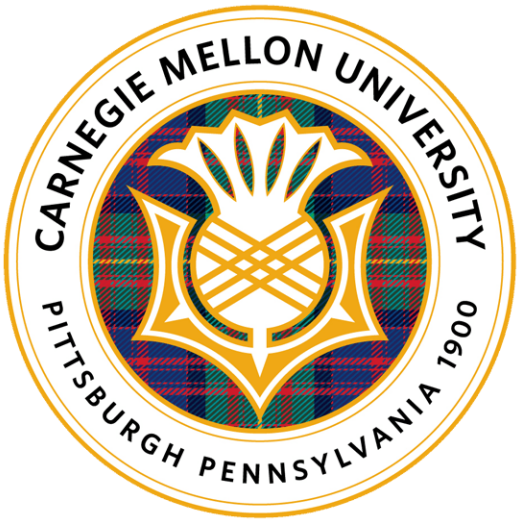
Improving Robotic Ultrasound AI using Optical Flow

FNU Abhimanyu

CMU-RI-TR-23-63

August 2023



The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Howie Choset, Chair
John Galeotti, Co-Chair
Ceci Morales

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Robotics.

Abstract

Ultrasound is an important modality for medical intervention such as vascular access because it is safe, portable, and low-cost. However, ultrasound scanning requires trained sonographers who are scarce, and it can be challenging to perform ultrasound examinations in disaster or battlefield scenarios. This motivates us to automate ultrasound scanning. One significant challenge in automating ultrasound scanning is performing the scan on a highly curved surface while simultaneously maintaining proper contact with the surface to capture high-quality images. Another major challenge while automating ultrasound scanning is that if an ultrasound probe is pressed too hard against the skin it causes significant anatomical deformations. Subsequently, these deformations present a major challenge in the generalization of tasks like segmentation and registration in ultrasound images. Therefore, in this work, I aim to improve the autonomy of a robotic ultrasound system and improve the generalizability of ultrasound imaging algorithms to work at different force values.

In the first part of the thesis, I present an enhanced strategy for ultrasound scanning using a robot with minimal expert guidance. Our methods demonstrate improved quality in the collected ultrasound images compared to existing approaches. In the subsequent section, I address the challenge of deformable registration in ultrasound images, particularly when these images are acquired at different force levels. Instead of relying on supervised learning methods, that require exhaustive and expensive ground truth calculations, I propose an unsupervised approach to estimate the displacement field between various stages of deforming anatomy in ultrasound images. Furthermore, by analyzing displacement field patterns for different compression forces, we develop an accurate physics model for ultrasound image compression. This model enables the generation of photo-realistic ultrasound images at different compression forces. Leveraging these photo-realistic images, I augment a vessel segmentation model to enhance its generalization capabilities, particularly for higher force values.

I further showcase the versatility of our method by successfully predicting accurate deformation fields for various medical ultrasound tasks, including respiratory motion cancellation and curved needle registration. The application of our approach to different medical scenarios highlights its effectiveness and broad applicability.

Acknowledgments

I would first like to thank my advisors, Professors Howie Choset and John Galeotti, for their support over my past two years at Carnegie Mellon University. I am eternally grateful for the feedback regarding my work on a weekly basis. I am also extremely grateful to the members of the Biorobotics Lab for providing me with valuable experiences, companionship, and feedback throughout my time here. I have learned a lot from my labmates which has helped me a lot in my research. I would like to specifically thank members of my research team: Tejas Zodage, Vinit Sarode, Ananya Bal, Raghavv Goel, Kirtan Patel, Nico Zevallos, Andrew Orekhov, Hans Kumar, Ed Chen, Evan Harber, Ceci and others for helping me out in my research.

Finally, I would like to thank my parents, and my sister for their endless support and their decision to send me to CMU for pursuing my dream. None of this would ever be possible without their support and blessings. I am forever indebted to their gratitude and care.

Contents

1 Introduction **1**

2 Related Works **3**
2.1 Robotic Ultrasound System (RUS) . . . 3
2.2 Force controllers for robotics . . . 3
2.3 Deformable registration for medical imaging . . . 4
2.4 Data augmentation for medical imaging . . . 5

3 Enabling Robotic Ultrasound Scanning **6**
3.1 Introduction . . . 6
3.2 Approach . . . 8
3.2.1 Experimental setup . . . 8
3.2.2 Reference trajectory generation . . . 8
3.2.3 Hybrid-force-position controller . . . 9
3.2.4 Sending commands to the robot . . . 10
3.3 Results . . . 11
3.3.1 Image quality comparison . . . 11
3.3.2 Force profile comparison . . . 12
3.4 Conclusion . . . 13

4 Unsupervised optical flow estimation in ultrasound images **14**
4.1 Introduction . . . 14
4.2 Approach . . . 16
4.2.1 U-RAFT network architecture . . . 16
4.2.2 Loss functions for unsupervised training . . . 16
4.3 Experimental setup and training details . . . 17
4.3.1 Experimental setup . . . 17
4.3.2 Training details . . . 19
4.4 Results and Discussion . . . 19
4.4.1 Deformable registration results with U-RAFT . . . 19
4.4.2 Analyzing and canceling respiratory motion with U-RAFT . . . 20
4.4.3 Registration of curved needles in ultrasound images . . . 24
4.5 Conclusion . . . 26

5 Improving segmentation in ultrasound images using optical flow **27**
5.1 Introduction 27
5.2 Methods 29
5.2.1 Synthetic data generation 29
5.3 Results 33
5.3.1 Photorealism of generated synthetic images 33
5.3.2 Improvement in segmentation using synthetic data generation . . . 36
5.4 Conclusion 38

6 Conclusion and future works **39**

Bibliography **41**

List of Figures

3.1 Robotic Ultrasound System (RUS), which consists of a 6-DoF UR3e serial manipulator and an ultrasound probe mounted to the robot end-effector . . . 7

3.2 Robotic Ultrasound System (RUS), which consists of a 6-DoF UR3e serial manipulator and an ultrasound probe mounted to the robot end-effector . . . 8

3.3 (a) Ultrasound Image from *blue-gel-phantom* with improper contact between the probe and the surface. (b) Ultrasound Image from *blue-gel-phantom* with proper contact between the probe and the surface. 11

3.4 Robotic Ultrasound System (RUS), which consists of a 6-DoF UR3e serial manipulator and an ultrasound probe mounted to the robot end-effector . . . 12

4.1 Pipeline used to generate image reconstructions for training U-RAFT in an unsupervised manner. Our proposed cyclic loss function in (4.3) improves registration quality by comparing the reconstructed reference image to the original reference image. 16

4.2 (a) The robot arm we used for capturing the ultrasound images with force-controlled scanning, together with a human tissue and vasculature phantom model we used to test U-RAFT. (b) Our U-RAFT model registers a deformed image to a reference image with RAFT, creating deformation field that is then used with a spatial transformer network (STN) to generate new deformed images, shown here with example *in-vivo* porcine images. 18

4.3 Example deformable registration result from the blue-gel dataset, showing (a) the reference image I_r, the deformed image I_d, and the reconstructed deformed image I'_d, and (b) the deformation field calculated between I_d and I_r using U-RAFT, with a displacement vector plotted on a 4x4 pixel grid. The vessel walls for the blue-gel images are manually annotated for visibility. 20

4.4 Example deformable registration result from robotic experiment on the blue-gel dataset, showing the fixed image I_f, the moving image I_m, and the reconstructed fixed image I'_f, and the deformation field calculated between I_f and I_m using U-RAFT, with a displacement vector plotted on a 4x4 pixel grid. Sub-figure (a) shows compression and (b) shows uncompression. 21

4.5 (a) Zoomed-in view of an example live-pig reconstructed deformed image I'_d using $\mathcal{L}_{\text{us}}$. (b) Zoomed-in view of the reconstructed deformed image I'_d using $\mathcal{L}_{\text{fa-cyclic-us}}$. The use of the feature-aware cyclic loss function helps remove the optical distortions observed in (a) for large deformations. 21

4.6 Live-pig results showing reference, deformed, and reconstructed deformed images for (a) small deformation, (b) medium deformation, (c) large deformation, (d) vessel collapse, and (e) an atypical case of a vessel decollapsing. The similarity between the deformed and the reconstructed deformed images shows the efficacy of U-RAFT. Registration performance drops for the case of a vessel decollapsing, but even in this atypical scenario our approach fails gracefully. 22
4.7 Fixed (I_f) and moving (I_m) image and their respective displacement fields . 23
4.8 Average pixel displacement before and after applying respiratory compensation 23
4.9 (a) Reference Image (I_r), (b) Target Image (I_t), (c) Unregistered I_r and I_t, (d) Predicted displacement field between I_r and I_t using U-RAFT (first row) and Farneback Optical Flow [1] (second row) and (e) Registered I_r and I_t. It's observed that the displacement field predicted by U-RAFT is more accurate and has fewer false positives, especially for static regions of the image. 24

5.1 PhyDNet architecture as shown in [2] 30
5.2 Physics Inspired Ultrasound Image Generator (Phy-UGen) network architecture 31
5.3 Ground truth, synthetic image generated using K_{const} model and synthetic image generated using K_{var} model. Ultrasound images are at 4N and 20N respectively 33
5.4 Comparison of SSIM values between synthetic images generated from the constant stiffness model and the variable stiffness model and the ground truth image at different force values 34
5.5 Comparison of PSNR values between synthetic images generated from the constant stiffness model and the variable stiffness model and the ground truth image at different force values 35
5.6 Comparison of IoU values between synthetic images generated from the constant stiffness model and the variable stiffness model and the ground truth image at different force values 36
5.7 Ground truth images, synthetic images generated using constant stiffness model and synthetic images generated using variable stiffness model at 2N, 6N, 10N, 14N, 18N, and 22N 37
5.8 (a) Ultrasound image at 10 N. (b) Ultrasound image with predicted segmentation mask using a U-Net model trained using real data collected at 2 N. (c) Ultrasound image with predicted segmentation mask using U-Net model train using real and synthetic data. The inclusion of synthetic data improves segmentation accuracy. 38

List of Tables

4.1 Comparison of registration error for different loss functions. $\mathcal{L}_{\text{fa-cyclic-us}}$ outperforms the other two loss functions in terms of both SSIM and F-SSIM. . 20

4.2 Average pixel displacement before and after applying respiratory compensation 21

4.3 Mean Needle Tip Error (NTE) on the *blue-gel-needle*. NTE is calculated after registering needle tips in the phantom's curved needle insertion. The method with the lowest mean NTE is highlighted. The mean NTE prior to registration was 9.44 pixels 25

4.4 Mean Needle Tip Error (NTE) on the *pig-lab-needle*. NTE is calculated after registering needle tips in the pig's curved needle insertion. The method with the lowest mean NTE is highlighted. The mean NTE prior to registration was 4.24 pixels 25

5.1 Comparing performance on the test datasets of a U-Net model with and without data augmention using the synthetic dataset. The results from the best performing dataset is highlighted. 37

Chapter 1

Introduction

Numerous medical imaging techniques, such as MRI [3], CT [4], and ultrasound [5] can be utilized to offer feedback during medical examinations. Among these techniques, ultrasound imaging offers several advantages. Unlike other methods, it does not involve ionizing radiation, ensuring patient safety. Ultrasound also requires a smaller workspace and has a lower cost compared to MRI and CT, and provides real-time images of the internal anatomy. Point-of-care ultrasound (POCUS) is particularly popular due to its affordability, portability, and widespread availability. Despite its numerous benefits, ultrasound imaging is heavily reliant on the skill of the trained sonographer. The sonographer must identify the appropriate area to scan, move the ultrasound probe within the region of interest and adjust the probe's pose to acquire diagnostic image quality and prevent patient injury. However, skilled sonographers may not be available in all locations, necessitating the development of autonomous robotic ultrasound systems.

However, automating ultrasound scanning presents significant challenges. One such challenge is scanning highly curved surfaces while maintaining physical contact between the probe and the subject to capture high-quality images. A failure in doing so leads to sub-optimal image quality and difficulty in accurately visualizing the underlying anatomy. Another major challenge in automating ultrasound scanning is the potential for excessive pressure applied by the ultrasound probe against the skin. This excessive pressure can cause significant anatomical deformations, distorting the internal structures of interest. These deformations pose a significant challenge when generalizing tasks like segmentation and registration in ultrasound images, as the appearance and location of anatomical structures are altered. To address these challenges, this thesis aims to enhance the autonomy of robotic ultrasound scanning and generalize the ultrasound imaging algorithms like deformable registration and vessel segmentation across different force values. By improving the robotic ultrasound scanning as well as accurately predicting the impact of anatomical deformations, we strive to enhance the reliability and applicability of automated ultrasound scanning in various clinical scenarios.

In the first chapter, we propose utilizing our robotic system to perform ultrasound scans on various regions of the subjects, including highly curved areas like the inguinal fold. We propose manual as well as semi-automated ways for selecting a reference trajectory based on the internal and the external anatomy of the scanning subject. To track this trajectory, we employ a Spherical Linear Interpolation (SLERP) controller, which is continuously up-

dated using force feedback readings. This approach effectively balances the approximated reference trajectory and force readings, ensuring optimal acoustic coupling during the scan. Notably, our work is the first to combine SLERP with a 6-axis hybrid force-position controller, demonstrating promising results for automating ultrasound scanning on complex surfaces.

In the second chapter, we propose an approach to predict anatomical deformations in ultrasound images under the application of different forces. Predicting anatomical deformation is a part of deformable registration, which involves the registration of a reference image to a deformed image by estimating a per-pixel deformation field. By calculating the deformation field, we can understand the nature of the anatomical changes, enabling applications such as motion compensation, image analysis, and surgical planning. We present an unsupervised deep-learning model called U-RAFT (Unsupervised Recurrent All-pairs Field Transforms) for ultrasound-to-ultrasound deformable image registration and synthetic ultrasound image generation. U-RAFT uses RAFT [6], a CNN for optical flow estimation, to register images and create a deformation field (DF). It then uses a Spatial Transformer Network (STN) [7] to generate new synthetic images as well as enable unsupervised training of RAFT. We also show the efficacy of our method for other medical imaging tasks like respiratory motion compensation in lung ultrasound images and curved needle registration in ultrasound images. The application of our approach to different medical scenarios highlights its effectiveness and broad applicability

In the third chapter, we propose an approach to use deformable registration for augmenting images for training vessel segmentation models with small training datasets. Capturing the precise dynamics is a crucial requirement when generating photo-realistic ultrasound images under the application of various forces. Previous methods, such as [8, 9] for brain MRI images, as well as [10] for lung MRI images, have employed data augmentation techniques but have not emphasized capturing accurate process dynamics. Consequently, these methods tend to generate out-of-distribution datasets that do not effectively contribute to improving vessel segmentation. Alternatively, physics-based simulators like SOFA [11] offer accurate deformable 3D physics simulations under the influence of forces. These 3D simulations can be used to deform medical phantoms/subjects and simulate anatomical deformations, but SOFA requires an accurate 3D ground truth model and accurate material properties. Thus, we propose an approach that emulates the compression process of ultrasound images without relying on explicit 3D models or material properties. Subsequently, we employ the generated ultrasound images to augment and enhance segmentation specifically at higher forces. By eliminating the need for 3D models and material properties, our method offers a more accessible and practical solution for improving vessel segmentation through image augmentation.

Chapter 2

Related Works

2.1 Robotic Ultrasound System (RUS)

Robotic Ultrasound Systems (RUS) can be categorized into teleoperated [12–14], collaborative assisting [15–17], and autonomous systems. The former, i.e. non-autonomous, systems have improved the image acquisition processes and are able to perform ultrasound scanning in remote areas, but these systems still require a human-in-the-loop (HITL).To minimize the need for human intervention, many autonomous ultrasound methods have been explored in recent years [18], [19], [20]. Huang et al.[21] demonstrated autonomous scanning of the coronal plane using an external depth sensor. Their system was shown to scan and reconstruct the non-flat patient surface. Their method plans the scanning trajectory based only on external surface features, which may not be sufficient for many interventional procedures. Hennersperger et al. [22] developed a RUS that autonomously generates trajectories based on the points selected by the physician marked in an MRI or CT scan. This system enables autonomy but is dependent on the MRI/CT scans which are expensive and may not be available at all times. Merouche et al.[23] presents an automatic vessel tracking strategy as an alternative to the teach mode, replay mode. The pipeline was applied to provide 3-D results of the lower limb arteries. The system could calculate the distance between the center of the vessel and the center element of the probe. However, the system performance would heavily depend upon the detection and tracking of the vessels. [24] provides a teleoperated RUS system with three modes for the operator: float, haptic and automatic. The "automatic" mode scans along a desired trajectory recorded in advance by human-in-loop, unlike our presently proposed method where the desired trajectory is found using Bayesian optimization. Lastly, Reinforcement Learning has been used to control the movement of the US probe as shown in [25, 26], but those methods have low success rates on new patients that were not in their training data.

2.2 Force controllers for robotics

Force controller defines relationship between force and position controller in robotics. Carriere et al.[27] uses force control to ensure compliance in a US scanning system therefore controlling the force applied to the tissue and reducing the exerted force from the

sonographer. Piwowarczyk et al.[28] proposed a method that utilizes an admittance controller to scale the force applied by the user on the robot based on the force exerted on the environment. Ferraguti et al.[29] discusses the stability of force-controlled robots and their ability to deal with different environmental forces. Li et al.[30] uses admittance control for an exoskeleton robot to create a reference trajectory based on measured force. The integration of neural networks and admittance control for robot trajectory tracking was introduced by Yang et al.[31] The method ensures trajectory tracking through the use of a neural network, while admittance control regulates torques to follow the desired trajectory. More applications of force controllers in robotics is discussed in Keemink et al.[32].

2.3 Deformable registration for medical imaging

Classical deformable registration work consisted of choosing a deformable model and then finding the optimal parameters of the model based on the optimization of the objective function. The choice of deformable model is made based on the required computational efficiency and the richness of the description. Broit [33], proposed a deformable model as an elastic membrane grid that is deformed under the influence of forces until the equilibrium is reached. An external force tries to deform the image such that matching is achieved while an internal one enforces the elastic properties of the material. Linear elastic models have also been used to register brain images based on sparse correspondences as demonstrated in Davatzikos [34]. But the use of a linear elastic model prevents in capturing large deformations accurately. Thus, Rabbitt [35], modeled large deformations as a non-linear elastic model. There are also methods that model the image deformation as a viscous fluid as shown in Christensen [36]. Wang and Staib [37], used fluid deformation models in an atlas-enhanced registration setting. Pennec et al. [38] studied image registration as an energy minimization problem and drew the connection of the Demons algorithm with gradient descent schemes. Apart from assuming a deformation model, there are some interpolation-based methods that have been used for medical image registration. Bookstein [39], proposed the use of thin-plate spline (TPS) for image registration, where TPS minimizes the bending energy. the solution to this is unique and in closed form. The global nature of TPS resulted in localization errors while estimating a dense deformation field. This led to the use of Free-form deformation models (FFDs). FFDs were first popularized in the computer graphics community [40], [41] but gained wide acceptance in the medical image analysis community when coupled with cubic-B splines [42].

Classical deformable registration methods had non-convex cost functions and were slow due to the large number of optimization parameters. To overcome these difficulties, deep-learning methods have also been presented. Balakrishnan et al.[43] designed a U-Net framework named VoxelMorph to perform DIR of brain MR images. Unlike conventional registration methods that calculate the deformation field for every pair of images, they formulate the deformation field as a global function that could be optimally parameterized with the trained convolutional neural network. Estienne et al. [44] used a shared encoder with a separate decoder named U-ResNet to compute the DVF. The network inputs paired fixed and moving images and aims to output their specific segmentation maps. Rohe et al. [45] adopted a U-Net-like network to predict the deformation field for 3D cardiac MRI

volume registration. Sokooti et al. [46] presented a multi-scale network to learn a deformation field of intra-subject 3D chest CT registration. They used random DVF as supervision. Uzunova et al. [47] designed a network for the registration between 2D brain MRI and 2D cardiac MRI. Their ground truth is generated utilizing statistical appearance models (SAMs). They adapted FlowNet [48] architecture and obtained outperforming results. Although we have a lot of deformable registration methods for modalities such as CT-to-MRI or CT-to-ultrasound, ultrasound-to-ultrasound deformable registration is still an unsolved challenge due to noise, speckle, shadows, and mirror image artifacts [49].

2.4 Data augmentation for medical imaging

The rapid advancements in deep learning can be attributed to the abundance of new data available for training models. But still, collecting large datasets for medical imaging is still a challenge due to privacy concerns and labeling costs. Data augmentation makes it possible to greatly expand the amount and variety of data available for training without actually collecting new samples. One of the most common forms of data augmentation is performing the affine transformation. Frameworks like Medical Open Network for AI (MONAI) [50], are developed for performing such transformations. However, affine transformation applies a constant transformation across all the pixels in the image, there are pixel-level transformations such as pixel erasing [51]. Also, Generative models have also been used to generate realistic images for medical imaging [52],[53]. However, GANs are always prone to mode collapse. Finally, there are a variety of augmentation methods that are governed by physical or biological models to generate new images or modify existing ones. For instance, these techniques have been used to simulate multiple sclerosis lesions in brain MR images [54] or to add cancer signs to breast mammography images [55]. Though these physics-inspired models are present for MR and CT images, very little work has been done in the ultrasound image domain.

Chapter 3

Enabling Robotic Ultrasound Scanning

Chapter 3 is adapted from the publication:
Raghavv Goel*, **FNU Abhimanyu***, Kirtan Patel, John Galeotti, Howie Choset, "Autonomous Ultrasound Scanning using Bayesian Optimization and Hybrid Force Control," International Conference of Robotics and Automation (ICRA), 2022
FNU Abhimanyu's contributions as the first author of the publication include: conducting an initial literature review, developing the algorithm and experiments, conducting the experiments, analyzing the data, writing the manuscripts, and responding to reviewers with revisions. Raghavv Goel and Kirtan Patel assisted with providing experimental suggestions and running actual experiments. Howie Choset and John Galeotti are the supervising faculty advisors.

3.1 Introduction

Ultrasound has emerged as a crucial medical imaging method for diagnostics and medical intervention procedures, owing to its real-time feedback, portability, and lack of radiation exposure. Despite its numerous benefits, ultrasound imaging is heavily reliant on the skill of the trained sonographer. The sonographer must identify the appropriate area to scan, move the ultrasound probe within the region of interest, adjust the probe's pose, and apply safe, accurate forces through the probe to maintain diagnostic image quality and prevent patient injury. However, skilled sonographers may not be available in all locations, necessitating the development of autonomous ultrasound systems.

The Robotic Ultrasound System (RUS) is a robotic system with an attached ultrasound station and scanning probe, as shown in Figure 3.1. By integrating robotic technology into ultrasound imaging, RUS offers improved accuracy, stability, repeatability, and maneuverability in image acquisition. Recent research has focused on improving the autonomy of RUS, but most existing systems still require human intervention to navigate the ultrasound probe to the region of interest. The development of truly autonomous RUS could eliminate the need for human operators, thereby increasing accessibility to high-quality ultrasound imaging in various scenarios. We propose an autonomous 6-DoF RUS for ultrasound scanning which employs a SLERP [56] + force feedback controller and leverages input from an RGB-D sensor. In this work, we propose to use this system to scan different parts of the

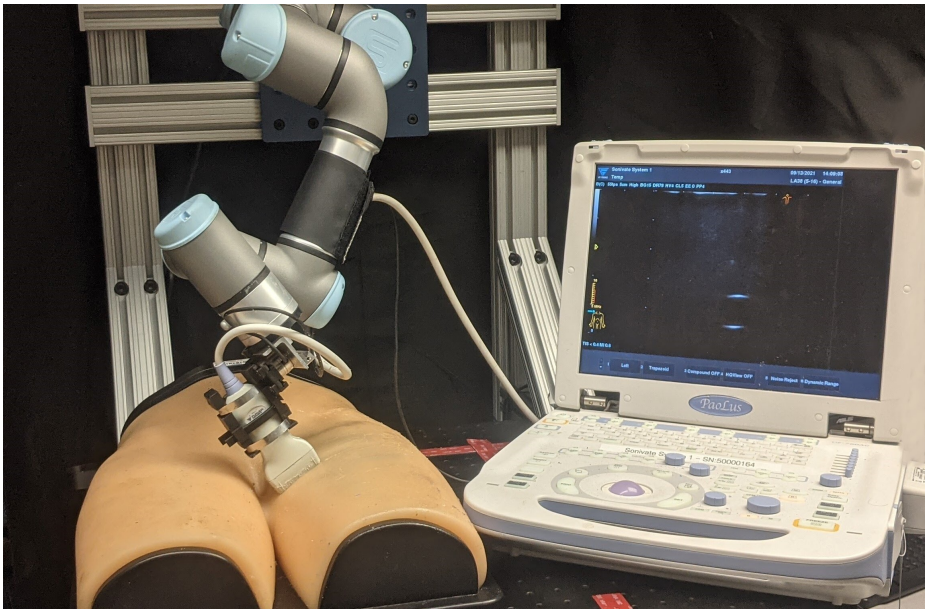


Figure 3.1: Robotic Ultrasound System (RUS), which consists of a 6-DoF UR3e serial manipulator and an ultrasound probe mounted to the robot end-effector

subjects including highly curved regions like the inguinal fold. For scanning, we start by generating the reference trajectory using 3 different methods. Once we have the initial and the final point for ultrasound scanning chosen by the user, we use the surface normals to calculate the initial and the goal pose in order to create a reference trajectory. To track this trajectory, we employ a Spherical Linear Interpolation (SLERP) controller, that is updated at every instance using noisy force feedback readings. This approach strikes a balance between the approximated reference trajectory and force readings, ensuring proper acoustic coupling during the scan. Notably, this is the first work to use SLERP to generate a reference trajectory and combine it with a 6-axis hybrid force-position controller. Our proposed approach shows promising results in automating ultrasound scanning on complex surfaces.

3.2 Approach

In this section, we talk about the experimental setup, the reference trajectory generation, and then the hybrid-force-position controller.

3.2.1 Experimental setup

Our RUS, as shown in Figure 3.1, consists of a PaoLus UF-760AG Portable Diagnostic Ultrasound Imaging Equipment (FUKUDA DENSHI,UK) using a 5-12 MHz 2D linear transducer mounted on the 6-DoF Universal Robot UR3e robot. Our robotic system also has an Intel Realsense D-435i [57] RGB-D sensor to capture the surface normals of the subject as well as a 6-axis ATI force sensor. We use Python to implement the controller and the kinematics of the robot. Furthermore, Robot Operating System (ROS) [58] was used to communicate with different sensors and the robot.

3.2.2 Reference trajectory generation

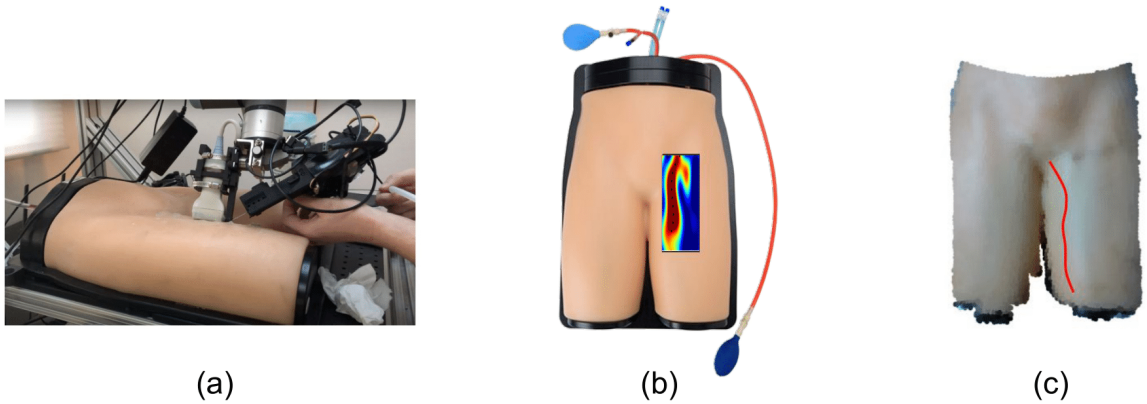


Figure 3.2: Robotic Ultrasound System (RUS), which consists of a 6-DoF UR3e serial manipulator and an ultrasound probe mounted to the robot end-effector

As our first step, we need to generate a reference trajectory on the subject's surface and for the ultrasound probe to follow. We developed three different methods to generate three different methods for generating the reference trajectory.

Our first method involves a manual mode, where a user manually positions the end-effector of the robot at the desired location. To facilitate this, we utilize the $freedrive$ mode of the UR3e robot, which allows unrestricted movement of the manipulator. By enabling the robot's $freedrive$ mode, the user can intuitively control and navigate the robot to the desired positions as shown in Figure 3.2-a. Subsequently, the user selects and saves the goal poses for the final scan, ensuring accurate and targeted data acquisition. More details on this can be found in [59].

Our second method focuses on determining the reference trajectory based on the precise location of the vessels within the subject. The underlying rationale behind this approach is to ensure that the robot's scanning trajectory effectively covers and scans the most densely

populated regions where vessels are located. To accomplish this, we palpate at multiple locations on the subject, which allows us to create an informative map as shown in Figure 3.2-b, indicating the positions of the vessels. We then utilize this vessel information map to generate a reference trajectory for the robot's scanning motion. In order to minimize the number of palpations, we use Bayesian optimization [60] to decide the next best location for palpation. By incorporating Bayesian optimization and utilizing the vessel information map, this method enables a higher degree of autonomy compared to the previous manual method. More details about this method can be found in [61].

The third method involves finding the reference trajectory based on the anatomical landmarks on the surface of the subject. In this method, we have an additional RGB-D sensor mounted on the end-effector of the robot which helps us in retrieving a point cloud of the subject's surface. After the scan, the generated stitched point cloud is processed with outlier removal using the DBSCAN clustering algorithm [62]. The resultant point cloud is then used for surface normal computation. Utilizing the point cloud of the subject's surface as a reference, we identify and mark important anatomical landmarks such as a knee or an inguinal fold. These landmarks serve as key points for generating the reference trajectory. By strategically selecting these landmarks as shown in Figure 3.2-c, we can ensure that the robot's scanning trajectory covers the desired regions of interest. More details can be found in [63].

In all these above-listed methods, we store the trajectory as an array of intermediate poses $P_0, P_1, ..., P_T$.

3.2.3 Hybrid-force-position controller

SLERP-based force controller

Based on the reference trajectory presented in the last section, SLERP based controller is used to track the orientations in the reference trajectory by computing the intermediate orientations to follow using

$$r_t = \frac{d(p_t, p_T)}{d(p_0, p_T)} \tag{3.1}$$

$$q_t = \frac{q_0 \sin((1 - r_t)\theta_t) + q_T \sin(r_t \theta_t)}{\sin(\theta_t)} \tag{3.2}$$

where $\theta_t = 2\cos^{-1}(q_0 \cdot q_T)$, $d(.,.)$ is the Euclidean distance, $P_0 = \{p_0, q_0\}$, $P_t = \{p_t, q_t\}$ and $P_T = \{p_T, q_T\}$ are the starting, instantaneous and final pose of the ultrasound probe. $p_t = [p_t^x, p_t^y, p_t^z]$ is the position, and $q_t = [q_t^w, q_t^x, q_t^y, q_t^z]$ is the quaternion with respect to the base frame of the robot.

After computing the quaternions $[q_0, \dots q_{t-1}, q_t, \dots q_T]$ the angular velocity is calculated in the end-effector frame as

$$\delta q_t = q_t^{-1} q_{t+1} \tag{3.3}$$

$$w_{slerp,t} = quat2euler(\delta q_t)/\Delta t \tag{3.4}$$

The 3 position/linear axis in the end-effector frame of the robot is controlled using the position controller mentioned in [64].

Force-based hybrid controller

In this controller, force feedback is used along all axis to control the angle made by the probe on the unknown surface to be normal to the surface, and to control the force exerted along the length of probe (*y-axis* in probe frame) to maintain a desired contact in the y-axis of the probe while scanning. This is done by updating the input linear (v_y) and angular velocities ($w_{\text{FF},x}, w_{\text{FF},z}$) to the robot end-effector as

$$v_{y,t} = -\, D_y(f_{y,t} - f_{y,d}) \tag{3.5}$$

$$\omega_{\text{FF},x,t} = -\, D_x(f_{x,t} - f_{x,d}) \tag{3.6}$$

$$\omega_{\text{FF},z,t} = -\, D_z(f_{z,t} - f_{z,d}) \tag{3.7}$$

where, the subscript 'FF' implies force-feedback. D_y, D_x and D_z are positive scalar controller gains. To match the surface normal, the desired forces are set as $f_{x,d} = 0N$, $f_{z,d} = 0N$ and $f_{y,d} = -8N$.

Combining the controllers

SLERP-based controller is good at following a given orientation trajectory but fails to adjust for abrupt changes in the orientation. Force-based orientation controller is good at adjusting for any on-the-fly change but can be noisy. To counter the cons of each method, we combine both of them using a Kalman filter-like approach [65], where the SLERP-based controller is part of a trivial process model while the force feedback is part of the measurement model based on Eq. (3.6), Eq. (3.7). The innovation/measurement residual is given as

$$\tilde{\omega}_t = \omega_{\text{FF},t} - \omega_{slerp,t} \tag{3.8}$$

The optimal gain is given as $K_t = (I_3 + R)^{-1}$, where $R \in \mathbb{R}^3$ is the covariance of the observation noise of the force sensor and I_3 is the identity matrix. Using Eq. (3.8) and K_t, we get the the updated angular velocity as

$$\omega_t = \omega_{slerp,t} + K_t\tilde{\omega}_t \tag{3.9}$$

Note that the gain K_t can also be kept fixed. Experiments were tried by keeping both a fixed and varying gain. Based on the assumptions mentioned earlier angular velocity along *y-axis* is 0.

3.2.4 Sending commands to the robot

1. In Section 4.2, we calculate the velocity of the ultrasound (us) tip. Then in order to command the robot, we need to calculate the velocity in the robot tool frame ($tool$). We calculate the velocity in the robot tool frame as,

$$\begin{bmatrix} \omega_{tool,t} \\ v_{tool,t} \end{bmatrix} = {}^{tool}\, Adj_{us} \begin{bmatrix} \omega_{us,t} \\ v_{us,t} \end{bmatrix} \tag{3.10}$$

2. Before commanding the desired velocity to the robot, we run a safety check to avoid singularity due to the limited workspace of the UR3e robot. We do this by calculating the determinant of the Jacobian matrix in that particular joint state of the robot.

3.3 Results

In this section, we assess the quality of robotic ultrasound controllers based on the collected ultrasound image quality and the applied force during scanning.

3.3.1 Image quality comparison

The quality of ultrasound images is directly proportional to the controllers' efficacy for scanning. Thus we quantify the quality of our controllers using the quality of the collected ultrasound images. We estimate the quality of ultrasound images using the zero-normal cross-correlation (ZNCC) value.

Zero-normalized cross-correlation (ZNCC) is a similarity measure commonly used in image processing and computer vision tasks [66]. It quantifies the similarity between the reference and target image by normalizing their cross-correlation coefficient with respect to their individual variances. The ZNCC equation is given by:

$$\text{ZNCC}(I_{ref}, I_{tar}) = \frac{\sum_{x,y}(I_{ref}(x,y) - \bar{I}_{ref})(I_{tar}(x,y) - \bar{I}_{tar})}{\sqrt{\sum_{x,y}(I_{ref}(x,y) - \bar{I}_{ref})^2 \sum_{x,y}(I_{tar}(x,y) - \bar{I}_{tar})^2}} \tag{3.11}$$

where I_{ref} and I_{tar} represent the reference and target image and $I(x, y)$ is the pixel values of the image at location x and y. $\bar{I}_{ref}$ and $\bar{I}_{tar}$ denote the mean pixel value for reference and target image respectively.

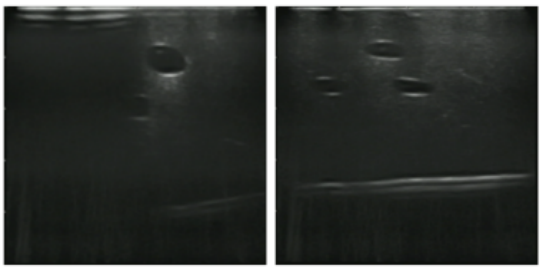


Figure 3.3: (a) Ultrasound Image from *blue-gel-phantom* with improper contact between the probe and the surface. (b) Ultrasound Image from *blue-gel-phantom* with proper contact between the probe and the surface.

We carry out experiments on the bluegel and the leg phantom, and compare the quality of the scans obtained using the following controllers: only SLERP, only FF, and combined SLERP + FF (ours). For the given trajectory $[P_0, \dots P_t, \dots P_T]$, we get a ZNCC score of 0.935, 0.933, 0.953 and 0.944 for SLERP, force-feedback (FF), combined SLERP+FF ($K_t = 0.5$) and combined SLERP+FF (varying K_t) respectively, where K_t is from Eq. (3.9). The upper bound of this metric is 0.993 when the robotic arm is locked in a fixed location with constant force (thus making good contact with the unknown surface as no sweeping is being performed which can shake the probe off the normal surface) and the

lower bound is 0.795 when the robotic arm force constraint is turned off thus leading to jittery contact with the surface. Qualitatively, Figure 3.3a and b, shows the ultrasound image collected without our combined controller and with the combined controller. Without the combined controller, we see unwanted artifacts in the ultrasound images.

3.3.2 Force profile comparison

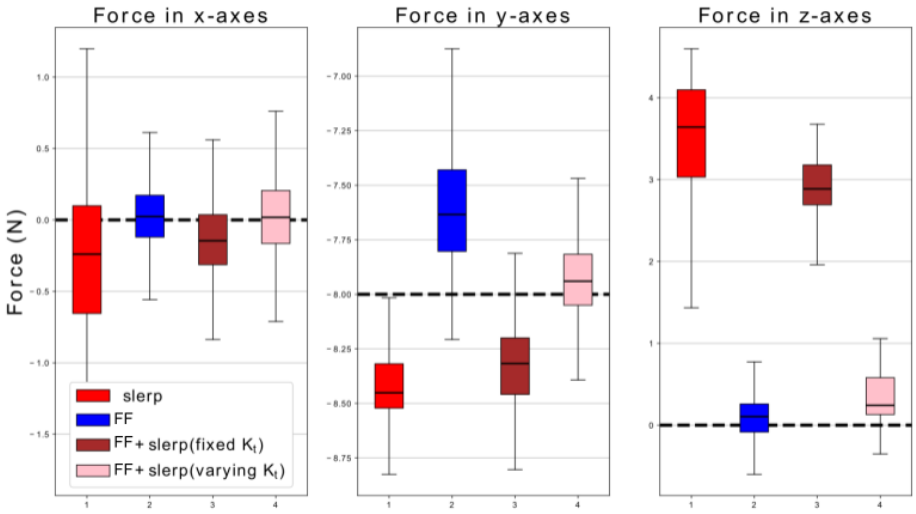


Figure 3.4: Robotic Ultrasound System (RUS), which consists of a 6-DoF UR3e serial manipulator and an ultrasound probe mounted to the robot end-effector

Besides the image quality, the force profile in the x, y, and the z axes helps us compare our controllers. For the purpose of our experiments $f_{x,d} = 0N$, $f_{y,d} = -8N$ and $f_{z,d} = 0N$. We run the ultrasound scanning controller and record the force values measured by the 6-axis force sensor. We show the measured force in each of the axes as box plots in Figure 3.4. As shown in the figure, SLERP + FF (variable α) performs the best as the obtained force along the desired axis (Y) is the closest to the desired force while the forces along the other two axes (X and Z) are close to zero which is desired. The variance in force values is also the lowest with the SLERP + FF controller.

The results show that using our proposed controllers can help us capture high-quality ultrasound images in comparison to the staple position controller used in [67] and [68]. Moreover, the combined controller helps us further improve the scanning process by compensating for the noisy reading by the force sensor as well as the surface normal estimation.

3.4 Conclusion

This chapter presents a solution for scanning using a Robotic Ultrasound System (RUS). I present three different trajectory generation method for RUS: 1) Manual reference trajectory selection by the user 2) Using Bayesian Optimization to estimate the reference trajectory based on internal anatomy and 3) Using RGB-D sensor to select a reference trajectory based on important anatomical landmarks on the subjects skin. Then I discuss the hybrid force position controller and the formulation of it. In the result section, I show the efficacy of the proposed controller proposed controller by showing the quality of collected ultrasound images and the force-profile recorded.

Chapter 4

Unsupervised optical flow estimation in ultrasound images

Chapter 4 is adapted from the manuscript:
FNU Abhimanyu, Andrew L. Orekhov, Ananya Bal, John Galeotti, Howie Choset, "Unsupervised Deformable Ultrasound Image Registration and Its Application for Vessel Segmentation"
FNU Abhimanyu's contributions to the manuscript include: conducting initial literature review, developing the algorithm and experiments, conducting the experiments, analyzing the data and writing the manuscripts. Andrew L. Okhelov and Ananya Bal assisted with providing experiments and writing the manuscript. Howie Choset and John Galeotti are the supervising faculty advisors.

4.1 Introduction

Ultrasound imaging is an important modality for vascular access because it is safe, portable, and low-cost. However, the ultrasound probe needs to press against the skin to maintain acoustic coupling contact while capturing images, causing anatomical deformations during the scan. These deformations present a challenge for training deep neural networks that segment vessels from ultrasound images.

In this chapter, we propose an approach to use deformable registration for augmenting images for training vessel segmentation models with small training datasets. Deformable registration is the problem of how to register pairs of images, one referred to as the *reference image* and the other as the *deformed image*, where the two images are of the same anatomy, but exhibit different deformations. By registering images captured at different forces, we will show that we can generate synthetic images at intermediate forces, ensuring that the augmented images are physically realistic. A similar idea was used in [8, 9] for brain MRI images, and in [10] with lung MRI images.

Previous approaches to deformable registration include hand-crafted, iterative nonlinear optimization methods with a variety of cost function definitions and parameterizations of deformations [49, 69, 70]. These methods, however, typically have non-convex cost functions and are slow due to the large number of optimization parameters. To overcome these

difficulties, deep-learning methods have also been presented [71–73], but work on deep-learning methods for ultrasound-to-ultrasound registration is limited. Compared to other higher-quality imaging modalities studied in most deformable registration work (e.g. CT-to-MRI or CT-to-ultrasound), ultrasound-to-ultrasound deformable registration is uniquely challenging due to noise, speckle, shadows, and mirror image artifacts [49]. Despite this, we will show that this problem is amenable to a deep-learning approach using our proposed model.

In this chapter, we present a deep-learning model called U-RAFT (Unsupervised Recurrent All-pairs Field Transforms) for ultrasound-to-ultrasound deformable image registration and synthetic ultrasound image generation. As shown in Fig. 4.2b, U-RAFT uses RAFT [6], a CNN for optical flow estimation, to register images and create a deformation field (DF). It then uses a Spatial Transformer Network (STN) [7] to generate new synthetic images. This approach allows for unsupervised training of U-RAFT, which we use to apply RAFT to ultrasound images for the first time, as well as its application for cancelling respiratory motion in the lung dataset as well as registering needle during insertion in the ultrasound images.

Compared to prior work, our work is unique in that it tackles ultrasound-to-ultrasound deformable registration, we are able to register images at a rate suitable for online use (~ 33 Hz), and our training is unsupervised. We note that there are multiple utilities of our deformable registration approach like data augmentation vessel segmentation, longitudinal studies/diagnosis, population studies, and intra-operative registration to anatomy [49, 73].

In the first section, we present the network architecture of U-RAFT and describe three loss functions we considered for training this network in an unsupervised manner. We present experimental results using U-RAFT on a benchtop silicone phantom model as well as *in-vivo* porcine images of femoral arteries and veins. We compare the registration quality among the three loss functions we defined. We also show the application of U-RAFT model for compensating respiratory motion in the lung dataset and also the registration of curved needles in ultrasound images.

4.2 Approach

This section discusses the network architecture used for predicting the deformation field (DF) and the loss function used to train this network in an unsupervised manner. Furthermore, we discuss the use of the DF to generate new synthetic ultrasound images and their use to improve vessel segmentation.

4.2.1 U-RAFT network architecture

Let I_r and I_d be the reference and deformed ultrasound images, respectively, collected at forces $F_r \in \mathbb{R}$ and $F_d \in \mathbb{R}$. We denote a DF as $u_{dr} = g_\theta(I_d, I_r)$, where g_θ is the function we seek to model with our network and the subscript θ denotes the network parameters used. Here, we use the state-of-art RAFT network[6] to model $g_\theta(I_d, I_r)$. We chose RAFT over other CNN-based networks like FlowNet [48], FlowNet2[74], and PWC-Net [75] because of its superior performance on the Sintel[76] and KITTI[77] datasets.

RAFT has been shown to outperform other optical flow methods in the RGB domain [48, 74, 75], but no prior work has shown the application of RAFT on medical ultrasound images, which are inherently noisier than RGB images[49]. RAFT is also a supervised method that needs a ground truth displacement field for training. Acquiring ground truth for ultrasound images is time-consuming and labor-intensive, so we seek to make the training unsupervised. We do this by passing the output of RAFT through a Spatial Transformer Network (STN) [7] to generate a reconstructed deformed image $I'_d = \text{STN}(u_{dr}, I_r)$. This enables us to incorporate the similarity of I'_d and I_d in our training loss function, which, as we will show below, provides improved registration performance. We refer to the RAFT architecture together with STN as U-RAFT.

4.2.2 Loss functions for unsupervised training

We now define three different choices of loss functions to train the U-RAFT network in an unsupervised manner. We discuss the formulation and the advantages/disadvantages of each and perform a quantitative comparison between them in the results section. The first

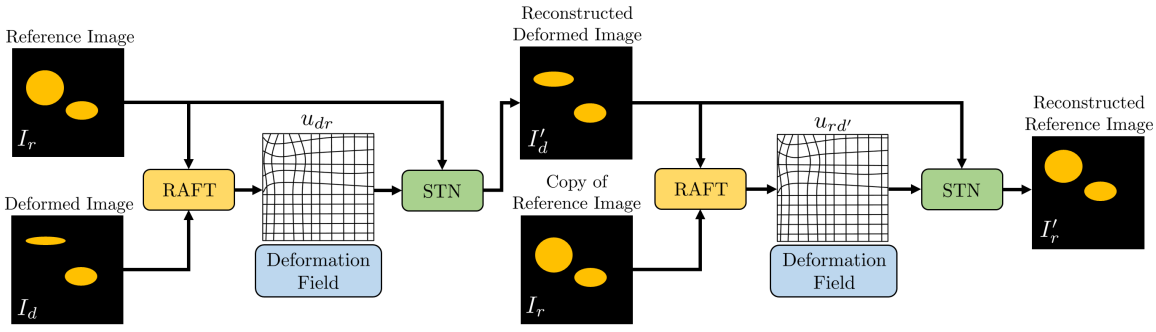


Figure 4.1: Pipeline used to generate image reconstructions for training U-RAFT in an unsupervised manner. Our proposed cyclic loss function in (4.3) improves registration quality by comparing the reconstructed reference image to the original reference image.

loss function we consider, denoted as $\mathcal{L}_{us}$, consists of two parts $\mathcal{L}_{\text{ssim}}$, a multi-scale structural similarity (SSIM) loss that penalizes the differences in appearance between I_d and I'_d, and $\mathcal{L}_{\text{smooth}}$, which penalizes abrupt changes in the neighboring pixels of I'_d (generated from u_{dr}):

$$\mathcal{L}_{\text{us}}(I_d, I'_d, u_{dr}) = \beta\mathcal{L}_{\text{ssim}}(I_d, I'_d) + (1-\beta)\mathcal{L}_{\text{smooth}}(u_{dr}) \tag{4.1}$$

where $\beta \in \mathbb{R}$ is a parameter to adjust the relative weight of $\mathcal{L}_{\text{ssim}}$ and $\mathcal{L}_{\text{smooth}}$. $\mathcal{L}_{\text{ssim}}$ and $\mathcal{L}_{\text{smooth}}$ are given by:

$$\begin{aligned}\mathcal{L}_{\text{ssim}}(I_d, I'_d) &= 1 - \text{SSIM}(I_d, I'_d) \\ \mathcal{L}_{\text{smooth}}(u_{dr}) &= \text{mean}_{xy}\left(\frac{\nabla^2 u_{dr}(x,y)}{\nabla x^2} + \frac{\nabla^2 u_{dr}(x,y)}{\nabla y^2}\right)\end{aligned} \tag{4.2}$$

where x, y are the pixel location of a 2D-deformation field, and mean_{xy} denotes the mean over all pixels.

The second loss function we consider is a cyclic version of (5.4), denoted as as $\mathcal{L}_{\text{us-cyclic}}$. In this loss function, we register the reference image I_r to the reconstructed deformed image I'_d to generate a new reconstructed reference image $I'_r = \text{STN}(g_\theta(I_r, I'_d), I'_d)$, as shown in Fig. 4.1. We then add to the loss function in (5.4) an additional term that calculates $\mathcal{L}_{\text{us}}$ for I_r, I'_r, and $u_{rd'}$.

$$\mathcal{L}_{\text{us-cyclic}} = \mathcal{L}_{\text{us}}(I_d, I'_d, u_{dr}) + \mathcal{L}_{\text{us}}(I_r, I'_r, u_{rd'}) \tag{4.3}$$

Finally, the third loss function we consider is designed to improve flow prediction in the vicinity of important anatomical features like veins, arteries, etc. We denote this feature-aware, cyclic, multi-scale SSIM loss function as $\mathcal{L}_{\text{fa-cyclic-us}}$. We use the scale-invariant feature transform (SIFT) algorithm [78], as implemented in OpenCV [79], to extract keypoints in an ultrasound image and construct a binary feature map around those keypoints. We then multiply each image by its binary feature map to create $\widetilde{I}_d$, $\widetilde{I}'_d$, $\widetilde{I}_r$, $\widetilde{I}'_r$ for the deformed, reconstructed deformed, reference, and reconstructed reference images, respectively. We then calculate the loss using the cyclic loss in (4.3) but with the images with features extracted:

$$\mathcal{L}_{\text{fa-cyclic-us}} = \mathcal{L}_{\text{us}}(\widetilde{I}_d, \widetilde{I}'_d, u_{dr}) + \mathcal{L}_{\text{us}}(\widetilde{I}_r, \widetilde{I}'_r, u_{rd'}) \tag{4.4}$$

4.3 Experimental setup and training details

4.3.1 Experimental setup

We validate our method for three different applications: 1. Predicting deformation in vessels under the application of force, 2. Respiratory motion compensation in lung ultrasound images and 3. Curved needle registration during insertion in ultrasound images. This section discusses the ultrasound dataset collection for each of these applications.

Predicting deformation in vessels under the influence of force We collected ultrasound images from two different subjects: a human tissue/vasculature gel phantom model (CAE Blue Phantom), which we refer to as the *blue-gel-vascular*, and two different live pigs,

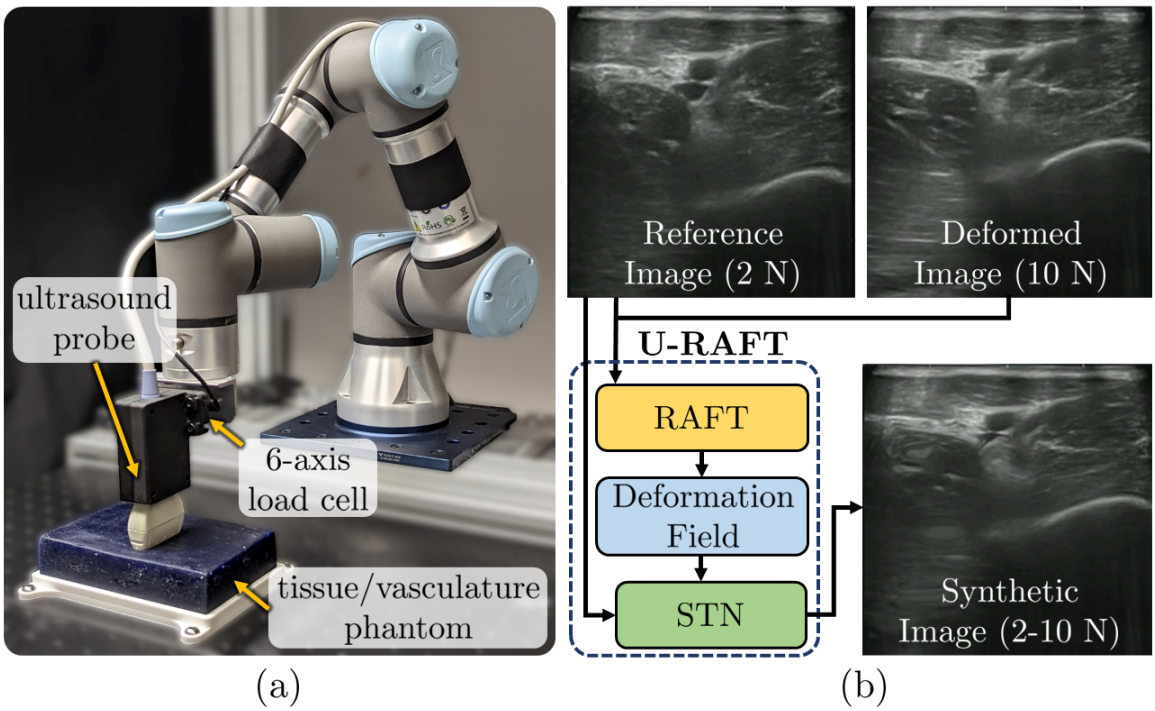


Figure 4.2: (a) The robot arm we used for capturing the ultrasound images with force-controlled scanning, together with a human tissue and vasculature phantom model we used to test U-RAFT. (b) Our U-RAFT model registers a deformed image to a reference image with RAFT, creating deformation field that is then used with a spatial transformer network (STN) to generate new deformed images, shown here with example *in-vivo* porcine images.

which we refer to as the *live-pig-vascular*. The IACUC-approved *in-vivo* porcine experiments were done in a controlled lab setting under the supervision of clinicians. The data was collected using a robotic ultrasound system which includes a UR3e manipulator (Universal Robots) with a PaoLus UF-760AG Portable Diagnostic Ultrasound Imaging Equipment (FUKUDA DENSHI,UK) using a 5-12 MHz 2D linear transducer and a six-axis force/torque sensor (ATI) mounted on the end effector, as shown in Fig. 4.2. The datasets were collected either in a "scanning mode", where the robot scanned between two predefined points on the surface of the subject with a hybrid force motion controller similar to the controller described in [61], or in "palpation mode", where the robot was commanded with a sinusoidal force profile at a single point on the skin surface. The minimum and maximum force used for both the modes were 2 N and 10 N, respectively. We collected 715 images per force for the *blue-gel-vascular* and 1200 ultrasound images for the *pig-lab-vascular*.

Respiratory motion compensation in lung ultrasound images We collected ultrasound video sequences over 3 different live pigs where each video sequences had 1000 images each. For data collection, we use the same setup as shown in 4.2, with the probe kept static near the lung of the live pig.

Curved needle registration in ultrasound images The data was collected using a Fukuda Denshi ultrasound and a linear transducer with a 51mm scanning width. The ultrasound

probe was held by hand, and all images were acquired from a human tissue/vasculature gel phantom model which we refer to as the *blue-gel-needle*. Four 2.5-inch echogenic needles were used, three of which were pre-bent by differing amounts. Each needle was inserted for 6 trials with an insertion angle between $30\,\mathrm{deg}$ to $60\,\mathrm{deg}$. Each insertion sequence had 50-80 frames. There were 2 test images for small curvature, 2 test images for medium curvature, and 3 test images for large curvature. We asked two graders to contour the boundary of the needle shaft in the images. We did in-vivo robotic needle insertion experiments into the legs of live-pig, and the illustration of the robotic mechanism is in Figure 3.1. The needles were inserted into the muscle or toward the femoral artery. The data was collected using the same ultrasound machine and transducer, with an imaging depth of 5cm. 527 test images, with 177 labeled images for evaluation which we refer to as the *live-pig-needle*.

4.3.2 Training details

We believe that finding the optimal displacement field is primarily influenced by the interaction between the tissue and ultrasound, rather than being dependent on artifacts present in the ultrasound image. This perspective allows our model to generalize effectively on larger datasets. In order to achieve this, we train our model exclusively on the vascular dataset from each subject. Additionally, we employ the model trained on the vascular dataset to address other applications such as respiratory motion compensation and needle registration. For training purposes, we utilize different datasets with varying sizes. Specifically, we train the model on 600 images and test it on 115 images from the *blue-gel-vascular* dataset. For the *live-pig-vascular* dataset, we employ 900 training images and 300 testing images. For all these datasets, the RAFT weights were initialized with pre-trained KITTI weights[77] and were trained for 150 additional epochs. The implementation is highly parallelized and performs full-batch gradient descent using the Stochastic Gradient Descent [80] optimizer in the Pytorch Autograd library[81], with a batch size of 12 with a learning rate of 0.0001.

4.4 Results and Discussion

4.4.1 Deformable registration results with U-RAFT

In this section, we evaluate U-RAFT's performance on registering ultrasound images from the blue-gel and live-pig datasets using the three loss functions described in Section 4.2-B. We use the image similarity between the original deformed image I_d and the reconstructed deformed image I'_d to measure the efficacy of our method. Figure 4.3 shows an example of reference and deformed images from the blue-gel and the live-pig datasets along with the reconstructed reference images from the U-RAFT model. We use two metrics to compare the different loss functions: 1) SSIM[82] and 2) a feature-aware-SSIM (F-SSIM), which is SSIM applied to the images after using SIFT to extract features as described in Section 4.2-B.

The results are summarized in Table 4.1. In both SSIM and F-SSIM, the cyclic loss function $\mathcal{L}_{\text{cyclic-us}}$ outperforms $\mathcal{L}_{\text{us}}$, and the feature-aware cyclic loss function $\mathcal{L}_{\text{fa-cyclic-us}}$

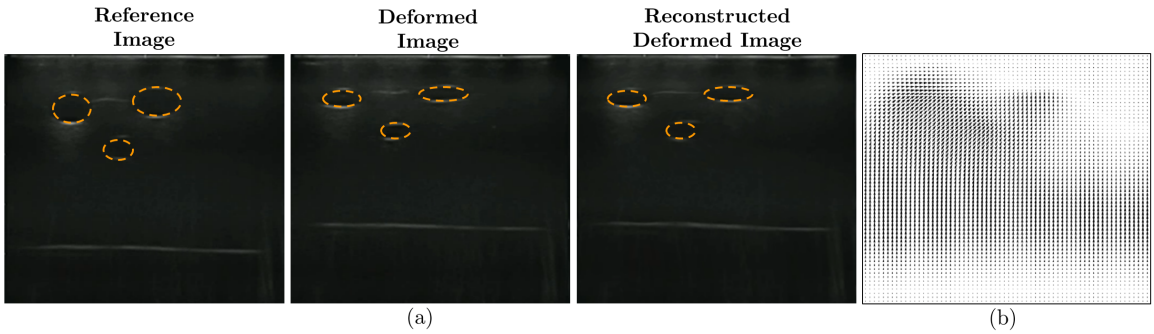


Figure 4.3: Example deformable registration result from the blue-gel dataset, showing (a) the reference image I_r, the deformed image I_d, and the reconstructed deformed image I'_d, and (b) the deformation field calculated between I_d and I_r using U-RAFT, with a displacement vector plotted on a 4x4 pixel grid. The vessel walls for the blue-gel images are manually annotated for visibility.

Table 4.1: Comparison of registration error for different loss functions. $\mathcal{L}_{\text{fa-cyclic-us}}$ outperforms the other two loss functions in terms of both SSIM and F-SSIM.

Loss Function	**SSIM (blue-gel)**	**F-SSIM (blue-gel)**	**SSIM (live-pig)**	**F-SSIM (live-pig)**
$\mathcal{L}_{\text{us}}$	0.905	0.966	0.870	0.918
$\mathcal{L}_{\text{cyclic-us}}$	0.907	0.967	0.883	0.927
$\mathcal{L}_{\text{fa-cyclic-us}}$	0.909	0.969	0.886	0.931

outperforms the cyclic loss function $\mathcal{L}_{\text{cyclic-us}}$. The cyclic function outperforms the multi-scale structural similarity as it adds a better regularization of the flow prediction. We have also observed qualitatively, as shown in the example in Fig. 4.5, that the combination of feature extraction and cyclic loss leads to improved registration, particularly for larger deformations.

4.4.2 Analyzing and canceling respiratory motion with U-RAFT

Let I_f and I_m be the fixed and moving ultrasound images, respectively from the same respiratory cycle. We denote a displacement field (DF) as $u_{fm} = g_\theta(I_f, I_m)$, where g_θ is the function we seek to model with our network and the subscript θ denotes RAFT's [6] network parameters. We use U-RAFT to generate a reconstructed fixed image $I'_f = \text{STN}(u_{fm}, I_m)$.

We use U-RAFT to track the displacement of every pixel in the video sequence and therefore capture the periodic movement of the tissue during respiratory motion. The first frame in the video sequence is I_f and the rest of the frames are the moving images I_m. For the frame i in the video sequence, we calculate the net displacement of the pixel at location x^*, y^* as, $d_{x^*,y^*,i} = ||u_{f,m_i}(x^*, y^*)||$, where $||.||$ is the L_2 norm. We then pass the sequence of moving images I_{m_i} to STN along with u_{f,m_i} to retrieve the static sequence of images $I'_{f_i} = STN(u_{f,m_i}, I_{m_i})$. Furthermore, we compute the Fast Fourier Transform (FFT) [83]

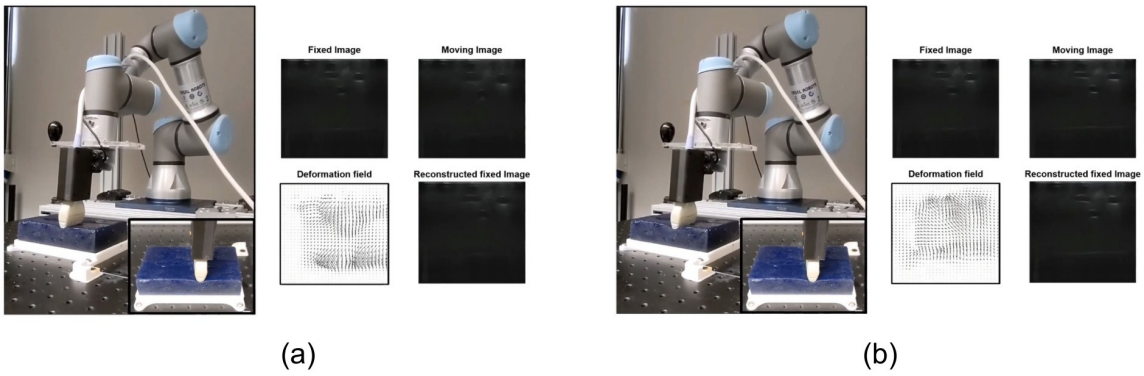


Figure 4.4: Example deformable registration result from robotic experiment on the blue-gel dataset, showing the fixed image I_f, the moving image I_m, and the reconstructed fixed image I'_f, and the deformation field calculated between I_f and I_m using U-RAFT, with a displacement vector plotted on a 4x4 pixel grid. Sub-figure (a) shows compression and (b) shows uncompression.

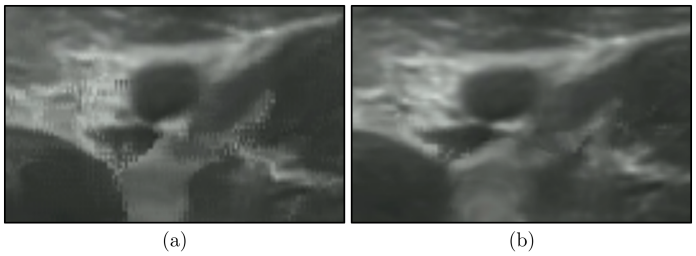


Figure 4.5: (a) Zoomed-in view of an example live-pig reconstructed deformed image I'_d using $\mathcal{L}_{\text{us}}$. (b) Zoomed-in view of the reconstructed deformed image I'_d using $\mathcal{L}_{\text{fa-cyclic-us}}$. The use of the feature-aware cyclic loss function helps remove the optical distortions observed in (a) for large deformations.

of d_{x^*,y^*} over time to calculate the per-pixel displacement frequency.

Table 4.2 provides the mean displacement of pixels for all the frames before and after applying for respiratory compensation. Furthermore, we also calculate the respiration rate, via the FFT, by tracking pixels near the rib cage in the ultrasound images. We calculated the respiration rate to be approximately 18 beats per minute (bpm), which is consistent with the recorded breathing rate of approximately 19 bpm during the experiment.

Table 4.2: Average pixel displacement before and after applying respiratory compensation

Dataset	**Avg. displacement before compensation (pixel)**	**Avg. displacement after compensation (pixel)**
Pig 1	2.09	0.32
Pig 2	2.32	0.65
Pig 3	1.75	0.49

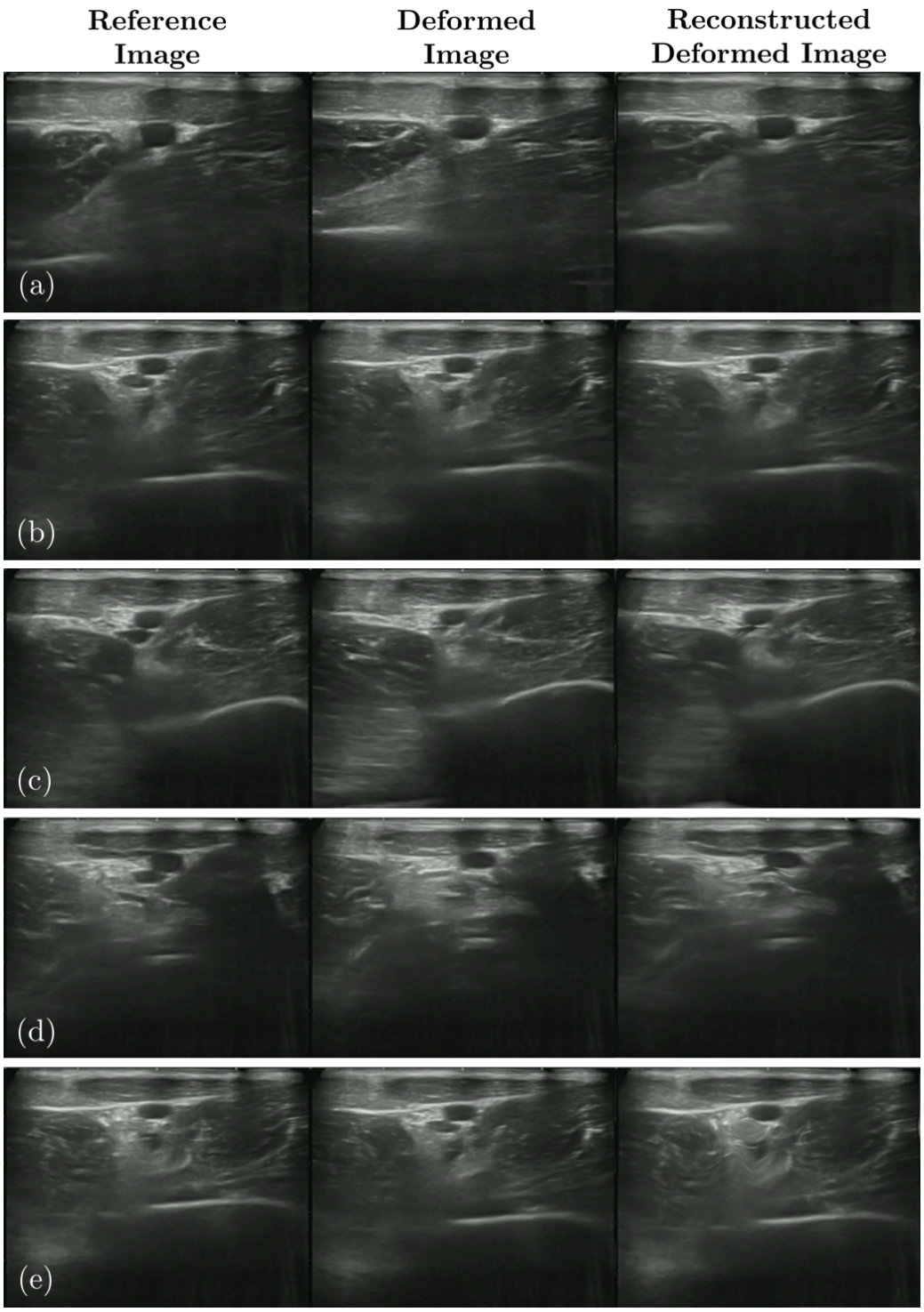


Figure 4.6: Live-pig results showing reference, deformed, and reconstructed deformed images for (a) small deformation, (b) medium deformation, (c) large deformation, (d) vessel collapse, and (e) an atypical case of a vessel decollapsing. The similarity between the deformed and the reconstructed deformed images shows the efficacy of U-RAFT. Registration performance drops for the case of a vessel decollapsing, but even in this atypical scenario our approach fails gracefully.

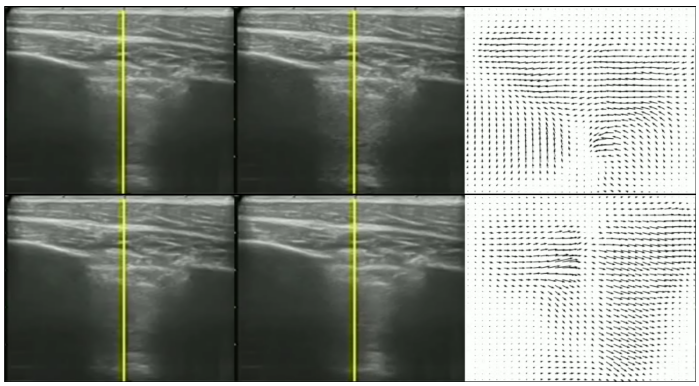


Figure 4.7: Fixed (I_f) and moving (I_m) image and their respective displacement fields

.

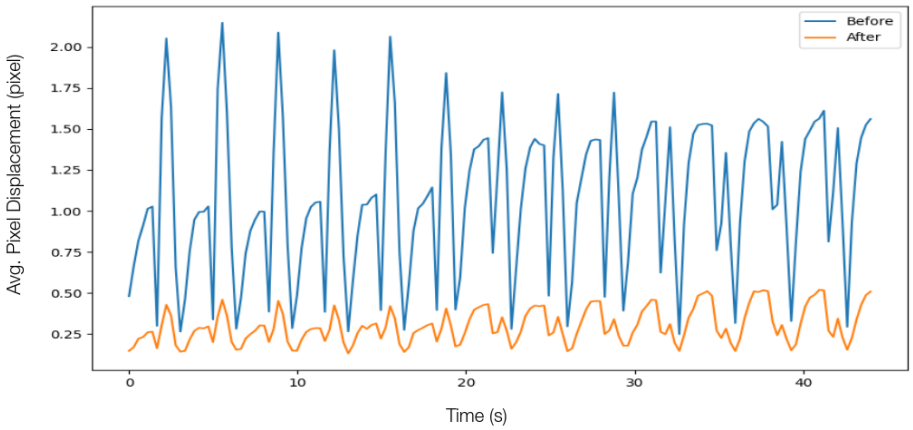


Figure 4.8: Average pixel displacement before and after applying respiratory compensation

.

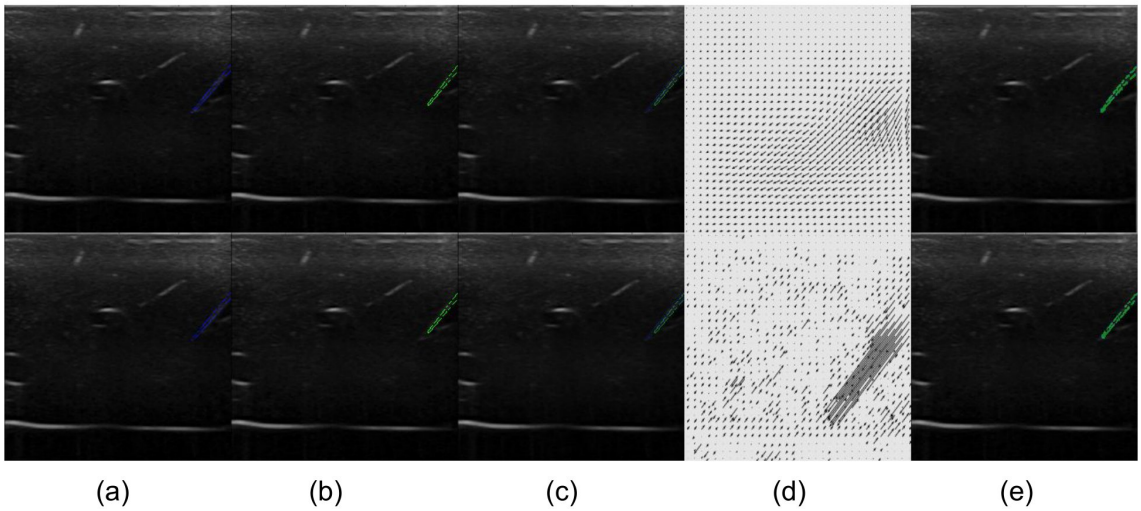


Figure 4.9: (a) Reference Image (I_r), (b) Target Image (I_t), (c) Unregistered I_r and I_t, (d) Predicted displacement field between I_r and I_t using U-RAFT (first row) and Farneback Optical Flow [1] (second row) and (e) Registered I_r and I_t. It's observed that the displacement field predicted by U-RAFT is more accurate and has fewer false positives, especially for static regions of the image.

4.4.3 Registration of curved needles in ultrasound images

Needle registration is a crucial task, and we employ the unsupervised RAFT algorithm to address this challenge specifically for registering curved needles in ultrasound images. Our test dataset comprises seven trials of needle data of *blue-gel-needle*, each consisting of video sequences with 50-80 frames each. Similarly, we have three trials of needle data of *live-pig-needle* with each trial having 200-400 frames. Additionally, for each needle dataset, we have annotations available for each needle during the insertion process.

In this experiment, we utilize U-RAFT to register the reference image (I_r) and the target image (I_t). The needle-tip-error(NTE) serves as our metric for comparing the performance of different methods. For this experiment, we fix the range between the I_r and the target image I_t to 5 frames. The needle-tip-error for the *blue-gel-needle* dataset is provided in Table 4.3 and the NTE for the *live-pig-needle* is provided in Table 4.4. In Tables 4.3 and 4.4, we compare different variants of RAFT, which differ in their training dataset and their initialization. The U-RAFT model which is fine-tuned on 10% (of the vascular dataset) needle dataset performs better than the RAFT model trained on the vascular dataset. In the U-RAFT model, initializing the flow map with the previous flow map yields superior performance compared to initializing it with a zero map, as it enables the model to incorporate temporal dependencies and improve flow predictions iteratively. We also compare the results of U-RAFT with Unsupervised Flownet [48] and Farneback optical flow [1] estimation using the OpenCV [79] implementation.

Table 4.3: Mean Needle Tip Error (NTE) on the *blue-gel-needle*. NTE is calculated after registering needle tips in the phantom's curved needle insertion. The method with the lowest mean NTE is highlighted. The mean NTE prior to registration was 9.44 pixels

Method	**Mean Needle Tip Error (NTE) in pixels**
U-RAFT (pre-trained on vascular data) + Initialized with last flow map	4.35
U-RAFT (pre-trained on vascular data) + Initialized with zero-flow map	4.56
U-RAFT (fine-tuned on needle data) + Initialized with last flow map	**4.21**
U-RAFT (fine-tuned on needle data) + Initialized with zero-flow map	4.42
Unsupervised Flownet [48]	6.23
Farneback Optical Flow estimation [1]	7.08

Table 4.4: Mean Needle Tip Error (NTE) on the *pig-lab-needle*. NTE is calculated after registering needle tips in the pig's curved needle insertion. The method with the lowest mean NTE is highlighted. The mean NTE prior to registration was 4.24 pixels

Method	**Mean Needle Tip Error (NTE) in pixels**
U-RAFT (pre-trained on vascular data) + Initialized with last flow map	3.12
U-RAFT (pre-trained on vascular data) + Initialized with zero-flow map	3.21
U-RAFT (fine-tuned on needle data) + Initialized with last flow map	**3.02**
U-RAFT (fine-tuned on needle data) + Initialized with zero-flow map	3.23
Unsupervised Flownet [48]	3.38
Farneback Optical Flow estimation [1]	3.43

4.5 Conclusion

In conclusion, this chapter proposes an unsupervised approach for training an ultrasound-to-ultrasound deformable registration model. We propose and compare three different loss functions and show that the function based on feature-aware cyclic loss performs best. The effectiveness of the proposed approach is demonstrated through experiments on medical phantom as well as multiple *in-vivo* porcine datasets. We also use the U-RAFT model for accurate pixel-tracking at online rates, which makes it suitable for compensating for tissue motion, such as motion due to respiration. We also show the effectiveness of the U-RAFT model for curved needle registration, where we are able to improve the registration results over other methods like [48] and [1].

Chapter 5

Improving segmentation in ultrasound images using optical flow

5.1 Introduction

Vascular access and subsequent placement of central venous and arterial catheters is an essential first step for delivering life-saving medical care to trauma patients, e.g. administering anesthesia, monitoring vitals, and delivering resuscitative treatments like Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA). Accessing a blood vessel, commonly done via the Seldinger technique [84], requires insertion of a needle into the center of the vessel, which is typically done by a highly skilled clinician using ultrasound to determine where to insert the needle.

The work in this chapter is motivated by the potential benefits of supporting human-guided vascular access with a robot so as to enable personnel away from centers of medical excellence to gain vascular access while avoiding vessel wall damage and hematomas caused by failed needle insertion attempts. This would be especially impactful on battlefields and in mass casualty disasters where there is limited access to trained medical personnel and hospital facilities. Examples of recent work towards the goal of robot-assisted femoral vessel access under ultrasound guidance include a hand-held device [85] and our group's system using a robot manipulator [59].

Convolutional neural networks (CNNs), U-Net [86], and variants of U-Net [87] are commonly used in medical image segmentation, including vessel segmentation in ultrasound [85, 88–90]. Training these models requires time-consuming labeling of the vessel contours in each image by personnel trained to interpret ultrasound images. Furthermore, the training set includes only a small subset of deformed vessel shapes that could occur, limiting the ability of the model to generalize to probe forces and deformed vessel shapes outside the training dataset. Although nonlinear warping augmentations could be applied to images [86], this technique is not guaranteed to generate physically realistic image augmentations.

In this chapter, we propose an approach to use deformable registration for augmenting images for training vessel segmentation models with small training datasets. However capturing the precise dynamics is a crucial requirement when generating photo-realistic

ultrasound images under the application of forces. Previous methods, such as [8, 9] for brain MRI images, as well as [10] for lung MRI images, have employed data augmentation techniques but have not emphasized capturing accurate process dynamics. Consequently, these methods tend to generate out-of-distribution datasets that do not effectively contribute to improving vessel segmentation. Alternatively, physics-based simulators like SOFA [11] offer accurate physics simulations for simulating anatomical deformations under the influence of forces. However, utilizing these simulators requires an accurate 3D model and knowledge of the material properties of the subject. Obtaining the 3D model and material properties of the subject can be an arduous and sometimes unfeasible task, rendering the use of such simulators more challenging.

In this chapter, we propose an approach that emulates the compression process of ultrasound images without relying on 3D models or material properties. Subsequently, we employ the generated ultrasound images to augment and enhance segmentation specifically at higher forces. By eliminating the need for 3D models and material properties, our method offers a more accessible and practical solution for improving vessel segmentation through image augmentation.

5.2 Methods

5.2.1 Synthetic data generation

Strategy 1: Constant stiffness (K_{const}) model

In this method, we assume a constant stiffness model K during the extent of compression of the subject. Suppose we have ultrasound images at F_r as I_r and at F_d as I_d. The simplified stiffness model enables us to use the displacement fields (DF) between deformed and reference ultrasound images and then utilize them to analytically calculate the deformation field at additional probe force values F_{new} calculated using,

$$\begin{aligned} \Delta F_{d,r} &= K u_{d,r} \\ \Delta F_{\text{new},r} &= K u_{\text{new},r} \\ u_{\text{new},r} &= \frac{\Delta F_{\text{new},r}}{\Delta F_{d,r}} u_{d,r} \end{aligned} \tag{5.1}$$

where, $\Delta F_{d,r} = F_d - F_r$ and $\Delta F_{new,r} = F_{\text{new}} - F_r$ We use the U-RAFT model from Chapter 3 to predict the displacement field between deformed and reference ultrasound image $u_{d,r}$. We then calculate $u_{\text{new},r}$ using Equation 5.1 and pass $u_{\text{new},r}$ through Spatially Transformer Network (STN) to generate $I_{\text{new}} = \text{STN}(u_{\text{new},r}, I_r)$. We then use synthetic images to augment the vessel segmentation dataset. We will show in Section 5.3.2 that this data augmentation technique helps a U-Net vessel segmentation model generalize to different forces.

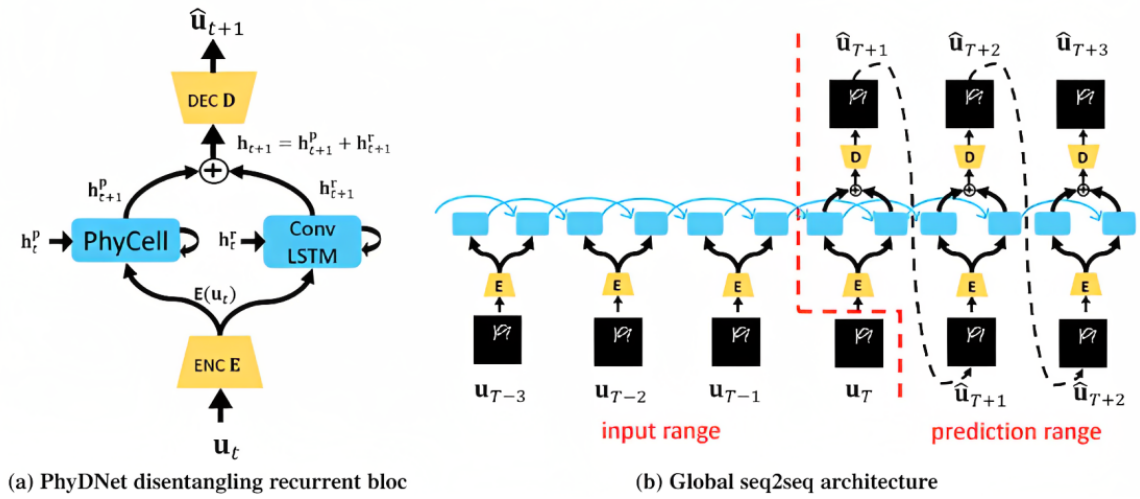
Strategy 2: Variable stiffness (K_{var}) model

This method acknowledges the fact that the stiffness model of the subject is not necessarily constant, as demonstrated in the previous section. This observation motivates us to implicitly model the stiffness of the subject using a neural network. Since the compression data is sequential in nature, we can employ a recurrent neural network (RNN) to effectively capture and model the stiffness variations throughout the compression process.

Let there be a sequence of N ultrasound images collected at a particular location of the phantom $I_1 \dots I_N$ along with the respective force values $F_1 \dots F_N$ at which these images are collected. We use the recurrent module PhyDNet [91] which leverages physics knowledge on dynamics and disentangles it from other unknown factors of variations necessary for future forecasting.

PhyDNet

The core of PhyDNet is a recurrent block projecting input images I_t into a latent space $\mathcal{H}$. The video evolution in the latent space $\mathcal{H}$ is thus governed by the following partial differential equation (PDE): $\delta h(t,x)/\delta t = \mathcal{M}_p(h^p, I) + \mathcal{M}_r(h^r, I)$. To reach this objective, we introduce a recurrent bloc which is shown in Figure 5.1. A video frame I_t at time t is mapped by a deep convolutional encoder E into a latent space representing the targeted



(a) PhyDNet disentangling recurrent bloc (b) Global seq2seq architecture

Figure 5.1: PhyDNet architecture as shown in [2]

space $\mathcal{H}$. $E(I_t)$ is then used as input for two parallel recurrent neural networks, incorporating this spatial representation into a dynamical model. The left branch in Figure 5.1 models the latent representation h^p fulfilling the physical part $\mathcal{M}_p(h^p, I)$. The PDE is modeled by our recurrent physical cell PhyCell, which leads to the computation of h^p_{t+1} from $E(I_t)$ and h^p_t. The right branch in Figure 5.1 models the latent representation hr fulfilling the residual part $\mathcal{M}_r(h^r, I)$. We use a generic recurrent neural network for this task, e.g. ConvLSTM [92] for learning the residual component, which computes h^r_{t+1} from $E(u_t)$ and residual hidden state h^r_r. $h_{t+1} = h^p_{t+1} + h^r_{t+1}$ is the combined representation processed by a deep decoder D to forecast the image I_{t+1}. More details on PhyDNet can be found in [2].

Physics Inspired Ultrasound Image Generator (Phy-UGen)

Phy-UGen is a dual-PhyDNet module that uses the video prediction capabilities of the PhyDNet module to understand the dynamics of vessel compression in ultrasound images and use it to generate synthetic ultrasound images simulating newer force compression. Also instead of having a time sequence data we have a *force-sequence* data where the data along the force axis are correlated. We have a dual-PhyDNet module, in which one module takes ultrasound image I_t at force F_t N as an input while the other module takes the displacement field $u_{t,1}$ as an input. $u_{t,1}$ is calculated from the U-RAFT model between the ultrasound image I_t and I_1. The intuition between using dual-PhyDNet is such that apart from learning the pixel-level dynamics from images, we can also learn relevant pixel-level dynamics from the displacement field, especially effective in highlighting the displacement of the pixel. Once the modes learn individual modules, we devise a physics aggregator block that combines the latent representation from the $h^p_{t+1,\text{img}}$ with the latent representation of $h^p_{t+1,\text{DF}}$. The aggregator learns a convolutional filter C, which aggregates using, $h^p_{t+1} = h^p_{t+1,\text{img}} + C \cdot (h^p_{t+1,\text{flow}} - h^p_{t+1,\text{img}})$. Furthermore, the aggregated pyhsics latent space h^p_{t+1} is concatenated with the $h^r_{t+1,\text{img}}$ and processed by the decoder D_{img} to output the ultrasound image at a newer force value I_{t+1}. Further, we modify the video-prediction nature of the

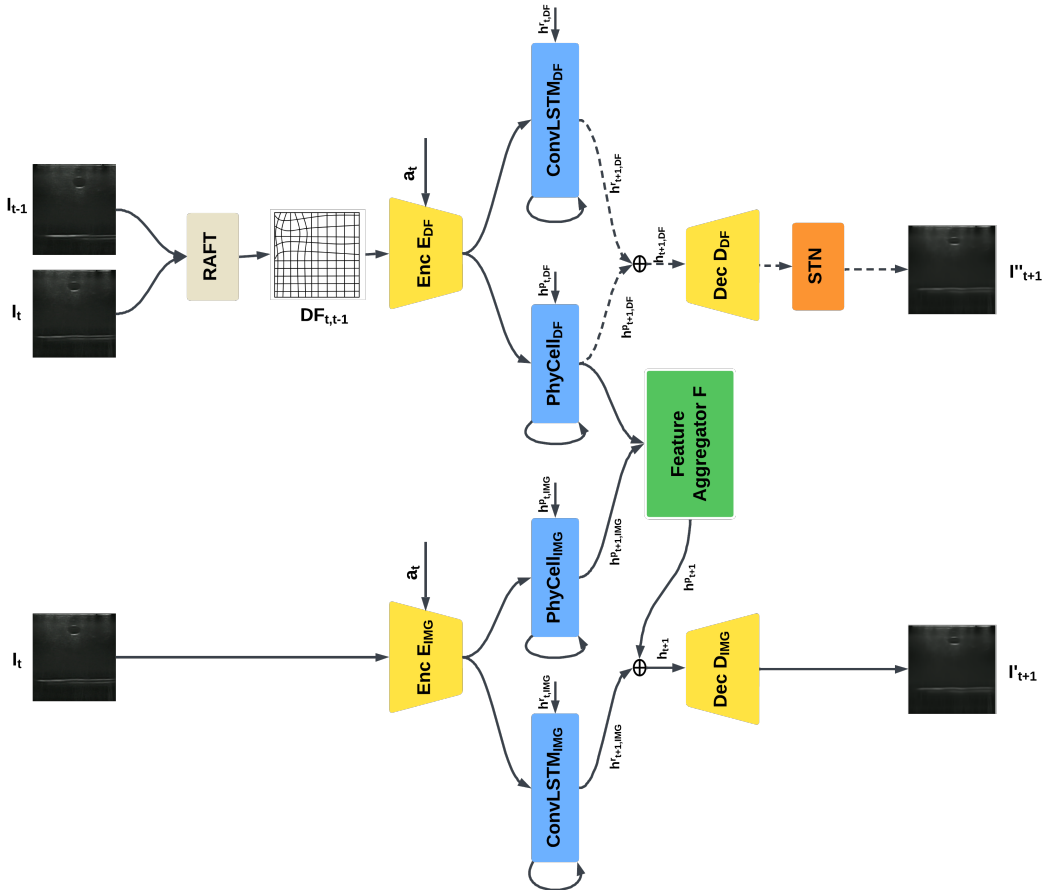


Figure 5.2: Physics Inspired Ultrasound Image Generator (Phy-UGen) network architecture

PhyDNet into an action-conditioned video prediction module, where the action is $a_{t+1} = F_{t+1} - F_t$ is passed as an additional input. To pass as input along with the image I_t, we map a_t to a 2D tensor of the same resolution as the input image. The 2D tensor action is concatenated with the input ultrasound image and the input displacement field respectively and passed onto the PhyDNet to predict the next ultrasound image I'_{t+1}.

Loss function for learning a variable stiffness model

In this section, we define the loss function to train the Phy-UGen module discussed in the last section. For simplicity of explanation, we divide our loss functions into 2 parts: 1. Reconstruction loss between the predicted image and the actual image. 2. Reconstruction and smoothness loss between the predicted displacement field and the actual displacement field by U-RAFT.

The similarity loss between two images denoted as $\mathcal{L}_{image}$, consists of two parts $\mathcal{L}_{\text{ssim}}$, a multi-scale structural similarity (SSIM) loss that penalizes the differences in appearance (luminance, contrast and structure) between the predicted and the actual ultrasound image in the sequence and $\mathcal{L}_{\text{mse}}$, which penalizes difference in pixel value between the predicted

and actual image:

$$\begin{aligned}\mathcal{L}_{\text{image}}(I_{t+1}, I'_{t+1,\text{img}}, I'_{t+1,\text{flow}}) = \alpha(\beta\mathcal{L}_{\text{ssim}}(I_{t+1}, I'_{t+1,\text{img}}) \\ + (1-\beta)\mathcal{L}_{\text{mse}}(I_{t+1}, I'_{t+1,\text{img}})) \\ + (1-\alpha)(\beta\mathcal{L}_{\text{ssim}}(I_{t+1}, I'_{t+1,\text{flow}}) \\ + (1-\beta)\mathcal{L}_{\text{mse}}(I_{t+1}, I'_{t+1,\text{flow}}))\end{aligned} \tag{5.2}$$

where $\alpha, \beta \in \mathbb{R}$ are parameters to adjust the relative weight between the two branches of Phy-UGen modules and between the $\mathcal{L}_{\text{ssim}}$ and $\mathcal{L}_{\text{mse}}$. $\mathcal{L}_{\text{mse}}$ and $\mathcal{L}_{\text{ssim}}$ are given by:

$$\begin{aligned}\mathcal{L}_{\text{mse}}(I, I') &= \sum_{x,y}(I(x,y) - I'(x,y))^2 \\ \mathcal{L}_{\text{ssim}}(I, I') &= 1 - \text{SSIM}(I, I')\end{aligned} \tag{5.3}$$

The second part of the loss function we consider, denoted as $\mathcal{L}_{flow}$, consists of two parts: $\mathcal{L}_{\text{mse}}$, that penalizes the differences in flow appearance between u_{t+1} and u'_{t+1}, and $\mathcal{L}_{\text{smooth}}$, which penalizes abrupt changes in the neighboring pixels of u'_{t+1} :

$$\mathcal{L}_{\text{us}}(u_{t+1}, u'_{t+1}) = \gamma\mathcal{L}_{\text{mse}}(u_{t+1}, u'_{t+1}) + (1-\gamma)\mathcal{L}_{\text{smooth}}\left(u'_{t+1}\right) \tag{5.4}$$

where $\gamma \in \mathbb{R}$ is a parameter to adjust the relative weight of $\mathcal{L}_{\text{mse}}$ and $\mathcal{L}_{\text{smooth}}$. $\mathcal{L}_{\text{mse}}$ is same as explained in Eq.5.3 and $\mathcal{L}_{\text{smooth}}$ are given by:

$$\mathcal{L}_{\text{smooth}}(u) = \text{mean}_{xy}\left(\frac{\nabla^2 u(x,y)}{\nabla x^2} + \frac{\nabla^2 u(x,y)}{\nabla y^2}\right) \tag{5.5}$$

where x, y are the pixel location of a 2D-deformation field, and mean_{xy} denotes the mean over all pixels.

Experimental setup and training details

We use the use robotic setup mentioned in Sec in Chapter 4 for collecting ultrasound images at different force values. For this experiment, we use palpation mode where we execute force-controlled sinusoidal motion ranging the force value from 2N to 22N at a frequency of 10 Hz. The minimum force value of 2N was chosen to ensure the minimum possible acoustic coupling between the phantom and the ultrasound probe. The choice of a maximum force value of 22N was determined by considering the highest force that could be safely applied before the occurrence of vessel collapse was observed. This decision aimed to establish a limit that would prevent any potential damage to the phantom under study. We collected 25 palpation sequences from uniformly sampled locations on the phantom to demonstrate our model's capabilities. For the constant stiffness (K_{const}) model, we use the ultrasound image collected at 2N and 10N to generate synthetic ultrasound images at all other forces. The variable stiffness (K_{var}) model, which needs a more sequential form of input, is trained using all images between 2N-10N. The remaining portion of the sequence (10N to 22N) was used for validation, ensuring the generalization of our methods to higher

force values. Additionally, we utilized 20 out of the 25 palpation sequences for training to evaluate the model's generalization on new phantom locations.

To accurately estimate the displacement field on the vascular dataset for the constant stiffness model, we rely on a trained U-RAFT model. This model, trained in Chapter 4 using the cyclic-loss approach, provides the necessary capability to estimate precise displacement fields. By utilizing the trained U-RAFT model, we ensure accurate and reliable results for the constant stiffness model analysis on the vascular dataset.

For training the Phy-UGen module, all the module weights were initialized with Lecun initialization [77] and were trained for 600 additional epochs. The implementation is highly parallelized and performs full-batch gradient descent using the Adam [93] optimizer in the Pytorch Autograd library[81], with a batch size of 4 with a learning rate of 0.001.

5.3 Results

5.3.1 Photorealism of generated synthetic images

In this section we evaluate the photo-realism of the generated ultrasound images using both the constant stiffness and the variable stiffness model described in the last section. We use three different metric for comparing the photo-realism of these two images: 1. SSIM metric 2. PSNR metric and 3. IoU score between the synthetic image and the ground truth image at the same force value. We evaluate our methods on 5 different palpation video sequence and at force values $[2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22]$N.

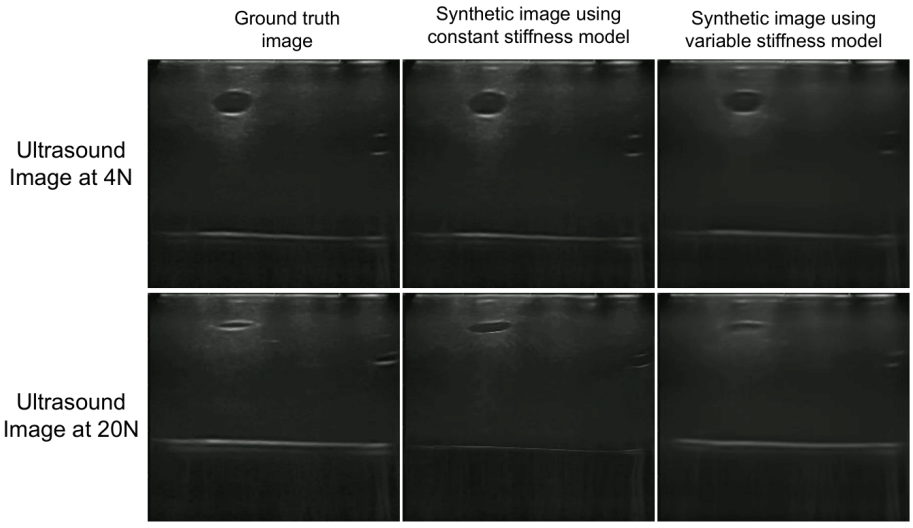


Figure 5.3: Ground truth, synthetic image generated using K_{const} model and synthetic image generated using K_{var} model. Ultrasound images are at 4N and 20N respectively

Structural Similarity Index

The Structural Similarity Index (SSIM) is a widely used image quality metric that quantifies the similarity between two images. It takes into account the luminance, contrast, and structural information of the images. The SSIM index is calculated using the following equation: $\text{SSIM}(x, y) = \frac{2\mu_x\mu_y+c_1}{\mu_x^2+\mu_y^2+c_1} \cdot \frac{2\sigma_{xy}+c_2}{\sigma_x^2+\sigma_y^2+c_2} \cdot \frac{\sigma_{xy}+c_3}{\sigma_x\sigma_y+c_3}$ In the equation, x and y represent the compared images, μ_x and μ_y are the mean values, σ_x and σ_y are the standard deviations, σ_{xy} is the cross-covariance, and c_1, c_2, and c_3 are small constants to ensure stability. The SSIM metric captures both global and local image similarities, making it robust to variations in illumination and contrast. Higher SSIM values indicate greater similarity between the images, while lower values indicate greater dissimilarity.

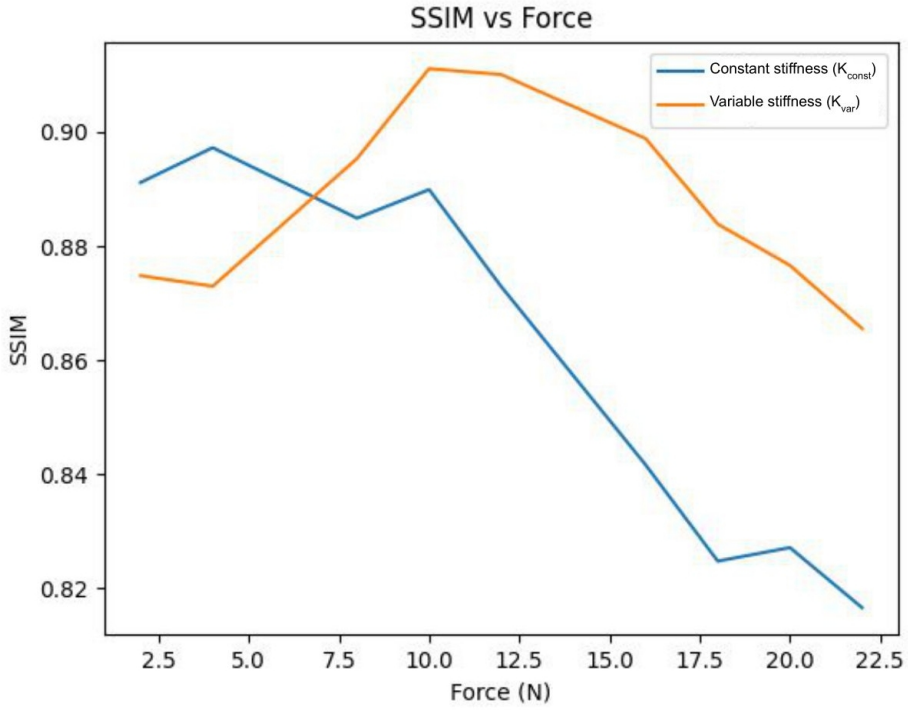


Figure 5.4: Comparison of SSIM values between synthetic images generated from the constant stiffness model and the variable stiffness model and the ground truth image at different force values

In our experiment, we validate constant and variable stiffness by generating images from a 2N and 10 N ultrasound image collected on the blue-gel-phantom. Figure 5.4 shows the SSIM value between the generated ultrasound images and ground truth images between 2N and 20N force values. The SSIM values for the K_{const} value shows better SSIM score for lower force values but shows a gradual decrease in the score as the force is increased. On the other hand, the K_{var} model outperforms the K_{const} model for higher force values showing higher photo-realism in generated images. Especially at higher forces the K_{const} model compresses the vessel in an unrealistic manner as seen in Figure 5.3.

Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is a widely used image quality metric that measures the fidelity of a reconstructed or compressed image compared to the original. It quantifies the pixel-wise difference using the Mean Squared Error (MSE) and provides a logarithmic representation of the signal-to-noise ratio. The PSNR is calculated as $\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right)$. Higher PSNR values indicate higher similarity and lower distortion between the images. The maximum possible PSNR value for 8-bit images (our resolution) is 48 dB.

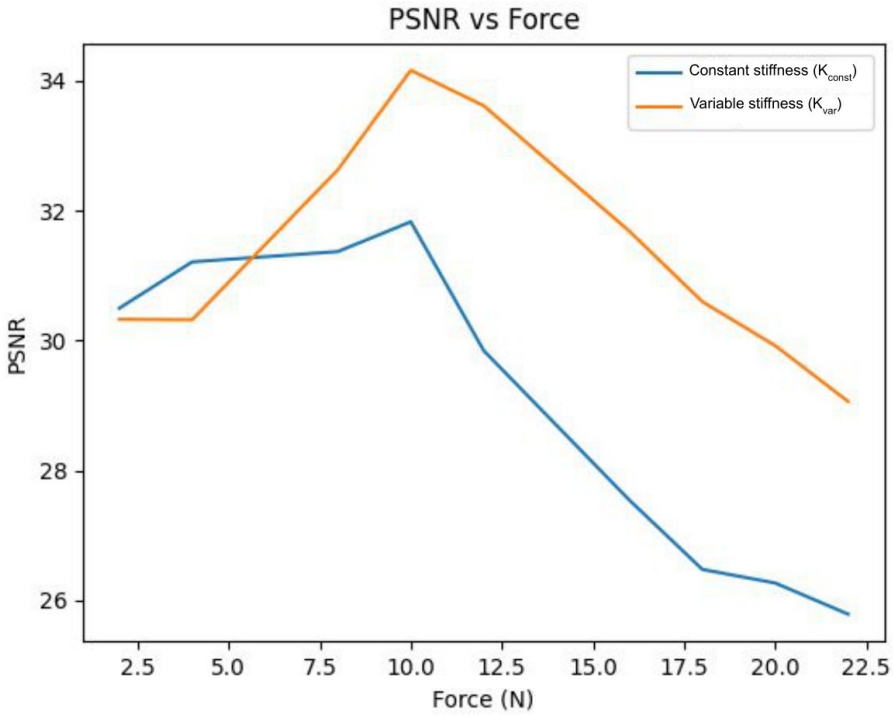


Figure 5.5: Comparison of PSNR values between synthetic images generated from the constant stiffness model and the variable stiffness model and the ground truth image at different force values

Both the K_{const} and the K_{var} model demonstrate high PSNR values. Especially for higher force values, the variable stiffness model performs superior performance demonstrating less noise-induced at higher force values.

Intersection-over-Union

Intersection over Union (IoU) is a popular metric for evaluating object detection and segmentation methods. It measures the overlap between predicted and ground truth regions. The IoU is calculated as the ratio of the intersection area to the union area: $\text{IoU} = \frac{\text{Intersection}}{\text{Union}}$. Higher IoU values between the labels of the generated and synthetic images indicate better alignment of the vessels.

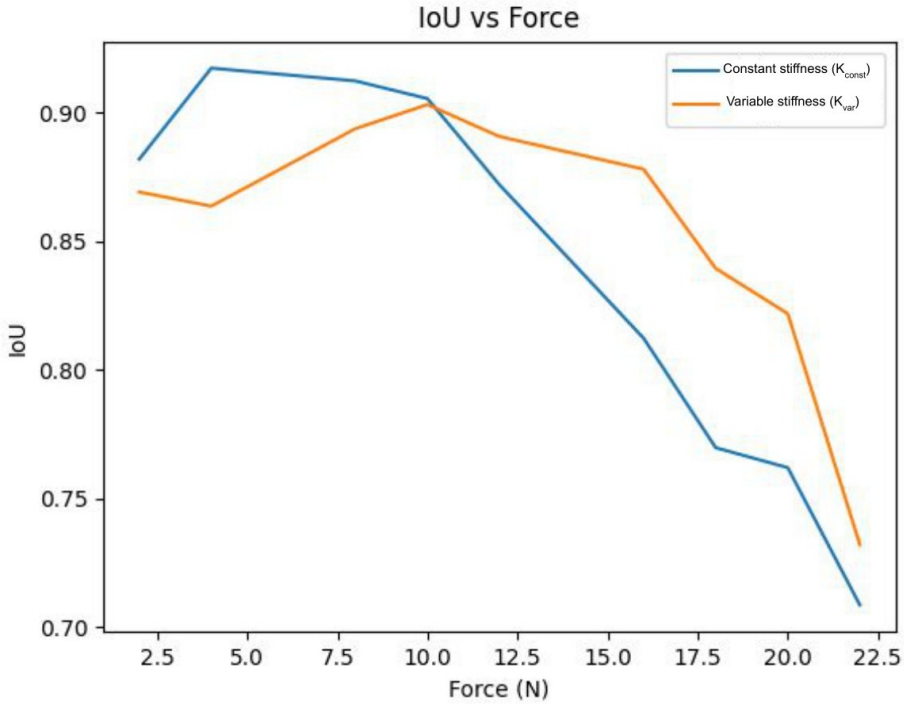


Figure 5.6: Comparison of IoU values between synthetic images generated from the constant stiffness model and the variable stiffness model and the ground truth image at different force values

As seen in Figure 5.6, the constant stiffness model demonstrates a high IoU value for lower-force images. But as the force increases, the unrealistic compression of the vessels with the method leads to poor IoU scores. On the other hand, the K_{var} model, shows moderately higher performance for all ranges of forces, demonstrating higher similarity between the ground truth and the generated images.

5.3.2 Improvement in segmentation using synthetic data generation

In this set of experiments, we evaluate the effect of using synthetic images to improve the results of vessel segmentation under tissue deformations. In the first experiment, we evaluated the realism of the synthetic ultrasound images created using the method described in Section 4.2-C on the *blue-gel-vascular*. Using the scanning mode of the robot, we collected $\sim$1600 images on the *blue-gel*. We then created synthetic images at 9 intermediate force values $F = 4, 6, 8, 12, 14, 16, 18, 20, 22$ N, denoted as $D_{4,6,8,12,14,16,18,20,22,\text{syn}}$, respectively. Example synthetic images are shown in Figure 5.7.

We now show how the synthetic images generated by U-RAFT can be used as a data augmentation technique to improve a vessel segmentation model with respect to the intersection-over-union (IoU) metric. We use the $\sim$ 1600 blue-gel images from each force value as our training dataset. We collected an additional $\sim$ 400 images each for testing and validation datasets.

For this experiment, we train individual U-Net models on $D^{\text{train}}_{2,\text{real}}$, $D^{\text{train}}_{10,\text{real}}$ and a combi-

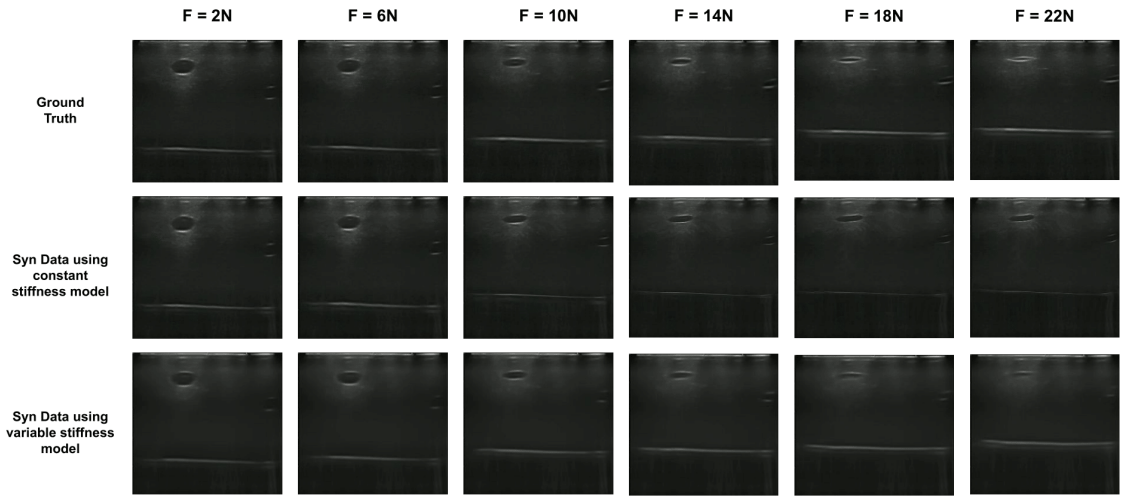


Figure 5.7: Ground truth images, synthetic images generated using constant stiffness model and synthetic images generated using variable stiffness model at 2N, 6N, 10N, 14N, 18N, and 22N

nation of $D^{\text{train}}_{2,\text{real}} + D^{\text{train}}_{10,\text{real}}$ dataset. Further, we augment the $D^{\text{train}}_{2,\text{real}} + D^{\text{train}}_{10,\text{real}}$ dataset with multiple-force synthetic data ($D_{4,6,8,12,14,16,18,20,22,\text{syn}}(K_{const}\text{and}K_{var})$) and use it to train a U-Net model. We also compare our augmentation technique to the random elastic augmentation mentioned in [86] to underscore the significance of augmenting using multiple-force synthetic data. We show the improvement in segmentation on the blue-gel

Table 5.1: Comparing performance on the test datasets of a U-Net model with and without data augmention using the synthetic dataset. The results from the best performing dataset is highlighted.

Dataset (Blue-gel)	**IoU**
$D^{\textbf{train}}_{2,\textbf{real}}$	0.55
$D^{\textbf{train}}_{10,\textbf{real}}$	0.57
$D^{\textbf{train}}_{2,10,\textbf{real}}$	0.58
$D^{\textbf{train}}_{2,10,\textbf{real}}$+ $D^{\textbf{train}}_{4,6,8,12,14,16,18,20,22,\textbf{syn}}(K_{const})$	0.61
$D^{\textbf{train}}_{2,10,\textbf{real}}$+ $D^{\textbf{train}}_{4,6,8,12,14,16,18,20,22,\textbf{syn}}(K_{var})$	**0.66**
$D^{\textbf{train}}_{2,10,\textbf{real}}$+ $D^{\textbf{train}}_{\textbf{rand-syn}}$	0.61

As shown in Table 5.1, augmenting the real dataset with multi-force K_{var} model synthetic data outperforms the training done using only real images. We also see it outperform the model trained with random elastic deformations, highlighting the need for realistic force-based augmentation using deformable registration. An example of this is shown in Fig. 5.8, where a model trained using only on 2 N images fails to segment a nearly collapsed vein, but the model trained with our synthetic images is able to segment the vessel.

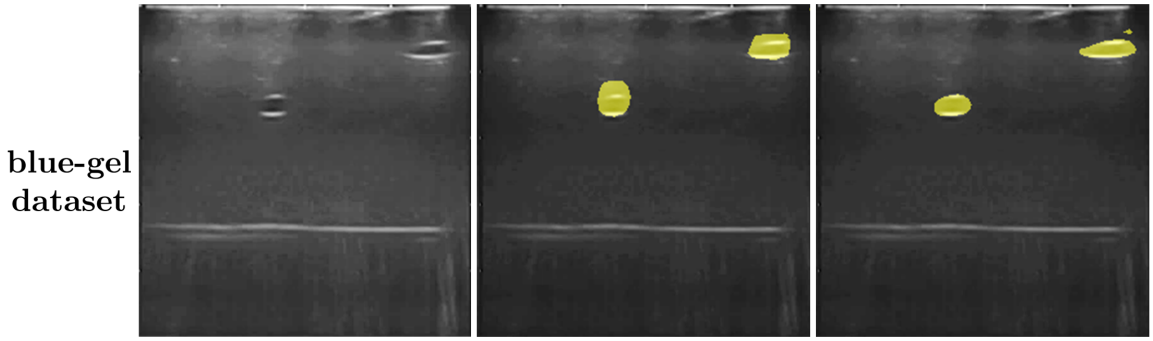


Figure 5.8: (a) Ultrasound image at 10 N. (b) Ultrasound image with predicted segmentation mask using a U-Net model trained using real data collected at 2 N. (c) Ultrasound image with predicted segmentation mask using U-Net model train using real and synthetic data. The inclusion of synthetic data improves segmentation accuracy.

5.4 Conclusion

We demonstrate how our augmentation approach (constant stiffness and variable stiffness) can be used to generate synthetic deformed images to expand the size of a vessel segmentation training dataset and improve vessel segmentation performance. We validate our approach on a benchtop human tissue/vessel phantom highlighting the practical application of our deformable registration model in real-world medical imaging tasks. Overall, this chapter presents an innovative solution to address the challenges associated with ultrasound imaging, particularly in the areas of ultrasound image segmentation which are critical for accurate vascular access.

Chapter 6

Conclusion and future works

In conclusion, this thesis addresses crucial challenges in the field of robotic ultrasound scanning and image analysis. Among many medical imaging techniques, ultrasound stands out for its safety, cost-effectiveness, real-time imaging capabilities, and portability, especially in point-of-care ultrasound (POCUS) applications. Despite its advantages, ultrasound scanning relies heavily on skilled sonographers, leading to challenges in certain scenarios. The thesis aims to overcome these challenges by enhancing the autonomy of robotic ultrasound systems and improving the generalizability of ultrasound imaging algorithms.

The first chapter presents an improved semi-automated force-controlled approach for robotic ultrasound scanning. The utilization of SLERP combined with a hybrid force-position controller demonstrates promising results for scanning highly curved surfaces accurately and efficiently. We quantify our scanning method using the quality of ultrasound images as well as the force profiles.

In the second chapter, the thesis introduces the U-RAFT model, a novel unsupervised deep-learning method for predicting deformations in ultrasound images. By employing deformable registration, the model accurately estimates per-pixel deformation fields, leading to diverse applications like synthetic image generation and registration tasks for respiratory motion compensation and curved needle registration. The versatility of the U-RAFT model across various medical imaging tasks highlights its effectiveness and broad applicability.

In the third chapter, the thesis proposes a novel approach to using deformable registration for augmenting training images and enhancing vessel segmentation models. Our method generates photo-realistic ultrasound images without relying on 3D models or material properties and improves segmentation capabilities at different compression levels.

Overall, this thesis contributes significantly to the advancement of robotic ultrasound scanning, deformable registration, and image augmentation in the ultrasound imaging domain. The proposed methods offer promising solutions to improve the quality and efficiency of ultrasound imaging and analysis, paving the way for more accessible and effective medical interventions.

As for future work, I plan to improve the proposed deformable registration algorithm U-RAFT to predict higher deformations especially for cases of vessel collapses. One of the proposed methods can be incorporating an iterative approach to predict larger deformations. Furthermore, I also want to improve the data generation method mentioned in Chapter 5. Currently, the method is able to generate accurate vessel shapes at higher force values.

But in doing so, it doesn't replicate the accurate ultrasound speckle pattern. This causes a decrease in SSIM, PSNR, and IoU scores for higher force values. I plan to improve this by adding an extra discriminator loss in Equation 5.2, to decrease the visual differences between the generated and ground truth images. Also, in the future, I want to validate the K_{var} method on more challenging datasets like the *live-pig-vascular*. In order to do so, I plan to collect more ultrasound images in the palpation mode on the live pigs.

Bibliography

[1] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*, pages 363–370. Springer, 2003. (document), 4.9, 4.4.3, 4.3, 4.4, 4.5

[2] Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction, 2020. (document), 5.1, 5.2.1

[3] Zhi-Pei Liang and Paul C Lauterbur. *Principles of magnetic resonance imaging*. SPIE Optical Engineering Press Belllingham, WA, 2000. 1

[4] Norbert J Pelc. Recent and future directions in ct imaging. *Annals of biomedical engineering*, 42:260–268, 2014. 1

[5] Peter NT Wells. Ultrasound imaging. *Physics in medicine & biology*, 51(13):R83, 2006. 1

[6] Zachary Teed and Jia Deng. RAFT: Recurrent all-pairs field transforms for optical flow. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 402–419. Springer, 2020. 1, 4.1, 4.2.1, 4.4.2

[7] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and koray kavukcuoglu. Spatial transformer networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. 1, 4.1, 4.2.1

[8] Egor Krivov, Maxim Pisov, and Mikhail Belyaev. MRI augmentation via elastic registration for brain lesions segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*, pages 369–380. Springer, 2018. 1, 4.1, 5.1

[9] Jakub Nalepa, Grzegorz Mrukwa, Szymon Piechaczek, Pablo Ribalta Lorenzo, Michal Marcinkiewicz, Barbara Bobek-Billewicz, Pawel Wawrzyniak, Pawel Ulrych, Janusz Szymanek, Marcin Cwiek, Wojciech Dudzik, Michal Kawulok, and Michael P. Hayball. Data augmentation via image registration. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4250–4254, September 2019. doi: 10.1109/ICIP.2019.8803423. 1, 4.1, 5.1

[10] Nicholas J. Tustison, Brian B. Avants, Zixuan Lin, Xue Feng, Nicholas Cullen, Jaime F. Mata, Lucia Flors, James C. Gee, Talissa A. Altes, John P. Mugler, and

Kun Qing. Convolutional neural networks with template-based data augmentation for functional lung image quantification. *Academic radiology*, 26(3):412–423, March 2019. ISSN 1076-6332. doi: 10.1016/j.acra.2018.08.003. 1, 4.1, 5.1

[11] Eulalie Coevoet, Thor Morales Bieze, Frederick Largilliere, Zhongkai Zhang, Maxime Thieffry, Mario Sanz Lopez, Bruno Carrez, Damien Marchal, Olivier Goury, Jeremie Dequidt, and Christian Duriez. Software toolkit for modeling, simulation and control of soft robots. *Advanced Robotics*, 31:1208–1224, November 2017. doi: 10.1080/01691864.2017.1395362. URL https://inria.hal.science/hal-01649355. 1, 5.1

[12] Jing Wang, Chengzhong Peng, Yan Zhao, Ruizhong Ye, Jun Hong, Haijun Huang, and Legao Chen. Application of a robotic tele-echography system for covid-19 pneumonia. *Journal of Ultrasound in Medicine*, 40(2):385–390, 2021. 2.1

[13] Sotiris Avgousti, Andreas S Panayides, Antonis P Jossif, Eftychios G Christoforou, Pierre Vieyres, Cyril Novales, Sotos Voskarides, and Constantinos S Pattichis. Cardiac ultrasonography over 4g wireless networks using a tele-operated robot. *Healthcare technology letters*, 3(3):212–217, 2016.

[14] Krzysztof Arent, Mateusz Cholewiński, W Domski, M Drwiega, Janusz Jakubiak, Mariusz Janiak, Bogdan Kreczmer, Adam Kurnicki, Bartłomiej Stańczyk, D Szcześniak-Stańczyk, et al. Selected topics in design and application of a robot for remote medical examination with the use of ultrasonography and ascultation from the perspective of the remedi project. *Journal of Automation Mobile Robotics and Intelligent Systems*, 11(2):82–94, 2017. 2.1

[15] Marie-Ange Janvier, Samir Merouche, Louise Allard, Gilles Soulez, and Guy Cloutier. A 3-d ultrasound imaging robotic system to detect and quantify lower limb arterial stenoses: in vivo feasibility. *Ultrasound in medicine & biology*, 40(1):232–243, 2014. 2.1

[16] Purang Abolmaesumi, Septimiu E Salcudean, Wen-Hong Zhu, Mohammad Reza Sirouspour, and Simon Peter DiMaio. Image-guided control of a robot for medical ultrasound. *IEEE transactions on robotics and automation*, 18(1):11–23, 2002.

[17] Zhongliang Jiang, Matthias Grimm, Mingchuan Zhou, Javier Esteban, Walter Simson, Guillaume Zahnd, and Nassir Navab. Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement. *IEEE Robotics and Automation Letters*, 5(2):1342–1349, 2020. 2.1

[18] Luís Santos and Rui Cortesão. A dynamically consistent hierarchical control architecture for robotic-assisted tele-echography with motion and contact dynamics driven by a 3d time-of-flight camera and a force sensor. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2931–2937. IEEE, 2015. 2.1

[19] Laure-Anaïs Chanel, Florent Nageotte, Jonathan Vappou, Jianwen Luo, Loïc Cuvillon, and Michel de Mathelin. Robotized high intensity focused ultrasound (hifu) system for treatment of mobile organs using motion tracking by ultrasound imaging: An in vitro study. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2571–2575. IEEE, 2015.

2.1

[20] Salvatore Virga, Oliver Zettinig, Marco Esposito, Karin Pfister, Benjamin Frisch, Thomas Neff, Nassir Navab, and Christoph Hennersperger. Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 508–513. IEEE, 2016. 2.1

[21] Qinghua Huang, Jiulong Lan, and Xuelong Li. Robotic arm based automatic ultrasound scanning for three-dimensional imaging. *IEEE Transactions on Industrial Informatics*, 15(2):1173–1182, 2018. 2.1

[22] Christoph Hennersperger, Bernhard Fuerst, Salvatore Virga, Oliver Zettinig, Benjamin Frisch, Thomas Neff, and Nassir Navab. Towards mri-based autonomous robotic us acquisitions: a first feasibility study. *IEEE transactions on medical imaging*, 36(2):538–548, 2016. 2.1

[23] Samir Merouche, Louise Allard, Emmanuel Montagnon, Gilles Soulez, Pascal Bigras, and Guy Cloutier. A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of b-mode images. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 63(1):35–46, 2015. 2.1

[24] François Conti, Jaeheung Park, and Oussama Khatib. Interface design and control strategies for a robot assisted ultrasonic examination system. In *Experimental Robotics*, pages 97–113. Springer, 2014. 2.1

[25] Richard Droste, Lior Drukker, Aris T Papageorghiou, and J Alison Noble. Automatic probe movement guidance for freehand obstetric ultrasound. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 583–592. Springer, 2020. 2.1

[26] Keyu Li, Jian Wang, Yangxin Xu, Hao Qin, Dongsheng Liu, Li Liu, and Max Q-H Meng. Autonomous navigation of an ultrasound probe towards standard scan planes with deep reinforcement learning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8302–8308. IEEE, 2021. 2.1

[27] Jay Carriere, Jason Fong, Tyler Meyer, Ron Sloboda, Siraj Husain, Nawaid Usmani, and Mahdi Tavakoli. An admittance-controlled robotic assistant for semi-autonomous breast ultrasound scanning. In *2019 International Symposium on Medical Robotics (ISMR)*, pages 1–7, 2019. doi: 10.1109/ISMR.2019.8710206. 2.2

[28] Jakub Piwowarczyk, Jay Carriere, Kim Adams, and Mahdi Tavakoli. An admittance-controlled force-scaling dexterous assistive robotic system. *Journal of Medical Robotics Research*, 05(01n02):2041002, 2020. doi: 10.1142/S2424905X20410020. URL https://doi.org/10.1142/S2424905X20410020. 2.2

[29] Federica Ferraguti, Chiara Talignani Landi, Lorenzo Sabattini, Marcello Bonfè, Cesare Fantuzzi, and Cristian Secchi. A variable admittance control strategy for stable physical human–robot interaction. *The International Journal of Robotics Research*, 38(6):747–765, 2019. doi: 10.1177/0278364919840415. URL https://doi.org/10.1177/0278364919840415. 2.2

[30] Zhijun Li, Bo Huang, Zhifeng Ye, Mingdi Deng, and Chenguang Yang. Physical human–robot interaction of a robotic exoskeleton by admittance control. *IEEE Transactions on Industrial Electronics*, 65(12):9614–9624, 2018. doi: 10.1109/TIE.2018.2821649. 2.2

[31] Chenguang Yang, Guangzhu Peng, Yanan Li, Rongxin Cui, Long Cheng, and Zhijun Li. Neural networks enhanced adaptive admittance control of optimized robot–environment interaction. *IEEE Transactions on Cybernetics*, 49(7):2568–2579, 2019. doi: 10.1109/TCYB.2018.2828654. 2.2

[32] Arvid QL Keemink, Herman van der Kooij, and Arno HA Stienen. Admittance control for physical human–robot interaction. *The International Journal of Robotics Research*, 37(11):1421–1444, 2018. doi: 10.1177/0278364918768950. URL https://doi.org/10.1177/0278364918768950. 2.2

[33] Chaim Broit. Optimal registration of deformed images. 1981. 2.3

[34] Christos Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Computer vision and image understanding : CVIU*, 66 2:207–22, 1997. 2.3

[35] Richard D. Rabbitt, Jeffrey A. Weiss, G. Christensen, and Michael I. Miller. Mapping of hyperelastic deformable templates using the finite element method. In *Optics & Photonics*, 1995. 2.3

[36] Gary E Christensen and Hans J Johnson. Consistent image registration. *IEEE transactions on medical imaging*, 20(7):568–582, 2001. 2.3

[37] Yongmei Michelle Wang and Lawrence H. Staib. Physical model-based non-rigid registration incorporating statistical shape information. *Medical image analysis*, 4 1: 7–20, 2000. 2.3

[38] Xavier Pennec, Pascal Cathier, and Nicholas Ayache. Understanding the "demon's algorithm": 3d non-rigid registration by gradient descent. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1999. 2.3

[39] F.L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6): 567–585, 1989. doi: 10.1109/34.24792. 2.3

[40] Thomas W. Sederberg and Scott R. Parry. Free-form deformation of solid geometric models. In *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '86, page 151–160, New York, NY, USA, 1986. Association for Computing Machinery. ISBN 0897911962. doi: 10.1145/15922.15903. URL https://doi.org/10.1145/15922.15903. 2.3

[41] William M. Hsu, John F. Hughes, and Henry Kaufman. Direct manipulation of free-form deformations. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '92, page 177–184, New York, NY, USA, 1992. Association for Computing Machinery. ISBN 0897914791. doi: 10.1145/133994.134036. URL https://doi.org/10.1145/133994.134036. 2.3

[42] Jérôme Declerck, Jacques Feldmar, Fabienne Betting, and Michael L. Goris. Auto-

matic registration and alignment on a template of cardiac stress and rest spect images. *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 212–221, 1996. 2.3

[43] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John V. Guttag, and Adrian V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *CoRR*, abs/1809.05231, 2018. URL http://arxiv.org/abs/1809.05231. 2.3

[44] Théo Estienne, Maria Vakalopoulou, Enzo Battistella, Alexandre Carré, Théophraste Henry, Marvin Lerousseau, Charlotte Robert, Nikos Paragios, and Eric Deutsch. Deep learning based registration usingÂ spatial gradients and noisy segmentation labels. In Nadya Shusharina, Mattias P. Heinrich, and Ruobing Huang, editors, *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data*, pages 87–93, Cham, 2021. Springer International Publishing. ISBN 978-3-030-71827-5. 2.3

[45] Marc-Michel Rohé, Manasi Datar, Tobias Heimann, Maxime Sermesant, and Xavier Pennec. Svf-net: Learning deformable image registration using shape matching. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention âˆ’ MICCAI 2017*, pages 266–274, Cham, 2017. Springer International Publishing. ISBN 978-3-319-66182-7. 2.3

[46] Hessam Sokooti, B. D. Vos, Floris F. Berendsen, Boudewijn P. F. Lelieveldt, Ivana Igum, and Marius Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017. 2.3

[47] Hristina Uzunova, Matthias Wilms, Heinz Handels, and Jan Ehrhardt. Training cnns for image registration from few samples with model-based data augmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017. 2.3

[48] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. 2.3, 4.2.1, 4.4.3, 4.3, 4.4, 4.5

[49] Chengqian Che, Tejas Sudharshan Mathai, and John Galeotti. Ultrasound registration: A review. *Methods*, 115:128–143, 2017. 2.3, 4.1, 4.2.1

[50] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208:106236, 2021. 2.4

[51] Bhushan D Patil, Vanika Singhal, Utkarsh Agrawal, Rajesh Langoju, Jiang Hsieh, Shobana Lakshminarasimhan, and Bipul Das. Deep learning based correction of low performing pixel in computed tomography. *Biomedical Physics Engineering Express*, 8(5):055027, aug 2022. doi: 10.1088/2057-1976/ac87b4. URL https://dx.doi.org/10.1088/2057-1976/ac87b4. 2.4

[52] Jue Jiang, Yu-Chi Hu, Neelam Tyagi, Pengpeng Zhang, Andreas Rimner, Gig S Mageras, Joseph O Deasy, and Harini Veeraraghavan. Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pages 777–785. Springer, 2018. 2.4

[53] Anmol Sharma and Ghassan Hamarneh. Missing mri pulse sequence synthesis using multi-modal generative adversarial network. *IEEE transactions on medical imaging*, 39(4):1170–1183, 2019. 2.4

[54] Mostafa Salem, Sergi Valverde, Mariano Cabezas, Deborah Pareto, Arnau Oliver, Joaquim Salvi, Àlex Rovira, and Xavier Lladó. Multiple sclerosis lesion synthesis in mri using an encoder-decoder u-net. *IEEE Access*, 7:25171–25184, 2019. 2.4

[55] Kenny H Cha, Nicholas Petrick, Aria Pezeshk, Christian G Graff, Diksha Sharma, Andreu Badal, and Berkman Sahiner. Evaluation of data augmentation via synthetic images for improved breast mass detection on mammograms using deep learning. *Journal of Medical Imaging*, 7(1):012703–012703, 2020. 2.4

[56] Ken Shoemake. Animating rotation with quaternion curves. *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, 1985. 3.1

[57] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, and Achintya Bhowmik. Intel realsense stereoscopic depth cameras, 2017. 3.2.1

[58] Stanford Artificial Intelligence Laboratory et al. Robotic operating system. URL https://www.ros.org. 3.2.1

[59] Nico Zevallos, Evan Harber, Kirtan Patel, Yizhu Gu, Kenny Sladick, Francis Guyette, Leonard Weiss, Michael R Pinsky, Hernando Gomez, John Galeotti, et al. Toward robotically automated femoral vascular access. In *2021 International Symposium on Medical Robotics (ISMR)*, pages 1–7. IEEE, 2021. 3.2.2, 5.1

[60] Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023. 3.2.2

[61] Raghavv Goel, Fnu Abhimanyu, Kirtan Patel, John Galeotti, and Howie Choset. Autonomous ultrasound scanning using bayesian optimization and hybrid force control. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8396–8402. IEEE, 2022. 3.2.2, 4.3.1

[62] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, page 226–231. AAAI Press, 1996. 3.2.2

[63] Ananya Bal, Ashutosh Gupta, Fnu Abhimanyu, John Galeotti, and Howie Choset. A curvature and trajectory optimization-based 3d surface reconstruction pipeline for ultrasound trajectory generation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2724–2730, 2023. doi: 10.1109/ICRA48891.2023.10161513. 3.2.2

[64] Maria Victorova, David Navarro-Alarcon, and Yong-Ping Zheng. 3d ultrasound imag-

ing of scoliosis with force-sensitive robotic scanning. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pages 262–265. IEEE, 2019. 3.2.3

[65] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960. ISSN 0021-9223. doi: 10.1115/1.3662552. URL https://doi.org/10.1115/1.3662552. 3.2.3

[66] Mojtaba Akbari, Jay Carriere, Tyler Meyer, Ron Sloboda, Siraj Husain, Nawaid Usmani, and Mahdi Tavakoli. Robotic ultrasound scanning with real-time image-based force adjustment: Quick response for enabling physical distancing during the covid-19 pandemic. *Frontiers in Robotics and AI*, 8, 2021. ISSN 2296-9144. doi: 10.3389/frobt.2021.645424. URL https://www.frontiersin.org/articles/10.3389/frobt.2021.645424. 3.3.1

[67] Felix von Haxthausen, Sven Böttger, Daniel Wulff, Jannis Hagenah, Verónica García-Vázquez, and Svenja Ipsen. Medical robotics for ultrasound imaging: current systems and future trends. *Current robotics reports*, 2:55–71, 2021. 3.3.2

[68] Mojtaba Akbari, Jay Carriere, Tyler Meyer, Ron Sloboda, Siraj Husain, Nawaid Usmani, and Mahdi Tavakoli. Robotic ultrasound scanning with real-time image-based force adjustment: quick response for enabling physical distancing during the covid-19 pandemic. *Frontiers in Robotics and AI*, 8:645424, 2021. 3.3.2

[69] Monan Wang and Li Pengcheng. A review of deformation models in medical image registration. *Journal of Medical and Biological Engineering*, 39, April 2018. doi: 10.1007/s40846-018-0390-1. 4.1

[70] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: a survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, July 2013. ISSN 1558-254X. doi: 10.1109/TMI.2013.2265603. Conference Name: IEEE Transactions on Medical Imaging. 4.1

[71] Hamid Reza Boveiri, Raouf Khayami, Reza Javidan, and Alireza Mehdizadeh. Medical image registration using deep neural networks: a comprehensive review. *Computers & Electrical Engineering*, 87:106767, October 2020. ISSN 00457906. doi: 10.1016/j.compeleceng.2020.106767. 4.1

[72] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. Deep learning in medical image registration: a review. *Physics in Medicine & Biology*, 65(20):20TR01, October 2020. ISSN 1361-6560. doi: 10.1088/1361-6560/ab843e.

[73] Jing Zou, Bingchen Gao, Youyi Song, and Jing Qin. A review of deep learning-based deformable medical image registration. *Frontiers in Oncology*, 12:1047215, December 2022. ISSN 2234-943X. doi: 10.3389/fonc.2022.1047215. 4.1

[74] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 4.2.1

[75] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. PWC-Net: CNNs for

optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 4.2.1

[76] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 4.2.1

[77] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 4.2.1, 4.3.2, 5.2.1

[78] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004. 4.2.2

[79] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 4.2.2, 4.4.3

[80] Léon Bottou. Stochastic gradient descent tricks. *Neural Networks: Tricks of the Trade: Second Edition*, pages 421–436, 2012. 4.3.2

[81] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. *NIPS 2017 Autodiff Workshop*, 2017. 4.3.2, 5.2.1

[82] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019. 4.4.1

[83] E. O. Brigham and R. E. Morrow. The fast fourier transform. *IEEE Spectrum*, 4(12): 63–70, 1967. doi: 10.1109/MSPEC.1967.5217220. 4.4.2

[84] Sven Ivar Seldinger. Catheter replacement of the needle in percutaneous arteriography: a new technique. *Acta radiologica*, 49(sup434):47–52, 2008. 5.1

[85] Laura Brattain, Theodore Pierce, Lars Gjesteby, Matthew Johnson, Nancy DeLosa, Joshua Werblin, Jay Gupta, Arinc Ozturk, Xiaohong Wang, Qian Li, Brian Telfer, and Anthony Samir. AI-enabled, ultrasound-guided handheld robotic device for femoral vascular access. *Biosensors*, 11(12):522, December 2021. ISSN 2079-6374. doi: 10.3390/bios11120522. 5.1

[86] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 5.1, 5.3.2

[87] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-Net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access*, 9:82031–82057, 2021. 5.1

[88] Edward Chen, Howie Choset, and John Galeotti. Uncertainty-based adaptive data augmentation for ultrasound imaging anatomical variations. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 438–442. IEEE, 2021.

5.1

[89] Nils Marahrens, Bruno Scaglioni, Dominic Jones, Raj Prasad, Chandra Shekhar Biyani, and Pietro Valdastri. Towards autonomous robotic minimally invasive ultrasound scanning and vessel reconstruction on non-planar surfaces. *Frontiers in Robotics and AI*, page 178, 2022.

[90] Cecilia Morales, Jason Yao, Tejas Rane, Robert Edman, Howie Choset, and Artur Dubrawski. Reslicing ultrasound images for data augmentation and vessel reconstruction. In *2023 International Conference on Robotics and Automation (ICRA)*. IEEE, 2023 (Accepted). 5.1

[91] Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. *CoRR*, abs/2003.01460, 2020. URL https://arxiv.org/abs/2003.01460. 5.2.1

[92] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai kin Wong, and Wang chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting, 2015. 5.2.1

[93] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. 5.2.1