, the performance of DopUS-Net$^{(1)}$ is significantly reduced, deteriorating the dice score by 0.24. We consider that this is because of the inherent characteristic instability of the Doppler signal. The results affirm the assumption that the Doppler signal can be seen as a good region proposal without spatial accurateness. In addition, compared to the final version DopUS-Net$^{(4)}$, DopUS-Net$^{(2)}$ and DopUS-Net$^{(3)}$ remove the RNN module and Doppler signal, respectively, from the top encoder. Although the experimental result demonstrated that DopUS-Net$^{(4)}$ achieved better performance than DopUS-Net$^{(2)}$ and DopUS-Net$^{(3)}$, the improvement is minor. All of these three networks significantly improve the state-of-the-art results (U-Net: 0.66, VesNet: 0.61) to 0.71, 0.72 and 0.73. Overall the insight can be taken, that the second encoder forces the network to emphasize on the Doppler effect. The results show that the intuitive inclusion and emphasized weight on the Doppler signal is justified in practice. More segmentation results of live US frames can be found in the online **video**.

C. Validation of the Artery Re-Identification Process

To further validate whether the proposed artery location re-identification approach can improve the robustness and final 3D compounding results, experiments were carried out on various volunteers. To quantitatively compare the difference between the scans with and without the re-identification process, two robotic scans with the same initial scanning trajectory were performed on the same volunteer. The initial trajectory was manually given by selecting the start and end points. The ulnar artery on the arm was selected as the anatomy of interest. Representative results of two scans with and without the re-identification process, respectively, on the same volunteer, are depicted in Fig. 7. To avoid misleading information, it is noteworthy that the segmentation noise of the DopUS-Net has been removed to focus on the vessel of interest. The heatmap intuitively shows the geometric offset between the segmentation result of the DopUS-Net and the manually annotated ground truth. Similar to the dataset used for training and evaluation of the proposed network structure, the ground truth annotation for the artery re-identification validation was performed under the supervision of an experienced sonographer.

TABLE III
SEGMENTATION RESULTS OVER VOLUME [MEAN (SD)]

Re-Identification	Dice Score	IoU
Enabled	0.86 (0.14)	0.78 (0.16)
Disabled	0.54 (0.39)	0.47 (0.37)

It can be seen from Fig. 7 that the re-identification process (case 2) can result in a relatively complete 3D compounding result, while the scan (case 1) without the re-identification process performed poorly. By monitoring the real-time segmentation masks and Doppler images, the system with the re-identification process can be aware of sub-optimal segmentation results. Thereby, automatic probe optimization is performed to achieve good segmentation results for an accurate 3D reconstruction. This characteristic is important to improve the robustness of the automatic US scanning program. Regarding case 2, three re-identification procedures were performed. Qualitatively, the colored geometric distance [see Fig. 7 (a) and (d)] demonstrate that the distance offsets of some segmentation results in case 1 are larger than $2\ mm$, while the maximum offset in Case 2 is around $0.5\ mm$.

To further quantitatively compare the results, the Dice Score and IoU are used as performance indicators. The scores for individual frames in the sweeps are depicted in Fig. 7 (b) and (e). The changing tendency in both indicators is consistent with the corresponding distance offset heatmaps. To statistically summarize the performance scores, we group the scores in ten bins ($[0, 0.1)$, $[0.1, 0.2)$,..., $[0.9, 1]$) as can be seen in Fig. 7 (c) and (d). Based on the histogram, there are 204 frames (62 for [0.8,0.9) and 142 for [0.9, 1.0]) of 242 frames in total (84.3%) which achieved a high dice score in case 2, while at the same time there are only 88 frames (38 for [0.8,0.9) and 50 for [0.9, 1.0]) of 183 respective total frames (48.1%) that achieved a high dice score in case 2. As for the IoU indicator,

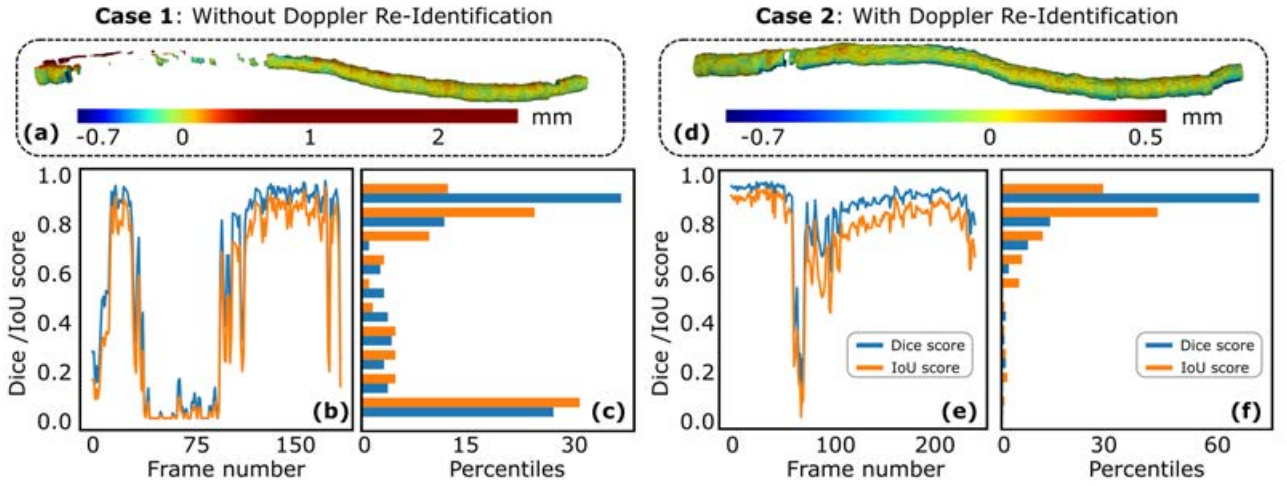

Fig. 7. A representative comparison between the robotic scans with and without artery location re-identification process. (a) is the heatmap of the geometry difference between the automatic segmentation results and the ground truth. (b) and (c) are the dynamic and statistically summarized Dice/IoU scores, respectively. (d), (e) and (f) are the results obtained when scanning with enabled re-identification process.

TABLE IV
OVERALL PERFORMANCE OF THE ROBOTIC SCREENING FRAMEWORK [MEAN(SD)]

Objects	Time efficiency		Re-identification			Sweep
	Total (s)	Compounding (s)	Occurrences	Re-identification steps	Success rate	distance (mm)
Volunteer 1	39(12)	0.41(0.17)	2.5(1.1)	1.1(0.3)	86%	103(14)
Volunteer 2	49(15)	0.54(0.17)	2.8(1.3)	1.4(0.9)	92%	113(9)

it has a similar tendency as the dice score. In addition, many results are distributed in the range of $[0, 0.5]$ for case 1, while the segmentation scores for case 2 are mainly distributed in the upper part. Finally, the average results are summarized in TABLE III. The experimental results demonstrated that the re-identification procedure can effectively improve the robustness and accuracy of the segmentation result (dice score: from 0.54 to 0.86; IoU: from 0.47 to 0.78).

D. Performance of the Proposed Quality-Aware Blood Vessel Screening Framework

To further quantitatively analyze the overall performance of the whole framework, we test the proposed quality-aware robotic screening approach on the forearm of two healthy volunteers. Ten independent experiments are carried out on each volunteer (in total 20 times). Considering the clinical scenarios, we investigate the effectiveness of the Doppler-based re-identification procedure and the time efficiency of the whole scanning. The results have been summarized in TABLE IV.

In order to evaluate the efficacy of the proposed method, experiments were conducted in which the start and end points of the scan trajectory were set manually. The average sweep distance of volunteer 1 and 2 were $103 \pm 14mm$ and $113 \pm 9mm$, respectively. The time efficiency was analyzed by measuring the compounding time for the 3D reconstruction and the total time required for the entire workflow. The greater sweep distance for volunteer 2 resulted in a longer scanning time ($49 \pm 15s$) compared to volunteer 1 ($39 \pm 12s$), and a proportional increase in compounding time ($0.54 \pm 0.17s$ vs $0.41 \pm 0.17s$). Overall, the experiments demonstrated that the robotic scans can be completed within a minute, meeting the standard of clinical practices in terms of time efficiency.

To validate the necessity of the re-identification procedure, we count how often the procedure was activated during scans and also how many out-of-plane adjustment steps were necessary to re-identify a good Doppler signal. It can be seen from TABLE-IV that the average occurrences and required steps in each re-identification procedure on volunteer 1 and volunteer 2 are 2.5 ± 1.1 times vs 2.8 ± 1.3 times and 1.1 ± 0.3 times vs 1.4 ± 0.9 times, respectively. These statistical numbers obtained independently from two different volunteers are very similar. The re-identification procedure was activated over two times on average per scan for both volunteers, emphasizing its crucial role in order to maintain a high segmentation performance. To ensure the accuracy and completeness of the reconstructed vessel (see Fig. 7), the Doppler-based re-identification is activated. The results demonstrate that after less than two steps of out-of-plane adjustment (1.1 and 1.4 for volunteers 1 and 2, respectively), the system can recover a good Doppler signal. In addition, if the system cannot revive good Doppler results after five searching steps, we consider the re-identification procedure failed. The success rate is 86% (successful / total = 22/25) and 92% (26/28) on volunteers 1 and 2, respectively. Factors affecting the success rate can include variations in the positioning of the probe over the artery, differences in blood pressure which affects the blood speed, and variations in tissue depth and composition

above the arteries. The experimental results demonstrate that the presented re-identification procedure can effectively and quickly adjust the probe to relocate to the orientation leading to good Doppler signal quality.

VI. Discussion

We present a novel quality-aware robotic US screening framework for tubular structures. Leveraging the Doppler signal, the vascular segmentation performance is improved using the proposed DopUS-Net, and the accuracy and completeness of the reconstructed 3D vessel of interest are significantly enhanced on multiple volunteers. The novel structure of DopUS-Net can also inspire other studies with multiple inputs, such as combining the frequency and image domain with OCT image segmentation [45]. However, there are some limitations that are worth noting. First, the patient's motion is not considered in this work, but such motion could result in disconnected blood vessels in the 3D view. To further address this challenge, previous work on motion-aware RUSS [30], [46], [47] could be integrated to monitor and compensate for comparatively large patient motion. Then advanced computational 3D compounding methods [48], [49] need to be developed to tackle the pixel-wise misplacement to guarantee local continuities. Second, the re-identification is only performed in the out-of-plane direction to recover the Doppler signal in this work. This is the compromised result to reduce the complexity of the re-identification solution, thereby ensuring the time efficiency to meet the requirement for further clinical translation. In the future, if we can access the low-level control of the used US machine, another promising solution is to maintain the probe normal to the constraint surface and directly change the emitted US wave direction. This will further improve the time efficiency by eliminating the need to adjust the probe orientation physically while maintaining stable contact between the probe and the surface. However, it must be noted that this approach requires specialized transducers and cannot be achieved using standard US probes.

In addition, data collection and labeling are often burdensome and expensive tasks, particularly in the field of medical image processing. To improve data efficiency, data augmentation is one of the most popular methods used [50]. To improve the generalization capability of a trained model on unseen data, representation disentanglement [51], [52] is investigated to explicitly disentangle the feature in latent space against the domain shift. This is particularly useful for US images because US images are very sensitive to the machine, US machine setting, and also real-time contact conditions. In addition, considering the shortage of data, an emerging concept of federated learning is proposed to enable the training across multiple decentralized servers holding local data examples [53]. Currently, real-time segmentation of ultrasound images is mainly performed on 2D images due to their superior accuracy in capturing local features like boundaries. However, once 3D ultrasound probes generate high-quality volumetric images, 3D segmentation algorithms [54], [55] can potentially improve segmentation performance by maintaining the accuracy of local features and enabling anatomical continuity in three dimensions.

VII. Conclusion

This work presents a novel approach using duplex images to facilitate the accurate segmentation of small blood vessels from cross-sectional US images on volunteers' limbs. To explore the most effective way to incorporate the Doppler signals, various versions of the DopUS-Net have been proposed and compared to the U-Net [14] and VesNet [15] in terms of dice score. The final version of DopUS-Net has two encoders with different inputs (two-channel "BD" and pure "D" images), and two convGRUs applied at the bottleneck layer to exploit the continuity properties of the anatomy from previous frames. In addition, based on the Doppler signal, an online artery re-identification module is developed to qualitatively assess the performance of the real-time image segmentation. Thereby, the quality-aware robotic screening program is developed to further improve the confidence and robustness of the image segmentation. The experimental results demonstrate that the overall performance of the scan with the re-identification process is significantly improved over the scan without the re-identification process on the same volunteer (dice score: from 0.54 to 0.86; IoU: from 0.47 to 0.78). The Doppler-based quality-aware screening significantly improved the accuracy and robustness of the 3D compounding results; thereby, we anticipate that it may further improve the acceptance of the autonomous US scanning programs in the future.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Med. Reza Ghotbi and Dr. Med. Angelos Karlas for their insightful discussions and valuable clinical feedback. Besides, we want to thank the Editors and reviewers for their implicit contributions to the improvement of this article.

References

[1] G. J. Hankey, P. E. Norman, and J. W. Eikelboom, "Medical treatment of peripheral arterial disease," *Jama*, vol. 295, no. 5, pp. 547–553, 2006.

[2] A. T. Hirsch, M. H. Criqui, D. Treat-Jacobson, J. G. Regensteiner, M. A. Creager, J. W. Olin, S. H. Krook, D. B. Hunninghake, A. J. Comerota, M. E. Walsh *et al.*, "Peripheral arterial disease detection, awareness, and treatment in primary care," *Jama*, vol. 286, no. 11, pp. 1317–1324, 2001.

[3] K. J. Rocha-Singh, T. Zeller, and M. R. Jaff, "Peripheral arterial calcification: prevalence, mechanism, detection, and clinical implications," *Catheterization and Cardiovascular Interventions*, vol. 83, no. 6, pp. E212–E220, 2014.

[4] E. Favaretto, C. Pili, and et al., "Analysis of agreement between duplex ultrasound scanning and arteriography in patients with lower limb artery disease," *J. Cardiovasc. Med.*, vol. 8, no. 5, pp. 337–341, 2007.

[5] J. Guo, C. Shi, and H. Ren, "Ultrasound-assisted guidance with force cues for intravascular interventions," *IEEE Trans. Autom. Sci. Eng.*, no. 99, pp. 1–8, 2018.

[6] M. W. Gilbertson and B. W. Anthony, "Force and position control system for freehand ultrasound," *IEEE Trans. Robot.*, vol. 31, no. 4, pp. 835–849, 2015.

[7] F. Pierrot, E. Dombre, E. Dégoulange, L. Urbain, P. Caron, S. Boudet, J. Gariépy, and J.-L. Mégnien, "Hippocrate: A safe robot arm for medical applications with force feedback," *Med. Image Anal.*, vol. 3, no. 3, pp. 285–300, 1999.

[8] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab, "Automatic force-based probe positioning for precise robotic ultrasound acquisition," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 11 200–11 211, 2020.

[9] J. Tan, B. Li, Y. Li, B. Li, X. Chen, J. Wu, B. Luo, Y. Leng, Y. Rong, and C. Fu, "A flexible and fully autonomous breast ultrasound scanning system," *IEEE Transactions on Automation Science and Engineering*, 2022.

[10] Z. Jiang, M. Grimm, M. Zhou, J. Esteban, W. Simson, G. Zahnd, and N. Navab, "Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1342–1349, 2020.

[11] M.-A. Janvier, L.-G. Durand, M.-H. R. Cardinal, I. Renaud, B. Chayer, P. Bigras, J. De Guise, G. Soulez, and G. Cloutier, "Performance evaluation of a medical robotic 3d-ultrasound imaging system," *Medical image analysis*, vol. 12, no. 3, pp. 275–290, 2008.

[12] S. Merouche, L. Allard, E. Montagnon, G. Soulez, P. Bigras, and G. Cloutier, "A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of b-mode images," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 63, no. 1, pp. 35–46, 2015.

[13] S. Beniwal, K. Bhargava, and S. K. Kausik, "Size of distal radial and distal ulnar arteries in adults of southern rajasthan and their implications for percutaneous coronary interventions," *Indian heart journal*, vol. 66, no. 5, pp. 506–509, 2014.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[15] B. Jiang, A. Chen, S. Bharat, and M. Zheng, "Automatic ultrasound vessel segmentation with deep spatiotemporal context learning," in *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer, 2021, pp. 3–13.

[16] D. Mishra, S. Chaudhury, M. Sarkar, and A. S. Soin, "Ultrasound image segmentation: a deeply supervised network with attention to boundaries," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1637–1648, 2018.

[17] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International conference on medical image computing and computer-assisted intervention*. Springer, 1998, pp. 130–137.

[18] E. Smistad and F. Lindseth, "Real-time automatic artery segmentation, reconstruction and registration for ultrasound-guided regional anaesthesia of the femoral nerve," *IEEE Trans. Med. Imaging*, vol. 35, no. 3, pp. 752–761, 2015.

[19] E. Karami, M. S. Shehata, and A. Smith, "Adaptive polar active contour for segmentation and tracking in ultrasound videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 4, pp. 1209–1222, 2018.

[20] P. Abolmaesumi, S. E. Salcudean, W.-H. Zhu, M. R. Sirouspour, and S. P. DiMaio, "Image-guided control of a robot for medical ultrasound," *IEEE Trans. Robot. Autom.*, vol. 18, no. 1, pp. 11–23, 2002.

[21] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[22] Z. Jiang, Z. Li, M. Grimm, M. Zhou, M. Esposito, W. Wein, W. Stechele, T. Wendler, and N. Navab, "Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback," *IEEE Transactions on Industrial Electronics*, 2021.

[23] Y. Huang, W. Xiao, C. Wang, H. Liu, R. Huang, and Z. Sun, "Towards fully autonomous ultrasound scanning robot with imitation learning based on clinical protocols," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 3671–3678, 2021.

[24] A. I. Chen, M. L. Balter, T. J. Maguire, and M. L. Yarmush, "Deep learning robotic guidance for autonomous vascular access," *Nature Machine Intelligence*, vol. 2, no. 2, pp. 104–115, 2020.

[25] T. S. Pheiffer, R. C. Thompson, D. C. Rucker, A. L. Simpson, and M. I. Miga, "Model-based correction of tissue compression for tracked ultrasound in soft tissue image-guided surgery," *Ultrasound in medicine & biology*, vol. 40, no. 4, pp. 788–803, 2014.

[26] B. Ihnatsenka and A. P. Boezaart, "Ultrasound: Basic understanding and learning the language," *Int. J. Shoulder Surg.*, vol. 4, no. 3, p. 55, 2010.

[27] M.-A. Janvier, S. Merouche, L. Allard, G. Soulez, and G. Cloutier, "A 3-d ultrasound imaging robotic system to detect and quantify lower limb arterial stenoses: in vivo feasibility," *Ultrasound in medicine & biology*, vol. 40, no. 1, pp. 232–243, 2014.

[28] S. Virga, O. Zettinig, M. Esposito, K. Pfister, B. Frisch, T. Neff, N. Navab, and C. Hennersperger, "Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*. IEEE, 2016, pp. 508–513.

[29] A. Karamalis, W. Wein, T. Klein, and N. Navab, "Ultrasound confidence maps using random walks," *Med. Image Anal.*, vol. 16, no. 6, pp. 1101–1112, 2012.

[30] Z. Jiang, H. Wang, and et al., "Motion-aware robotic 3d ultrasound," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2021.

[31] Y. Bi, Z. Jiang, Y. Gao, T. Wendler, A. Karlas, and N. Navab, "Vesnet-rl: Simulation-based reinforcement learning for real-world us probe navigation," *IEEE Robotics and Automation Letters*, 2022.

[32] G. Bradski, "The openCV library." vol. 25, no. 11, pp. 120–123.

[33] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," *Multimedia Tools and Applications*, vol. 79, no. 19, pp. 12 777–12 815, 2020.

[34] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] K. Cho, B. van Merriënboer, Ç. Gulçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.

[37] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *NIPS 2014 Workshop on Deep Learning*, 2014.

[38] N. Ballas, L. Yao, C. Pal, and A. Courville, "Delving deeper into convolutional networks for learning video representations," in *Int. Conf. Learn. Represent (ICLR)*, 2016.

[39] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[40] J. Bertels, T. Eelbode, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2019, pp. 92–100.

[41] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2017, pp. 240–248.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[43] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh *et al.*, "Mixed precision training," in *International Conference on Learning Representations*, 2018.

[44] C. Hennersperger, B. Fuerst, and et al., "Towards MRI-based autonomous robotic us acquisitions: a first feasibility study," *IEEE Trans. Med. Imaging*, vol. 36, no. 2, pp. 538–548, 2016.

[45] A. Farshad, Y. Yeganeh, P. Gehlbach, and N. Navab, "Y-net: A spatiospectral dual-encoder network for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*. Springer, 2022, pp. 582–592.

[46] Z. Jiang, N. Danis, Y. Bi, M. Zhou, M. Kroenke, T. Wendler, and N. Navab, "Precise repositioning of robotic ultrasound: Improving registration-based motion compensation using ultrasound confidence optimization," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[47] Z. Jiang, Y. Gao, L. Xie, and N. Navab, "Towards autonomous atlas-based ultrasound acquisitions in presence of articulated motion," *IEEE Robotics and Automation Letters*, 2022.

[48] C. Hennersperger, M. Baust, D. Mateus, and N. Navab, "Computational sonography," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 459–466.

[49] A. L. Y. Hung and J. Galeotti, "Good and bad boundaries in ultrasound compounding: preserving anatomic boundaries while suppressing artifacts," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 11, pp. 1957–1968, 2021.

[50] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu *et al.*, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.

[51] Y. Bi, Z. Jiang, R. Clarenbach, R. Ghotbi, A. Karlas, and N. Navab, "Misegnet: Mutual information-based us segmentation for unseen domain generalization," *arXiv preprint arXiv:2303.12649*, 2023.

[52] Q. Meng, J. Matthew, V. A. Zimmer, A. Gomez, D. F. Lloyd, D. Rueckert, and B. Kainz, "Mutual information-based disentangled neural networks for classifying unseen categories in different domains: Application to fetal ultrasound imaging," *IEEE transactions on medical imaging*, vol. 40, no. 2, pp. 722–734, 2020.

[53] C. I. Bercea, B. Wiestler, D. Rueckert, and S. Albarqouni, "Federated disentangled representation learning for unsupervised brain anomaly detection," *Nature Machine Intelligence*, vol. 4, no. 8, pp. 685–695, 2022.

[54] G. Wang, Y. Yang, H. Zhang, Z. Liu, and H. Wang, "Spherical interpolated convolutional network with distance–feature density for 3-d semantic segmentation of point clouds," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13 546–13 556, 2021.

[55] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6411–6420.

Zhongliang Jiang (Member, IEEE) received the M.Eng. degree in Mechanical Engineering from the Harbin Institute of Technology, Shenzhen, China, in 2017, and Ph.D. degree in computer science from the Technical University of Munich, Munich, Germany, in 2022. From January 2017 to July 2018, he worked as a research assistant in the Shenzhen Institutes of Advanced Technology (SIAT) of the Chinese Academy of Science (CAS), Shenzhen, China. He is currently a senior research scientist at the Chair for Computer Aided Medical Procedures (CAMP) at the Technical University of Munich.

His research interests include medical robotics, robotic learning, human-robot interaction, and robotic ultrasound.

Felix Duelmer received the B.Sc. degree in Mechanical Engineering in 2018, and M.Sc. degree in Mechatronics and Robotics in 2023, from the Technical University of Munich, Munich, Germany. He is currently working toward the Ph.D. degree in Computer Science with the Technical University of Munich, Munich, Germany.

His research interests include robotic ultrasound, imaging process, robotic learning, and computer vision.

Nassir Navab (Fellow, IEEE) received the Ph.D. degree in computer and automation with INRIA, and the University of Paris XI, Paris, France, in 1993.

He is currently a Full Professor and the Director of the Laboratory for Computer-Aided Medical Procedures with the Technical University of Munich, Munich, Germany, and an adjunct professor at Johns Hopkins University, Baltimore, MD, USA. He has also secondary faculty appointments with the both affiliated Medical Schools. He enjoyed two years of a Postdoctoral Fellowship with the MIT Media Laboratory, Cambridge, MA, USA, before joining Siemens Corporate Research (SCR), Princeton, NJ, USA, in 1994.

Dr. Navab is a fellow of the Academy of Europe, MICCAI, IEEE, and Asia-Pacific Artificial Intelligence Association (AAIA). He was a Distinguished Member and was the recipient of the Siemens Inventor of the Year Award in 2001, at SCR, the SMIT Society Technology award in 2010 for the introduction of Camera Augmented Mobile C-arm and Freehand SPECT technologies, and the "10 years lasting impact award" of IEEE ISMAR in 2015. He is the author of hundreds of peer-reviewed scientific papers, with more than 54,400 citations and enjoy an h-index of 104 as of August 11, 2022. He is the author of more than thirty awarded papers including 11 at MICCAI, 5 at IPCAI, and three at IEEE ISMAR. He is the inventor of 50 granted US patents and more than 50 International ones.