

Dissertation

Physics-Inspired Machine Intelligence for Robotic Ultrasound Imaging

Maria Tirindelli



School of Computation
Information and Technology
Technische Universität München





Technische Universität München
TUM School of Computation, Information and Technology

Physics-Inspired Machine Intelligence for Robotic Ultrasound Imaging

Maria Tirindelli

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Julien Gagneur

Prüfer der Dissertation: Prof. Dr. Nassir Navab

Prof. Dr. Purang Abolmaesumi

Die Dissertation wurde am 21.02.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 13.07.2023 angenommen.

Maria Tirindelli

Physics-Inspired Machine Intelligence for Robotic Ultrasound Imaging

Dissertation, Version 0.0

Technische Universität München

School of Computation, Information and Technology

Boltzmannstraße 3

85748 and Garching bei München

Abstract

Ultrasound imaging is a well-established imaging technique allowing for real-time tissue imaging without exposure to ionizing radiation. The automation of ultrasound acquisitions through medical robots is a promising technique that has been recently explored in the scientific literature. The advantages of automatizing ultrasound acquisitions are manifold. Firstly, it potentially constitutes a precious aid for novice radiologists, thus mitigating the operator dependence in ultrasound acquisition and interpretation. Secondly, it frees the operator from the burden to manipulate the ultrasound probe while performing surgical tasks.

Compared to other medical robots, the design and implementation of robotic-ultrasound systems present many challenges. Manual ultrasound acquisitions and interpretation, are complex tasks that require understanding and identification of anatomical features in presence of noise and artifacts and different image appearances, correct mapping of these features in an anatomical consistent fashion, and finally capability of navigating the probe to maximize the image quality. All these skills are acquired by sonographers during their studies and careers, and all of them need to be transferred to the robotic-ultrasound system, to enable smooth navigation of the ultrasound transducer, maximizing the quality and information content of the acquired data.

The aim of this dissertation is to examine the challenges of robotics-ultrasound systems. Within this work, we explore different methodologies for ultrasound image processing and their integration within visual servoing systems for robotic navigation. Additionally, we discuss the limitations of using limited and variable ultrasound data for network training and we analyze potential solutions through data augmentation are discussed.

Zusammenfassung

Die Ultraschallbildgebung ist ein bewährtes bildgebendes Verfahren, das die Darstellung von Gewebe in Echtzeit und ohne die Exposition von ionisierender Strahlung erlaubt. Die Automatisierung von Ultraschalluntersuchungen durch Medizinroboter ist eine neue vielversprechende Technik mit vielfältigen Vorteilen. Erstens stellt sie eine wertvolle Hilfe für unerfahrene Radiologen dar und verringert so die Variabilität zwischen Untersuchungen von verschiedenen Bedienern für die Erfassung und Auswertung. Zweitens wird der Bediener von der Last befreit, die Ultraschallsonde während der Durchführung chirurgischer Aufgaben manuell zu führen.

Im Vergleich zu anderen medizinischen Robotern stellen sich bei der Entwicklung und Implementierung von Roboter-Ultraschallsystemen viele Herausforderungen. Manuelle Ultraschallaufnahmen und -interpretationen sind komplexe Aufgaben, die das Verständnis und die Identifizierung anatomischer Merkmale in Gegenwart von Rauschen, Artefakten und unterschiedlichen Bilderscheinungen erfordern. Zusätzlich erfordert es die korrekte Abbildung dieser Merkmale in einer anatomisch konsistenten Weise und schließlich die Fähigkeit zur Navigation der Sonde um die Bildqualität zu maximieren. All diese Fähigkeiten werden von Sonographen während ihrer Ausbildung und ihrer beruflichen Laufbahn erworben und müssen auf das Roboter-Ultraschallsystem übertragen werden, um eine reibungslose Navigation des Ultraschallwandlers zu ermöglichen und die Qualität und den Informationsgehalt der erfassten Daten zu maximieren.

Das Ziel dieser Dissertation ist es, die Herausforderungen von Roboter-Ultraschallsystemen zu untersuchen. Im Rahmen dieser Arbeit untersuchen wir verschiedene Methoden der Ultraschallbildverarbeitung und deren Integration in visuelle Servosysteme für die Roboternavigation. Darüber hinaus werden die Limitationen der Verwendung begrenzter und variabler Ultraschalldaten für das Netzwerktraining erörtert und mögliche Lösungen durch Datenerweiterung analysiert.

Acknowledgement

This work was supported by the Bayerische Forschungsstiftung, under Grant DOK-180-19.

Contents

I	Introduction	1
1	List of Authored and Co-authored Publications	3
2	Introduction	5
2.1	Motivation and Problem Statement	6
2.2	Thesis Structure	7
3	Related Work	11
4	Robot Control	15
4.1	Robot Arm Manipulators as a Dynamical System	15
4.1.1	Useful Definitions	16
4.1.2	Lagrange derivation of the Robot manipulator dynamic equations	17
4.2	Control Strategies for Robot manipulators	21
4.2.1	Position Control	21
4.2.2	Stiffness Control	21
4.2.3	Impedance Control	22
4.2.4	Force Control	23
4.2.5	Hybrid Control	24
4.3	Visual Servoing in Robotics	24
5	Ultrasound Imaging	27
5.0.1	Ultrasound Imaging - Image Formation Principles	27
5.0.2	Ultrasound Imaging - Noise and Artifacts	35
5.0.3	Ultrasound Imaging and Deep Learning - Data Augmentation	39
5.1	Ultrasound 3D Volume Compounding	41
5.1.1	Voxel-based methods	41
5.1.2	Pixel-based methods	41
II	Contributions	43
6	Visual-Tactile Fusion for Autonomous Robotic Scanning	45
6.1	Introduction and Motivation	46
6.1.1	Current State of Automatic Vertebral Classification and Robotic Ultrasound for Spinal Injections	46

6.1.2	Fusing Haptic and Visual Feedback for Robotic-Ultrasound Scanning	47
6.2	Fusing Haptics and Vision for Robot Control - Method	48
6.2.1	Useful Notation and Anatomy	48
6.2.2	System Design	48
6.2.3	Autonomous Scanning Pipeline	50
6.2.4	Data Acquisition for Parameters Tuning and Model Training	50
6.2.5	Force Data Extraction	52
6.2.6	Ultrasound Data Extraction and Processing	54
6.2.7	Haptic-Visual Data Fusion	55
6.2.8	Facet Plane identification	57
6.3	Experiments	57
6.3.1	Ultrasound Signal Processing	58
6.3.2	Force-Ultrasound Data Fusion	58
6.3.3	Facet Plane Identification	61
6.4	Conclusions	61
7	Physics-Inspired Augmentation	63
7.1	Introduction and Motivation	64
7.1.1	Current State of Augmentation Approaches in Ultrasound Imaging	64
7.1.2	Intelligent Augmentation by Leveraging Physics of Ultrasound	65
7.1.3	Contributions	66
7.2	Physics-inspired Data Augmentation	66
7.2.1	Augmentation by Deformation	67
7.2.2	Augmentation by Reverberation	69
7.2.3	Augmentation by SNR Modification	69
7.3	Experiments	71
7.3.1	Classification	71
7.3.2	Segmentation	71
7.3.3	Dataset Acquisition	72
7.3.4	Implementation Details	72
7.3.5	Neural Network Architectures and Training	73
7.3.6	Evaluation Metrics	73
7.4	Results and Discussion	73
8	Path Optimization	75
8.1	Robotic Path Optimization	76
8.1.1	Introduction and Motivation	76
8.2	Methodology	77
8.2.1	Volume Coverage Optimizer	78
8.2.2	Shadow Reduction Optimizer	79
8.3	Experiments and Results	80

8.3.1 Simulation Experiments	81
8.3.2 Real Robot Experiment	83
8.4 Conclusion	84

III Conclusions 85

9 Conclusion and Future Directions 87

Bibliography 89

List of Figures 101

List of Tables 105

Part I

Introduction

List of Authored and Co-authored Publications

Included in this Dissertation for Grading

- [91] **Tirindelli M***, Victorova M*, Esteban J, Kim ST, Navarro-Alarcon D, Zheng YP, Navab N. Force-ultrasound fusion: Bringing spine robotic-us to the next “level”. *IEEE Robotics and Automation Letters*. 2020 Jul 14;5(4):5661-8.
- [115] **Tirindelli M***, Eilers C*, Simson W, Paschali M, Azampour MF, Navab N. Rethinking ultrasound augmentation: A physics-inspired approach. In *International Conference on Medical Image Computing and Computer-Assisted Intervention 2021 Sep 27* (pp. 690-700). Springer, Cham.
- [121] Sutedjo V*, **Tirindelli M***, Eilers C*, Simson W, Busam B, Navab N. Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark. *IEEE Robotics and Automation Letters*. 2022 Jan 10;7(2):1808-15.

Other Articles

- [61] Denasi A, Khan F, Boskma KJ, Kaya M, Hennersperger C, Göbl R, **Tirindelli M**, Navab N, Misra S. An observer-based fusion method using multicore optical shape sensors and ultrasound images for magnetically-actuated catheters. In *2018 IEEE International Conference on Robotics and Automation (ICRA) 2018 May 21* (pp. 50-57). IEEE.
- [82] Hase H, Azampour MF, **Tirindelli M**, Paschali M, Simson W, Fatemizadeh E, Navab N. Ultrasound-guided robotic navigation with deep reinforcement learning. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2020 Oct 24* (pp. 5534-5541). IEEE.
- [119] Ronchetti M, Rackerseder J, **Tirindelli M**, Salehi M, Navab N, Wein W, Zettinig O. PRO-TIP: Phantom for ROBust automatic ultrasound calibration by TIP detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention 2022* (pp. 84-93). Springer, Cham.

Introduction

Ultrasound imaging is a simple, flexible, and comparatively cheap medical imaging modality for image-based guidance and diagnostics [4], [14], [26], [31], [68], [118]. Compared to other imaging modalities such as CT and X-ray, ultrasound imaging offers a range of advantages including flexibility, lower cost, the absence of ionizing radiation, and dedicated rooms. However, it also presents a number of challenges, such as the presence of noise and artifacts [1], [2] and the low signal-to-noise ratio, which can make the data interpretation difficult. Moreover, the manual acquisition of ultrasound data renders the acquisition procedure highly dependent on the skills of the operator and also affects their capability to perform other tasks simultaneously, such as surgical resection or needle injections [15], [105].

Robotics-ultrasound systems have been proposed as a way to address the operator dependence of traditional ultrasound, improve repeatability and reduce fatigue [6]. These systems typically consist of an ultrasound probe mounted on a robotic arm, automatically driven to acquire a target and provide real-time feedback to the robot during a procedure. This eliminates the need for the operator to manually handle the ultrasound probe, allowing them to focus on other tasks [6], [10], [62].

Robotic-ultrasound systems face a number of challenges. While improving the repeatability of the ultrasound acquisition, such systems must also ensure patient comfort and safety and avoid losing image quality [9]. This requires careful monitoring of the robot-patient interactions and the implementation of a control strategy that ensures the safety of the procedure. Additionally, the robot's control strategy must allow for fine-tuning of its motion based on multiple external feedback sources, such as images and sensor data, to achieve a high-quality of the acquired data that is sufficient for proper processing and interpretation [52], [102]. Gaining an understanding of how to define a control strategy for the robot is not trivial. To explore potential solutions and gain insight into this challenge, it may be helpful to examine the ways in which expert sonographers control and coordinate their motions while performing ultrasound acquisitions. It is worth noting that expert sonographers do not simply follow a predetermined trajectory when manipulating the ultrasound probe; rather, they leverage a variety of factors during the acquisition process, including visual information displayed on the monitor, their variation with respect to the probe orientation and motion, and other sensory inputs like tactile feedback [27], [36]. For instance, they

utilize their understanding of anatomy to anticipate where certain features should appear, using this understanding to guide probe navigation. In addition, they take advantage of the knowledge they have acquired through experience with ultrasound systems, such as the recognition that maintaining a perpendicular orientation to highly reflective surfaces like bone can enhance image contrast and reduce blurriness [27]. The operator may also complement the visual feedback by utilizing techniques such as palpation to identify specific features that may not be clearly visible on the ultrasound data and adjust the probe accordingly [30], [58]. Finally, the operator has developed expertise in interpreting and "reading" ultrasound data, which requires a high level of experience and the ability to identify features irrespective of the particular scanner and parameters utilized, patient anatomy, and the presence of noise and artifacts.

In this dissertation, we examine and discuss how the incorporation of and leveraging physical properties present during ultrasound acquisition, in a manner similar to how sonographers do in practice.

2.1 Motivation and Problem Statement

The main objective of this thesis is to analyze the various components and aspects that make up a robotic ultrasound system, with a focus on ultrasound data processing and visual servoing. In doing so, we take inspiration from the strategies used by sonographers and we aim to identify the skills that we need to impart to our system.

To achieve this, we focus on the following key questions:

- How can we integrate visual and non-visual feedback into the robotic control in a way that allows the robot to effectively utilize all available information to navigate the patient's body? In this dissertation work, we address this question by proposing a new controller design, taking inspiration from the way sonographers integrate tactile and visual information in order to gain robust and reliable insights into the patient's anatomy. Similarly, our proposed controller exploits temporal deep learning techniques to leverage both vision and force feedback to achieve this.
- How can we ensure that the system is generalizable, even with a limited amount of data, and how can we address the problem of limited data in the case of ultrasound imaging? In this dissertation, we address the issue of data augmentation by proposing a new technique that defines data-specific augmentations based on the physics of the generation of ultrasound data. We further discuss the potential of augmentation techniques that aims at more realistically increasing

the diversity of the data and simulating the sources of variability encountered in clinical practice.

- How can the robot learn to use and refine the anatomical prior based on the information gathered from ultrasound acquisition, and how can we teach the system to navigate the ultrasound probe based on image feedback in order to maximize image quality? In this dissertation work, we address this question by designing a new trajectory generation method, that estimates the optimal probe orientations, ensuring both the best volume coverage and minimizing the effect of acoustic shadowing on the compounded 3D volume.

2.2 Thesis Structure

This thesis is divided into five chapters.

Chapter 1, 2, 3: List of Authored and Co-authored Publications, Introduction and Related Work

The first chapter presents the list of publications of the candidate, including those utilized for the grading of this thesis as well as those not included. The second chapter offers an introduction to the work presented, including a general overview of the area of robotic ultrasound and a brief introduction to the open research questions that are addressed or discussed within this manuscript.

Chapter 3: Robot Control

Chapter 3 presents a more detailed description of the more common control modes in robotics, with a particular focus on control modes commonly used in medical imaging and robotic ultrasound. The aim of the chapter is to provide an understanding of the mathematical equations defining robot dynamics and how control modes are constructed on top of these equations to achieve desired robot behaviors. The integration of such control modes with visual servoing methodologies is also introduced. Finally, the chapter discusses how visual/haptic feedback can be further exploited to refine the robot control, and how this relates to control strategies adopted by sonographers.

Chapter 4: Ultrasound Imaging

Chapter 4 provides an overview of ultrasound image formation, starting from the generation and recording of ultrasound waves and proceeding to the compounding of 2D images and 3D volumes. The chapter begins by discussing the physics of ultrasound image formation and its relationship to noise and artifacts characterizing this imaging modality. It then introduces the challenges of training deep learning models on ultrasound data, given the limited availability of (public) ultrasound datasets and the wide variability of such data, and finally discusses the impact and the potential of data augmentation as a means of addressing these limitations.

Chapter 5: Force-Ultrasound Fusion

Chapter 5 discusses visual-tactile integration for robot control, in relation to the operator's multisensory control of the ultrasound probe. The work focuses on the specific application of spinal injections. The chapter presents a system that aims to automatically navigate the robot and identify vertebral levels, with the goal of ensuring safe and precise lumbar injections [91]. During the robot scanning, both force and imaging data are acquired. The method leverages the capability of ultrasound to provide information about internal anatomy and of force sensors to provide information on the tactile appearance of the spine anatomy. To account for the temporal nature of both data, temporal convolutional networks are utilized to fuse the data modalities. Compared to state-of-the-art methods, the proposed method utilizes a combination of ultrasound and force data, which allows it to provide the robot with both visual and tactile feedback during the scanning, while other methods rely solely on either ultrasound or X-ray data, and do not provide tactile feedback. Secondly, the use of deep learning for processing ultrasound data and a temporal convolutional network for data fusion allows for improved accuracy and robustness in the identification of vertebral levels. Overall, the proposed method has the potential to improve the precision and safety of lumbar spinal injections, while reducing the reliance on ionizing radiation and the operator's manual skills.

Chapter 6: Physics-Inspired Augmentation

Chapter 6 discusses the challenges of training deep learning models on ultrasound data, given the limited availability of open datasets and the high variability of ultrasound data. To address these challenges, the use of data augmentation techniques that accounts for the physical characteristics of ultrasound images is proposed and discussed in relation to standard augmentation techniques which are instead optimized for photographic images [115]. The proposed techniques include deformation,

reverb, and signal-to-noise ratio adjustments, and are applied to US B-mode images to improve the generalization and performance of deep neural networks. The suitability of these physics-inspired data augmentation techniques for ultrasound images is compared to more commonly used techniques, which do not reflect the physical properties of ultrasound and can generate unrealistic images if used improperly. The effectiveness of these techniques is evaluated for the tasks of bone classification and segmentation on a spine ultrasound dataset. The potential applications of the proposed method are manifold, including the creation of autonomous applications such as robotic ultrasound acquisitions.

Chapter 7: Path Optimization

Chapter 7 discusses optimization strategies to update robot trajectories in order to maximize the quality of the ultrasound images for 3D compounding. Specifically, it focuses on the challenges posed by the occurrence of acoustic shadowing, which occurs in presence of tissue interface with very different acoustic impedance and can obscure the structures beneath. This is a particular concern in diagnostic imaging or image guidance for the heart and liver, such as radiotherapy and radiosurgery. When these procedures are performed manually, clinicians typically identify optimal probe positions to minimize shadowing artifacts caused by the rib cage. When it comes to automated, robotic acquisitions, smart trajectory planning methods potentially improve image quality while reducing ultrasound acquisition time, thus enabling its use in new medical tasks. Chapter 7 presents a novel method for reducing the effect of acoustic shadowing in ultrasound image acquisition and volume compoundings through the optimization of robotic probe poses [121]. The proposed pipeline is modular and adaptable to specific tasks, regions of interest, or imaged structures, and does not require physician assistance, external sensors, or pre-acquired data. The major contributions of the proposed method include the development of a modular optimization pipeline for robotic-ultrasound acquisitions and a volume coverage optimizer for maximizing the ultrasound coverage of the anatomy of interest. The proposed method is evaluated against standard (i.e., perpendicular) and random ultrasound scans in simulations, and initial results show the feasibility of translation to real case scenarios.

Chapter 8: Conclusions and Future Directions

Chapter 8 presents the conclusion and discussion of the works presented within this dissertation.

Related Work

Medical robotics is a well-established field that is experiencing significant growth and development across a wide range of applications [10], [18], [53], [81], [99]. Medical robots are commonly used for diagnostics and treatments, both in telerobotic mode and as semi- or fully autonomous systems [59], [100]. Robotics-ultrasound systems are a subset of medical robots and are mainly used for image-based diagnostics and guidance in surgery [8], [11]. The application of robotic-ultrasound systems has been described in the literature in various fields such as orthopedics [102], neurosurgery [62], [91], vascular imaging [21], [39], [48], [73], [104], [112], [114], cardiac imaging [49], as well as breast [69], [94], prostate [29], [54], and abdominal imaging [32], [49], [94].

One of the key challenges in the development of robotics-ultrasound systems is the design of an effective control strategy that combines a suitable force control setup with a visual servoing methodology enabling the system to correctly interpret ultrasound data and respond (i.e. move/act) appropriately based on this interpretation. Researchers have therefore been seeking to develop techniques that enable the systems to interpret ultrasound data and adjust their behavior accordingly, as well as ensure the safety and comfort of the patient [9], [34], [108].

Robotic ultrasound systems typically consist of a robotic arm with an ultrasound transducer mounted on top of it via a holder. Some systems also incorporate sensors such as RGB-D cameras to provide visual feedback before contact with the patient's skin, thereby avoiding potential collisions [32], [43], [44], [56], [63], [103], [114]. The interaction between the system and the patient is a delicate matter, as the contact between the transducer and the patient's skin must be properly tuned to ensure both patient comfort and proper coupling between the ultrasound probe and the skin. To achieve this, many researchers have suggested using a force control scheme along the z-axis of the robot [44], [48], [73], [82], [104], [112]. Force feedback can be further utilized to receive tactile feedback, and/or to estimate the mechanical properties of the examined tissues, thus better characterizing them [47], [57], [63], [65], [85], [86], [104]. The control of the other degrees of freedom varies more widely depending on the specific application. Generally, an initial trajectory or pose is computed based on anatomical priors obtained from pre-operative MR or CT imaging, or by anatomical landmarks previously selected by the operator [44], [49], [62],

[94]. Once the robot has reached the predefined position, the feedback coming from the ultrasound image is used to adjust the robot's position. The way the image is processed to compute a desired input robot position highly depends on the application. Some authors propose to exploit the noise and artifacts characterizing ultrasound imaging, to compute the control variables to navigate the robot [25], [37], [98]. As an example, confidence maps can be used to quantify acoustic shadowing, in a way to navigate the robot to poses where the tissues are more visible [37], [85]. In other cases, image segmentation and/or feature extracted from the ultrasound data are used to compute the relative displacement from a desired plane/pose, thus to compute the necessary robot motion to reach it [23], [28], [29], [33], [39], [42], [46], [73], [91]. Deep reinforcement learning methods have also been proposed for navigating the robot based on input b-mode images [82], [107], [109], [112], [116]. In these cases, the state vector is one or more b-mode data, and the reward for each action is based on the distance gained from the target location. With the advent of deep learning, many authors suggested using deep networks to process ultrasound images and extract the robot control parameters [91], [96], [123]. Deep Learning methods are also often used within robotic ultrasound systems for segmenting the acquired ultrasound data, in a way to ease the operator in data interpretation [75], [97], [103], [104], [123].

Deep Learning for ultrasound data segmentation, classification, and analysis has been widely described in the literature [74], [88], [93], [120], [122]. However, training deep learning models on ultrasound data is a challenging fact, due to different factors: *i)* Ultrasound data acquisition is time-consuming, and there exists very few openly available datasets [70]. *ii)* The appearance of ultrasound data largely varies among ultrasound scanners, between patients, and depending on the acquisition parameters and operator skills [113]. Therefore, it is challenging to train a model able to generalize well over the whole ultrasound data distribution. *iii)* Data annotation is challenging due to the already described difficulties in interpreting ultrasound data. To overcome these problems, some authors proposed to use augmentation techniques [76], to artificially increase the size of the ultrasound dataset [117]. Data augmentation techniques can be categorized into two main groups: synthetic data generation and image modification. Synthetic data generation involves the use of models such as generative adversarial networks (GANs) to generate additional data [70], [79], [90], [95], [106], [110], [111]. However, the effectiveness of this approach may be limited by the amount of available data for training the GAN. In contrast, image modification involves applying classical image modifications, such as translation, scaling, rotation, flipping, brightness variations, and noise addition, to the existing dataset. This approach has been employed in the context of ultrasound imaging for tasks such as bone segmentation [51], [60], [64], [67], [71], [78], [80], [83], [87], [89], [92].

It is worth noting that the presence of robotic arms in the system allows for the acquired ultrasound data to also include tracking information. As a result, several researchers have proposed combining these images into 3D data in order to create a volumetric representation of the explored anatomy and enable 3D to 3D registration with pre-operative data [21], [39], [43], [50], [65], [94], [103], [104], [114], [123]. The process of combining ultrasound images into 3D data, known as 3D volume compounding, has been widely studied in both free-hand and robotic ultrasound [3]. It involves recombining the B-mode data into a volumetric representation, which has the advantage of reducing noise and providing enhanced information from the acquisition. Additionally, it allows for registration with other 3D data [5]. However, ultrasound compounding can come at the cost of introducing new artifacts or enhanced shadowed regions [17], [101].

Robot Control

In the context of robot control, we typically refer to strategies for controlling the joint torque in order to achieve the desired configuration of the robot joints. To effectively implement these strategies, it is necessary to understand the dynamics of the robot, thus to derive its dynamics equations. This chapter presents a description of the derivation of the dynamic equations for a robot arm manipulator with a fixed extremity, followed by an overview of the main control strategies: position, stiffness, impedance, and force control. In addition, the chapter covers the integration of visual servoing into the control scheme and how sensor information can be used in combination to gain insights into tissue anatomy.

4.1 Robot Arm Manipulators as a Dynamical System

In order to understand the dynamics of a robot and manipulate its joints to achieve a specific behavior, it is necessary to analyze the robot as a dynamical system. A dynamical system can be defined as a system that changes over time and can be represented mathematically by a set of differential equations. The derivation of these equations is essential for analyzing the system's behavior and its relationships with the physical quantities involved. For instance, in the case of a point mass subjected to a constant force F , Newton's laws dictate that the dynamics of the system can be described by the following differential equation:

$$\mathbf{F} = m\mathbf{a} = m\frac{\partial\mathbf{v}}{\partial t} = m\frac{\partial^2\mathbf{x}}{\partial t^2} \quad (4.1)$$

where \mathbf{x} , \mathbf{v} , \mathbf{a} are the point position, velocity, and acceleration, respectively, and m the point mass. The differential equation in Equation 4.1 describes how the position of a point mass will change over time in response to the applied force F . While robots are not point masses and their dynamics are typically more complex, understanding the dynamics equations governing their motion is crucial for predicting their response to input joint forces (or torques). These equations, which uniquely describe the behavior of the robot over time, can be derived using Lagrange mechanics. It is worth noting

that both Newton and Lagrange mechanics can be used to derive these equations and that the latter can be derived from the former. For further information on the derivation of Lagrange mechanics, the reader is referred to [13].

4.1.1 Useful Definitions

Before proceeding to the derivation of the dynamic equations for a robot, it is necessary to introduce some key concepts and definitions related to the use of Lagrange's mechanics.

In a multi-body system consisting of n particles, the following quantities can be defined:

Generalized Coordinates

A set of parameters that completely specify the configuration of a physical system

Generalized Forces

Given a set of forces acting on the particles of the system, generalized forces are defined as:

$$Q_j = \sum_{i=1}^n F_i \frac{\partial r_i}{\partial q_j} \quad \text{for } j = 1, \dots, m \quad (4.2)$$

where r_i is the position vector of i_{th} particle of the system, F_i the resultant of the forces acting on the i_{th} particle, and q_1, \dots, q_m are the generalized coordinate for that system.

Note: In simpler terms, generalized forces can be thought of as an extension of the Newtonian concept of forces for multi-body systems, where it is useful to describe the system using a set of generalized coordinates. For example, a robot arm manipulator can be viewed as a multi-body system, and a convenient choice of generalized coordinates is the joint angles. These angles satisfy the definition of generalized coordinates, as knowing their values allows us to uniquely determine the configuration of the robot. The concept of generalized forces is less intuitive, but they can be understood as representing the component (or projection) of the forces acting on the robot along the generalized coordinates. As an example, the generalized forces acting on a robot manipulator might be the torques applied at each of its joints (that is, the rotational forces applied by the joint motors) or the torques generated by the gravitational force acting on the robot's joints.

Jacobian Matrix

In the context of a robot arm manipulator, what does the Jacobian matrix represent? Consider the i_{th} joint and let \mathbf{r}_i be the location of the link centroid in the base coordinate system. Let the set of generalized coordinates q_1, \dots, q_m represent the link angles. The Jacobian matrix defines the transformation between the generalized coordinates q_1, \dots, q_m and the locations \mathbf{r}_i of the i_{th} robot's links.

$$d\mathbf{r}_i = \mathbf{J}_i d\mathbf{q} \quad (4.3)$$

where \mathbf{J} is the Jacobian for the i_{th} link and $\mathbf{q} = [q_1, \dots, q_m]$. We can rewrite equation 4.4 to make the Jacobian explicit as:

$$\mathbf{J}_i = \frac{d\mathbf{r}_i}{d\mathbf{q}} \quad (4.4)$$

Furthermore, if we define both sides of the equation by dt we obtain that

$$\dot{\mathbf{r}}_i = \mathbf{v}_i = \mathbf{J}_i \dot{\mathbf{q}} \quad (4.5)$$

Note: In simple terms, we can say that the (k, j) element of the Jacobian quantifies how much effect a variation in the generalized coordinate k will have on the displacement of a joint i along the j_{th} cartesian coordinate. In fact, for a robotic arm with m links, where we defined the set of generalized coordinates $\{q_1, \dots, q_m\}$, we can write

$$d\mathbf{r}_i = \begin{bmatrix} dr_{i_x} \\ dr_{i_y} \\ dr_{i_z} \end{bmatrix} = \mathbf{J}_i d\mathbf{q} = \begin{bmatrix} \frac{\partial r_{i_x}}{\partial q_1} & \dots & \frac{\partial r_{i_x}}{\partial q_m} \\ \frac{\partial r_{i_y}}{\partial q_1} & \dots & \frac{\partial r_{i_y}}{\partial q_m} \\ \frac{\partial r_{i_z}}{\partial q_1} & \dots & \frac{\partial r_{i_z}}{\partial q_m} \end{bmatrix} \begin{bmatrix} dq_1 \\ \dots \\ dq_m \end{bmatrix} = \begin{bmatrix} \frac{\partial r_{i_x}}{\partial q_1} dq_1 + \dots + \frac{\partial r_{i_x}}{\partial q_m} dq_m \\ \frac{\partial r_{i_y}}{\partial q_1} dq_1 + \dots + \frac{\partial r_{i_y}}{\partial q_m} dq_m \\ \frac{\partial r_{i_z}}{\partial q_1} dq_1 + \dots + \frac{\partial r_{i_z}}{\partial q_m} dq_m \end{bmatrix} \quad (4.6)$$

4.1.2 Lagrange derivation of the Robot manipulator dynamic equations

Let's consider a robot manipulator system, and let's call T and U the kinetic and potential energy of the system, respectively. We further define a set of **generalized coordinates** $\{q_1, \dots, q_m\}$, corresponding to the joint angles, and a set of **generalized forces** $\{Q_1, \dots, Q_m\}$. For such a dynamic system, Lagrangian dynamic equations are defined as:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = Q_j \quad \text{for each } j = 1, \dots, m. \quad (4.7)$$

where $L = T - U$. If we rewrite 4.7 as a function of K and U we obtain:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{d}{dt} \frac{\partial U}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} + \frac{\partial U}{\partial q_j} = Q_j \quad \text{for each } j = 1, \dots, m. \quad (4.8)$$

Let's now try to derive each of these quantities (T , U and their partial derivatives).

T : The Kinetic Energy of the system

For the robot manipulator system described above, we can define the kinetic energy for each link i as:

$$T_i = \frac{1}{2} m_i \mathbf{v}_{ci}^T \mathbf{v}_{ci} + \frac{1}{2} \boldsymbol{\omega}_i^T \mathbf{I}_i \boldsymbol{\omega}_i \quad (4.9)$$

where m_i is the mass of the i_{th} link, \mathbf{v}_{ci} the linear speed of the i_{th} link centroid, $\boldsymbol{\omega}_i$ the angular velocity of the i_{th} link centroid and \mathbf{I} the Inertia matrix of the i_{th} link. The total kinetic energy is defined by the sum of the kinetic energies of each link:

$$T = \sum_{i=1}^m T_i = \sum_{i=1}^m \left(\frac{1}{2} m_i \mathbf{v}_{ci}^T \mathbf{v}_{ci} + \frac{1}{2} \boldsymbol{\omega}_i^T \mathbf{I}_i \boldsymbol{\omega}_i \right) \quad (4.10)$$

Note: The reader might have noticed that the two terms of the above formulation for a solid body (the link) resembles the definition of a point mass translational and rotational kinetic energy, according to Newton's mechanics:

$$\begin{aligned} T_i^{lin} &= \frac{1}{2} m_i \mathbf{v}_i^T \mathbf{v}_i \\ T_i^{rot} &= \frac{1}{2} \boldsymbol{\omega}_i^T \mathbf{I}_i \boldsymbol{\omega}_i \end{aligned} \quad (4.11)$$

where m_i is the mass of the particle, \mathbf{v}_i its linear velocity, $\boldsymbol{\omega}_j$ its angular velocity and \mathbf{I} its moment of inertia. In fact, a rigid body can be seen as an ensemble of points, for each of which we can define kinetic energy according to 4.11. Equation 4.9 derives from the integration of the kinetic energy (defined according to 4.11) over all of the points constituting a solid body (as, for example, the robot i_{th} link). In simple terms, we can consider the formulation in 4.9 as a generalized formula for the kinetic energy for a generic rigid body, where the linear and angular velocities \mathbf{v}_i and $\boldsymbol{\omega}_j$ are referred to the body centroid, and \mathbf{I} is an inertial matrix which depends on the geometry of the body.

As we want to work with the set of generalized coordinates $\{q_1, \dots, q_m\}$, we now have to refer to these quantities with respect to such coordinates:

$$\begin{aligned} \mathbf{v}_{ci} &= \mathbf{J}_j^L \dot{\mathbf{q}} \\ \boldsymbol{\omega}_i &= \mathbf{J}_j^A \dot{\mathbf{q}} \end{aligned} \quad (4.12)$$

where \mathbf{J}_i^L is the Jacobian matrix relating the linear velocity of the i_{th} link centroid and the joint angles $\{q_1, \dots, q_m\}$, and \mathbf{J}_i^A is the Jacobian matrix relating the linear velocity

of the i_{th} link centroid and the joint angles $\{q_1, \dots, q_m\}$. Therefore, we can rewrite 4.13 as

$$T = \frac{1}{2} \sum_{i=1}^m m_i (\mathbf{J}_i^L \dot{\mathbf{q}})^T (\mathbf{J}_i^L \dot{\mathbf{q}}) + (\mathbf{J}_i^A \dot{\mathbf{q}})^T \mathbf{I}_j (\mathbf{J}_i^A \dot{\mathbf{q}}) = \frac{1}{2} \dot{\mathbf{q}}^T \mathbf{H} \dot{\mathbf{q}} \quad (4.13)$$

where \mathbf{H} is called **Multi-Body Inertial Matrix**. Now, we have to compute the derivatives of the kinetic energy, to substitute in Equation 4.8:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} - \frac{d}{dt} \frac{\partial U}{\partial \dot{q}_j} + \frac{\partial U}{\partial q_j} = Q_j \quad \text{for each } j = 1, \dots, m. \quad (4.14)$$

Which yields:

$$\begin{aligned} \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} &= \frac{d}{dt} \frac{\partial}{\partial \dot{q}_j} \left(\frac{1}{2} \dot{\mathbf{q}}^T \mathbf{H} \dot{\mathbf{q}} \right) - \frac{\partial}{\partial q_j} \left(\frac{1}{2} \dot{\mathbf{q}}^T \mathbf{H} \dot{\mathbf{q}} \right) \\ &= \sum_{i=1}^m H_{ik} \ddot{q}_i + \sum_{i=1}^m \left(\frac{d}{dt} H_{ik} \right) \dot{q}_i - \frac{\partial}{\partial q_j} \left(\frac{1}{2} \sum_{i=1}^m \sum_{k=1}^m H_{ik} \dot{q}_i \dot{q}_k \right) \end{aligned} \quad (4.15)$$

In matrix form, we can rewrite equation 4.15 as

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{\mathbf{q}}} - \frac{\partial T}{\partial \mathbf{q}} = \mathbf{H} \ddot{\mathbf{q}} + \dot{\mathbf{q}}^T \mathbf{C} \dot{\mathbf{q}} \quad (4.16)$$

(\mathbf{C}) is a matrix that depends both on the joint position and on their velocity. \mathbf{C} is associated with centrifugal and Coriolis forces applied to the robot.

U : The Potential Energy of the System

If gravity is the only conservative force, then the potential energy of the system is given by:

$$U = - \sum_{i=1}^m m_i \mathbf{g} \mathbf{r}_{ci} \quad (4.17)$$

As for the kinetic energy, we have to compute the derivatives of the kinetic energy, to substitute in Equation 4.8:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} - \frac{d}{dt} \frac{\partial U}{\partial \dot{q}_j} + \frac{\partial U}{\partial q_j} = Q_j \quad \text{for each } j = 1, \dots, m. \quad (4.18)$$

We notice that the U does not depends on \dot{q}_j , therefore:

$$\frac{d}{dt} \frac{\partial U}{\partial \dot{q}_j} + \frac{\partial U}{\partial q_j} = - \sum_{i=1}^m m_i \mathbf{g} \frac{\partial \mathbf{r}_{ci}}{\partial q_j} \quad \text{for each } j = 1, \dots, m. \quad (4.19)$$

As $\frac{\partial r_{ci}}{\partial q_j}$ only depends on q , we can rewrite equation 4.19 as:

$$\boxed{\frac{d}{dt} \frac{\partial U}{\partial \dot{q}} + \frac{\partial U}{\partial q}} = \tau_g \quad (4.20)$$

where the j – th component of τ_g is given in equation 4.19

Q: The Generalized forces

When analyzing generalized forces applied to a robotic arm, we consider both the torques applied to the robot's joints and the external forces applied to the robot (usually to its end-effector). Therefore, the external generalized forces can be written as:

$$Q_j = \tau_j + J_{j,..}^T \mathbf{F} \quad \text{for each } j = 1, \dots, m \quad (4.21)$$

where τ_j is the torque acting on the j -th joint and $J_{j,..}^T$ the j th line of the Jacobian matrix, \mathbf{F} the external force acting on the robot, $J_{j,..}^T \mathbf{F}$ the generalized force component due to the external force on the joint j .

Combining all the terms together, and considering them for all the m links in the robot, we can derive the following equation

$$\mathbf{H}\ddot{q} + \dot{q}^T \mathbf{C}\dot{q} + \tau_g(q) = \tau + J^T F \quad (4.22)$$

which we can write in its final, known, form as:

$$\boxed{\mathbf{H}\ddot{q} + \mathbf{S}\dot{q} + \tau_g = \tau + J^T F} \quad (4.23)$$

We can express the same equation in cartesian coordinates as:

$$\boxed{\mathbf{H}_x \ddot{x} + \mathbf{S}_x \dot{x} + g_x = \mathbf{J}^{-T} \tau + F} \quad (4.24)$$

where \mathbf{H}_x , \mathbf{S}_x and g_x can be computed by substituting $\mathbf{J}^T x$ in the previous equations.

It is important to keep in mind that τ constitutes our control variable. That is, regardless of the behavior we want to enforce on the robot, the only thing we can act on are the torques applied to the joints. We can therefore rewrite equation 4.24 as:

$$\boxed{\tau = \mathbf{J}^T (\mathbf{H}_x \ddot{x} + \mathbf{S}_x \dot{x} + g_x - F)} \quad (4.25)$$

4.2 Control Strategies for Robot manipulators

Now that we have a clear understanding of how to describe the dynamic behavior of a robot manipulator, let's consider the various strategies that can be used to control its motion in space and with respect to the surrounding environment.

One approach to controlling the motion of a robotic arm is to specify a target trajectory $q(t)$ and compute the desired joint velocity \dot{q} and accelerations \ddot{q} by taking the first and second derivative of the trajectory. If we can estimate the Coriolis and centrifugal forces and the inertial matrix of the robot, we can compute the necessary torques to follow the desired trajectory. However, if the robot is interacting with objects in the environment, such as a wall that impedes its motion in a certain direction, we must also consider the reaction forces generated between such objects and the robot. More precisely, we need to understand how we want the robot to interact with the external environment. Some common control schemes include position control, stiffness control, impedance control, force control, and mixed force-impedance control. Visual servoing and other sensing modalities can be further included in the control scheme.

4.2.1 Position Control

It is important to note that pure position control is not typically achievable in real-world situations. However, for the sake of understanding, let's consider how pure position control might work in the scenario described earlier, where the robot is interacting with objects in the environment. In a pure position control scenario, the robot would attempt to follow the target trajectories at all costs, generating as much force as necessary to counteract the reaction forces applied by the environment. In other words, it would "break through" the wall if necessary to achieve the desired position.

4.2.2 Stiffness Control

In the case of stiffness control, we require the robot to behave like a pure spring (Fig. 4.1). In mathematical terms, this is equivalent to say that we enforce that the robot effector behaves according to:

$$-K_m(\mathbf{x} - \mathbf{x}_d) = \mathbf{F} \quad (4.26)$$

Where \mathbf{F} is the external force, \mathbf{x} the measured position of the robot end-effector, \mathbf{x}_d the desired position of the robot end-effector, K_m the spring stiffness, which is set

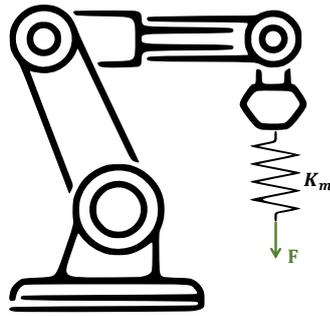


Fig. 4.1: Behavior of a robot under stiffness control, where the robot behaves as if a spring with stiffness K_m were attached to its end-effector

depending on the desired behavior. This means that in the presence of external forces F that deviate the robot from its original trajectory, the robot will apply torques to its joints in a way that "pushes against" the direction of the force, with an intensity that linearly increases with the magnitude of the external force. It is worth noting that pure position control is equivalent to a stiffness control with infinite stiffness.

While pure stiffness control offers a certain degree of freedom, its application in the medical domain can still lead to undesirable behaviors, such as the application of undesired forces on the patient body.

4.2.3 Impedance Control

To address the limitations of pure stiffness control, we can modify the robot model by adding a damper in parallel with the spring, effectively treating the robot as a mass-spring-damper system (see Figure 4.2).

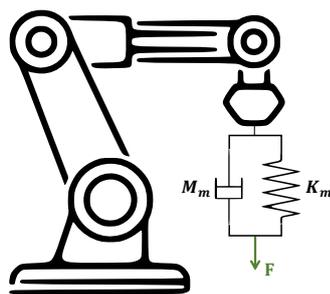


Fig. 4.2: Behavior of a robot under impedance control, where the robot behaves as if a spring with stiffness K_m and a damper with damping coefficient $Dm-m$ were attached to its end-effector

The purpose of the damper is to limit the speed of the robot to the desired speed. In the presence of sudden external forces, the robot will apply a force proportional to the

displacement from its desired position in an attempt to return to that position, but it will do so without sudden changes in speed, avoiding abrupt and unwanted motion.

$$M_m(\ddot{\mathbf{x}} - \ddot{\mathbf{x}}_d) + D_m(\dot{\mathbf{x}} - \dot{\mathbf{x}}_d) + K_m(\mathbf{x} - \mathbf{x}_d) = \mathbf{F} \quad (4.27)$$

Using the mass-spring-damper model described above, we can substitute the corresponding equations into the general equation for the robot's torques (equation 4.28), resulting in:

$$\begin{aligned} \boldsymbol{\tau} = & \mathbf{H}\mathbf{J}^{-1}\{\ddot{\mathbf{x}} + \dot{\mathbf{J}}\dot{\mathbf{q}} + \mathbf{M}_m^{-1}[D_m(\dot{\mathbf{x}} - \dot{\mathbf{x}}_d) + K_m(\mathbf{x} - \mathbf{x}_d)]\} \\ & + \mathbf{S}\dot{\mathbf{q}} + \boldsymbol{\tau}_g + \mathbf{J}^T[\mathbf{H}_x\mathbf{M}_m - \mathbf{I}]\mathbf{F} \end{aligned} \quad (4.28)$$

4.2.4 Force Control

In some cases, it is important to be able to directly control the interaction forces between the robot and the environment. In force control, the external forces are set to a desired value:

$$\boldsymbol{\tau} = \mathbf{J}^T(\mathbf{H}_x\ddot{\mathbf{x}} + \mathbf{S}_x\dot{\mathbf{x}} + \mathbf{g}_x - F_{desired}) \quad (4.29)$$

Force control is often used to control the robot in a direction where the motion is almost zero, such as in the case of robotic ultrasound, where it is applied only in the z direction with the assumption of negligible velocity and acceleration. Thus, equation 4.30 can be simplified to:

$$\boldsymbol{\tau} = \mathbf{J}^T(\mathbf{g}_x - F_{desired}) \quad (4.30)$$

This provides a reasonable force control law when the robot is stationary. However, it is often beneficial to incorporate feedback from external force sensors to continuously monitor and correct any deviations from the desired force due to simplifications in the model. With the inclusion of force sensors as feedback, equation 4.30 can be rewritten as:

$$\boldsymbol{\tau} = \boldsymbol{\tau}_g - \mathbf{J}^T(F_{desired} + K_{fp}F_e + K_{fi}\int_t F_e(t)dt) \quad (4.31)$$

where F_e is the difference between the desired force and the measured force, K_{fp} and K_{fi} are constant gains.

4.2.5 Hybrid Control

As mentioned previously, pure force control is rarely used in practice. Instead, it is more common to use a combination of force and motion (e.g. impedance) control, as shown in Figure 4.3. In this approach, force control is used along the axes where full control over the applied forces is desired, while impedance control is used along the remaining axes to control the trajectory. This hybrid control approach allows for the desired level of control to be applied in each specific direction.

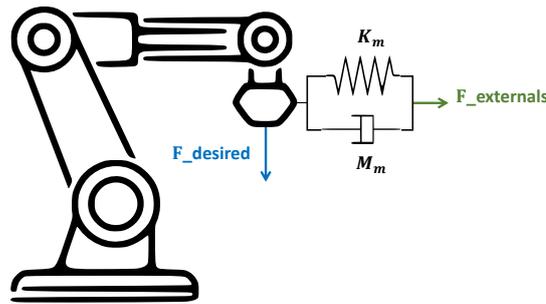


Fig. 4.3: Behavior of a robot under hybrid impedance -force control, where impedance control is used along the horizontal axis while force control is used along the vertical axis.

4.3 Visual Servoing in Robotics

In the previous section, we discussed how to control the joint torques of a robot to achieve a desired behavior, including the use of hybrid force/impedance control to maintain a constant applied force in certain directions while allowing the robot to move along a predefined trajectory in other degrees of freedom. However, in some situations, such as in the case of ultrasound acquisitions, defining an offline trajectory beforehand may not be optimal. This is because differences in the anatomy between the time when pre-operative data was taken and the time of the acquisition, as well as the potential for limited visibility of the anatomy due to artifacts such as shadowing, can lead to errors in the offline trajectory. To address these issues, it is often necessary to incorporate visual feedback into the robot control loop. This can be achieved by processing ultrasound images to extract relevant features, using these features to compute the desired location of the robot relative to the current one, and using the computed displacement to update the desired position input to the controller. The specific strategies for implementing visual servoing will vary depending on the specific application and the organ being scanned (Fig. 4.4).

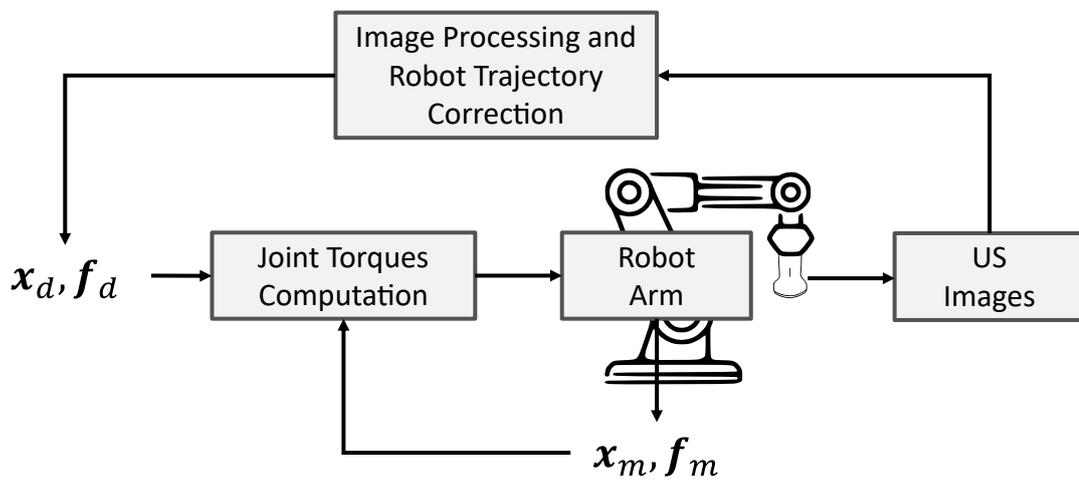


Fig. 4.4: Schematic representation of the incorporation of visual servoing into the robot control scheme.

Ultrasound Imaging

In this chapter, we will examine the fundamental concepts of ultrasound imaging, including the definition and characteristics of ultrasound waves, their generation, and the physical principles underlying the formation of ultrasound images. We will also discuss the impact of noise and artifacts on ultrasound images and the challenges of using deep learning techniques for ultrasound image processing, with a particular focus on the issue of data augmentation.

5.0.1 Ultrasound Imaging - Image Formation Principles

Ultrasound refers to acoustic waves that have frequencies higher than 20 KHz. Acoustic waves are defined as the propagation of variations in the density of a medium through space and time. For example, when a violinist plays a chord on a violin, the chord vibrates periodically, causing the surrounding air to compress and rarefy periodically. These areas of compression and rarefaction propagate through the air, constituting an acoustic wave (Fig. 5.1).

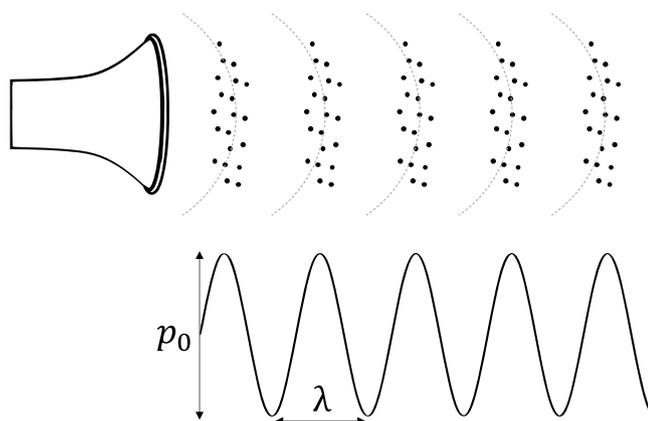


Fig. 5.1: Top: The areas of air compression and rarefaction produced by a sound wave. Bottom: The sinusoidal representation of the acoustic wave, where p_0 is the wave amplitude and λ is the wavelength

From a mathematical viewpoint, we can describe the behavior of such perturbations in space and time as:

$$p(t, \mathbf{x}) = p_0 \sin(\omega t + kx) \quad (5.1)$$

Where $w = 2\pi/T$ is the angular frequency, $k = 2\pi/\lambda$ is the wave number, T is the period of the wave and λ is the wavelength. We can derive the velocity of propagation of the wave as

$$v = \frac{\Delta space}{\Delta time} = \frac{\lambda}{T} \quad (5.2)$$

As for the acoustic waves caused by the vibrations of the violin chord, ultrasound waves are produced by mechanical vibrations that propagate through a medium as areas of compression and rarefaction. However, unlike the acoustic waves produced by the vibrations of a violin chord, the vibrations that generate ultrasound occur at a much higher frequency and are induced in a class of materials called piezoelectric materials.

Piezoelectricity

In the second half of the 19th century, the Curie brothers discovered that certain materials could produce an electrical field when subjected to mechanical strain or pressure: this phenomenon is called piezoelectricity. In this chapter, we will explore the role of piezoelectricity in the creation of ultrasound images. First, we will examine the molecular basis of piezoelectricity and how materials respond to electric fields. We will then delve into how an electric field can be generated by and produce strain, and how this property is utilized in the construction of ultrasound transducers. The interested reader can refer to [16] for further information on piezoelectricity.

Behavior of Materials in an Electric Field

What happens to a solid when we place it in an electric field, that is, when it is subject to a potential difference (Fig. 5.2)?

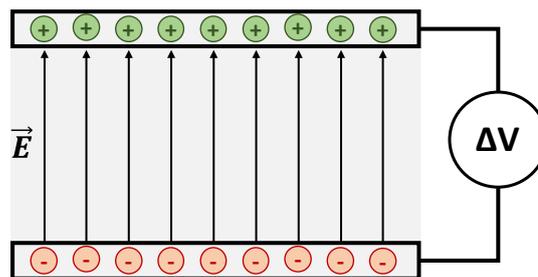


Fig. 5.2: The electrical field generated by the presence of a potential difference ΔV . Negative charges (red circles) cannot travel towards positive charges (green circles) due to the presence of insulating material in between them.

When a solid is placed in an electric field, a potential difference is established, which can result in the generation of a current in metallic materials. This current can lead to the uneven distribution of charge within the material, causing a polarization of the material. However, for insulating materials, such as dielectrics, charges are not able

to freely move within their structures and therefore a current is not generated when exposed to an electric field. However, when subject to an electrical field, they are still able to **polarize**. That is, due to the generation and/or alignment of their electrical dipoles to the electric field, they undergo a separation of their positive and negative charges. To better understand these properties, it is worth it to spend a few words on the concept of the electrical dipole of a molecule.

An electric dipole is a system (in our case the molecule constituting the material) where the centers of positive and negative charges are separated. In nature, there are molecules that possess a moment of dipole *even in absence of an electric field* (polar molecules) and molecules that do not (non-polar molecules) (Fig. 5.3).

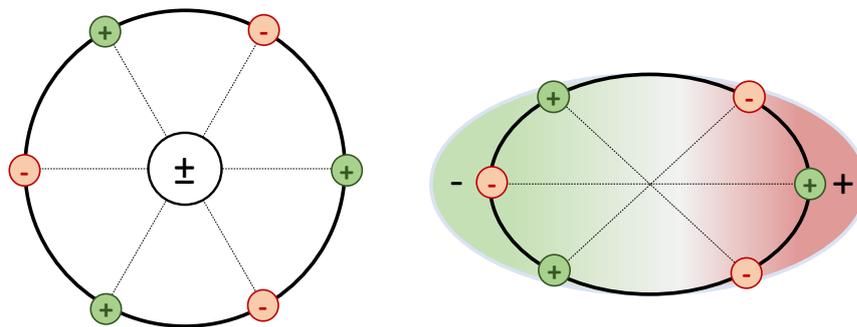


Fig. 5.3: (left) a non-polar molecule (right) a polar molecule. It can be appreciated how in the non-polar molecule the negative and positive charge center coincides, due the molecule geometry. The elongated geometry in the non-polar molecule causes the centers of positive and negative charges to differ, causing a non-null dipole moment.

Similarly, dielectrics can be categorized into polar and non-polar dielectrics depending on the nature of their constituting molecules.

Polarization in non-polar dielectrics is caused by a modification of the of the arrangement of the molecule, due to the presence of an electric field, with the consequent generation of an electric dipole which aligns with the electric field (Fig. 5.4). In polar dielectrics, the polarization arises from the re-alignment of the existing electric dipoles (which, in absence of an electric field, are randomly distributed) with the electric field (Fig. 5.5).

Piezoelectric Materials

Piezoelectric materials are a category of materials that is able to polarize in presence of a mechanical strain. Similarly to what is described for dielectrics, the polarization following a mechanical strain is induced by the generation of a molecular dipole due to molecules deformation (for non-polar piezoelectrics) (Fig. 5.6) or by the re-alignment

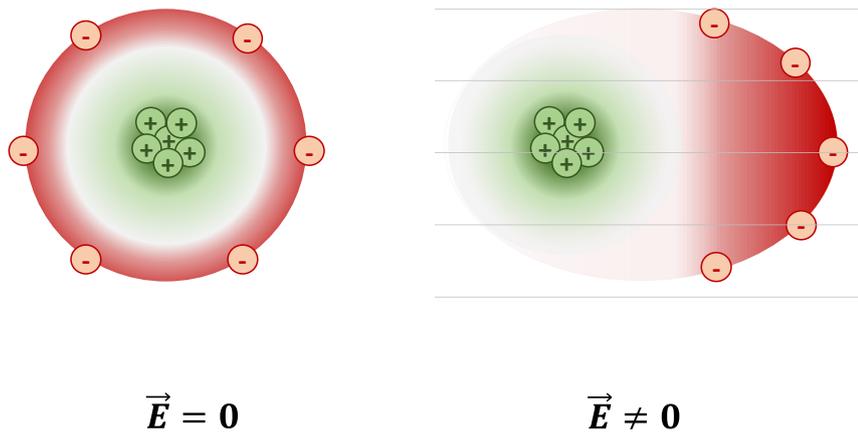


Fig. 5.4: A non-polar dielectric in absence of an electric field. (right) A non-polar dipole undergoing polarization due to the presence of an electric field. It can be noticed how the presence of an electric field causes a rearrangement of the charges within the molecule, thus the split of positive and negative charge centers and consequently the generation of a non-null dipole moment.

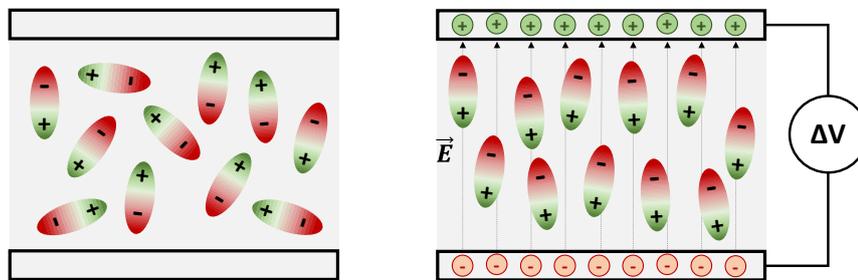


Fig. 5.5: Molecules in a dielectric cannot freely move in the solid. However, their structure allows for minor motions as changes in their orientation align to an applied electrical field. This figure shows the example of a polar dielectric, which undergoes polarization due to the re-orientation of its electrical dipoles.

of electrical dipoles of the material induced by the mechanical deformation (Fig. 5.7).

Notably, some piezoelectric materials also show the reverse effect. That is, mechanical deformation is induced when the solid is subject to an external electrical field. This is called the inverse piezoelectric effect.

Ultrasound Transducers

So far, we have seen that thanks to the discovery of the Curie brothers in 1880, there exists materials that are able to:

- a. Deform when subject to an external electrical field, i.e. to an external voltage (direct piezoelectric effect)
- b. Generate an electrical field, i.e. a voltage, when subject to mechanical deformation (inverse piezoelectric effect)

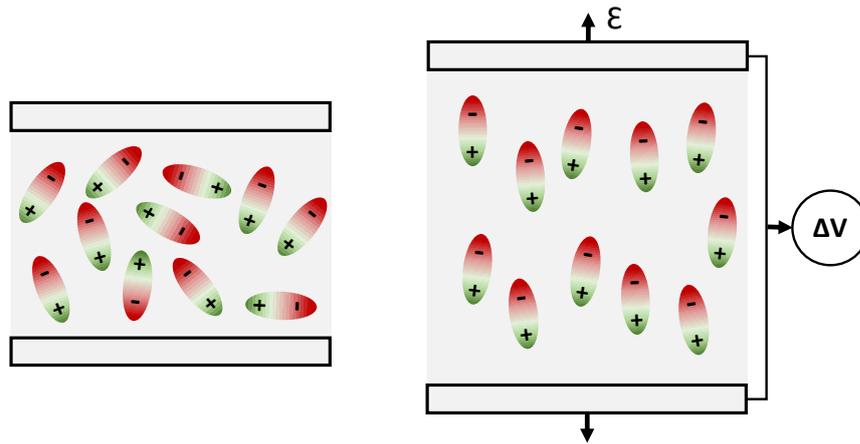


Fig. 5.6: Polarization occurring in a polar piezoelectric material due to the application of a mechanical strain ϵ . The polarization leads to the generation of a net electric field within the dipole, thus of a voltage difference ΔV which can be measured at its side

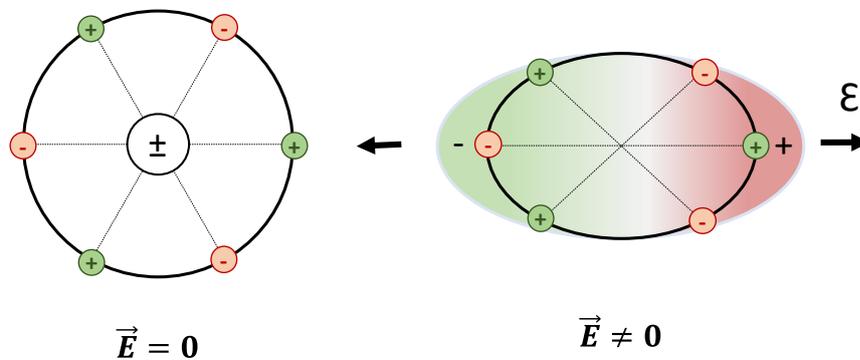


Fig. 5.7: Polarization occurring in a polar piezoelectric material due to the application of a mechanical strain ϵ . The polarization leads to the generation of a net electric field within the dipole, thus of a voltage difference ΔV which can be measured at its side

As we already know the relation between a vibrating object and the generation of an acoustic wave, we can see how:

- a. The application of a certain voltage at a frequency TX on a piezoelectric element will generate a sound wave at a frequency TX.
- b. An incoming ultrasound wave with a frequency RX reaching the piezoelectric element will generate a vibration (thus a mechanical deformation) of the piezoelectric element at a frequency RX, leading to the generation of an electrical voltage at the extremity of the piezoelectric element.

Furthermore, we know that when an acoustic wave reaches a reflective surface, it is reflected back. Let's assume to have one piezoelectric transducer, placed at a distance d from a point P_i . Both transducer and point are immersed in a medium where sound propagates with a constant velocity v . When the transducer emits an ultrasound wave, the propagating wave will travel in the medium until it encounters the reflective point. Here part of the wave will be absorbed, the part will pass through

the surface and part will be reflected. The reflected wave, which we call *echo* travels back towards the piezoelectric element. When it reaches the piezoelectric element, a mechanical deformation is induced in the element, and a voltage difference develops at its extremities (Fig. 5.8).

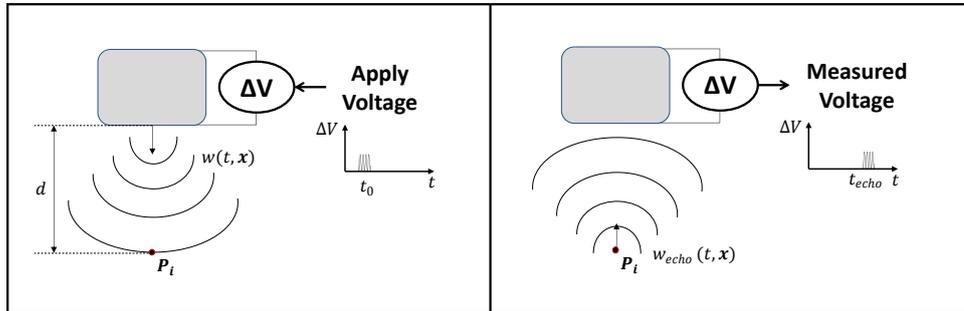


Fig. 5.8: The generation of an ultrasound wave and the generation of an echo from a point P reached by the ultrasound wave

If we record the time at which we measure the voltage caused by the echo, we can compute the distance between the transducer and point P as:

$$d = \frac{t_{echo}}{2v} \quad (5.3)$$

Similarly, if we look at the voltage recording at time t_{echo} , we can gather information on the reflectiveness of the point P_i located at a distance d from the piezoelectric element:

$$r_i(d) = Voltage(t = 2dv) \quad (5.4)$$

where $r_i(d)$ is a measure of the reflectivity of the point i located at a distance d from the piezoelectric element.

Although this simple experiment shows the potential of using a piezoelectric transducer to obtain spatial information, it is evident that it presents many shortcomings as we introduce more complex scenarios. As an example, let's imagine introducing multiple reflective points in the scene instead of one (Fig. 5.9)

We can see that points located at the same distance from the transducer will generate echos at the same time instant, making it impossible for the system to disentangle echo signals coming from the different spatial locations (Fig. 5.9, left). Moreover, if the multiple points are located at different distances from the transducer, their contribution might cancel due to the destructive interference of the two waves. Given these considerations, how do we make sure that what we are recording is uniquely

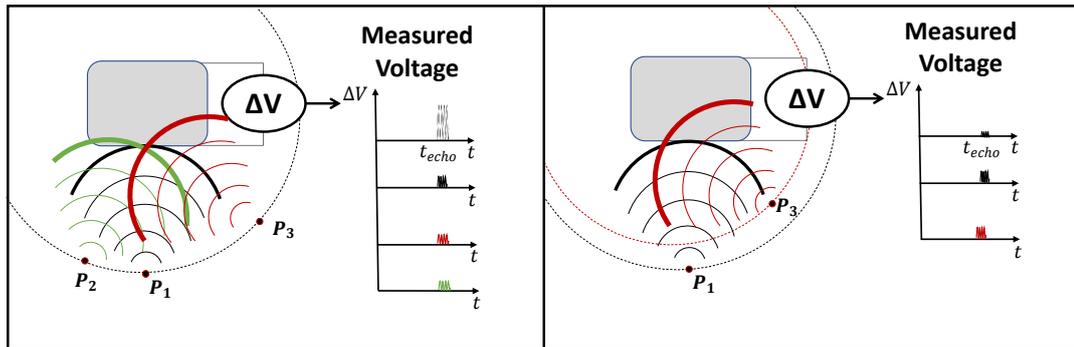


Fig. 5.9: Left: generation and recording of simultaneous echoes generated from point masses located at equal distances from the ultrasound source. Right: generation and recording of echoes undergoing destructive interference

associated with a given point P ? Firstly, we have to focus the wave propagation along a given direction, in order to "isolate the source of the echo" and stimulate one point at a time. There are different techniques to focus the ultrasound beam. The most commonly used one is the **electronic focusing**. In electronic focusing, multiple piezoelectric elements are activated with a delay pattern, in a way that the acoustic waves generated by the different elements sum up in a constructive manner only in a limited area, thus creating a focused beam. The line along which such beams travel is called scanline (Fig. 5.10)

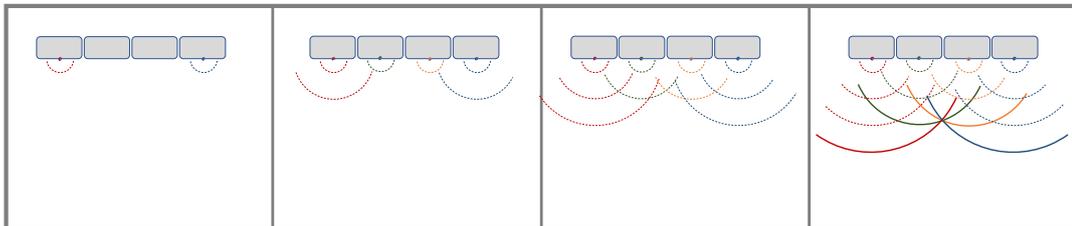


Fig. 5.10: The transmit delay pattern to achieve focusing (i.e. constructive interference) at a focus point

Notice that in all the other positions outside of the beam, the waves have a "random" phase shift with respect to each other. If we introduce a sufficient number of transducer elements, we can reasonably assume that the contributions of the waves outside of the scan line cancel each other due to phase misalignment.

A given point P lying on the scanline, will generate an echo that reaches the transducer i at the time:

$$t_{echo}^i(P) = 2vd_i \quad (5.5)$$

where $t_{echo}^i(P)$ is the echo generated by the point P on the transducer i and d_i is the distance between P and the transducer element i . Therefore, we can apply delays

in the receiving signals, so to isolate, for each transducer, the echo generated by the same points in space. After we have isolated the contribution of a given point along the scanline on each element of the transducer, we can sum up the (delayed) signals to improve the overall signal-to-noise ratio (Fig. 5.11). This procedure of emitting a set of pulses with a transmit delay pattern, receiving the echo with a receive delay pattern, and summing up the recorded data is called Delay-and-Sum (DAS).

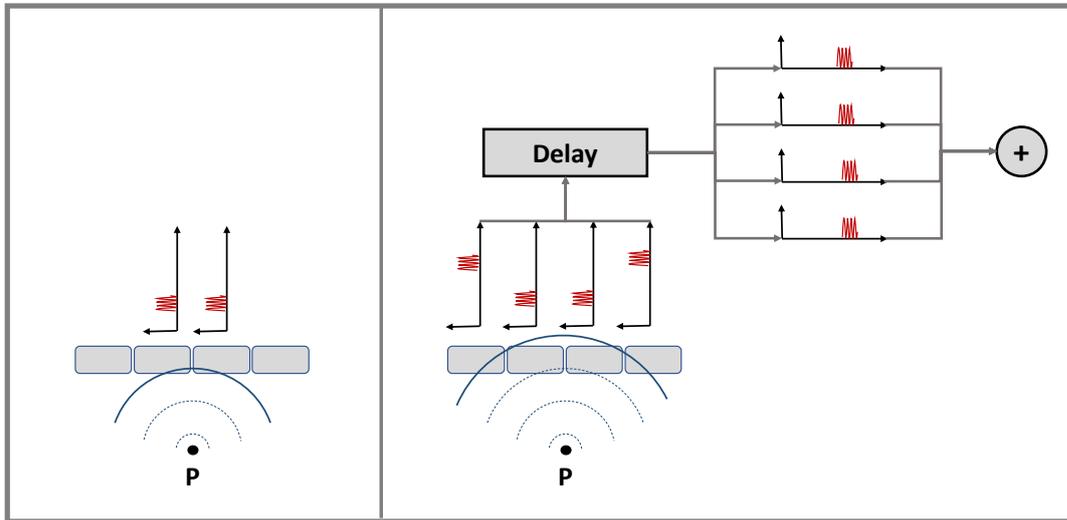


Fig. 5.11: The delayed pattern applied to the transducer elements to synchronize them to the reception of the echoes coming from point P

To account for tissue absorption, a Time-Gain-Compensation is typically applied to the signal upon recording. That is, signals are weighted by an exponential factor in space, to account for the fact that as the ultrasound wave travels in the tissue it also gets absorbed according to the:

$$I = I_0 e^{-\beta x} \quad (5.6)$$

where β is called the absorption coefficient of the medium.

Until now, we have been looking at ultrasound waves and ultrasound image formation with a "magnifying lens", through which we could analyze interactions between waves and points at a microscopic scale, that is, at the scale of the ultrasound wavelengths. We have seen how these wavefronts interact with each other to create constructive and destructive interference, and how point sources behave when reflecting incoming ultrasound waves. To be able to analyze other characteristics of the ultrasound image and image formation, we need now to move to a larger scale, and understand how the ultrasound beams created via electronic focusing interact with other materials/surfaces.

When an ultrasound beam traveling along a certain direction encounters a surface, its interaction with the surface can be mainly described by two phenomena:

1. Reflection
2. Refraction

Reflection causes an echo signal to be generated and travels back to the source along a given propagating direction which depends on the incident angle between the incoming beam and the surface. Refraction is related to the propagation of the beam into the tissue, with an angle that depends on the properties of the media the beam travels through (Fig. 5.12).

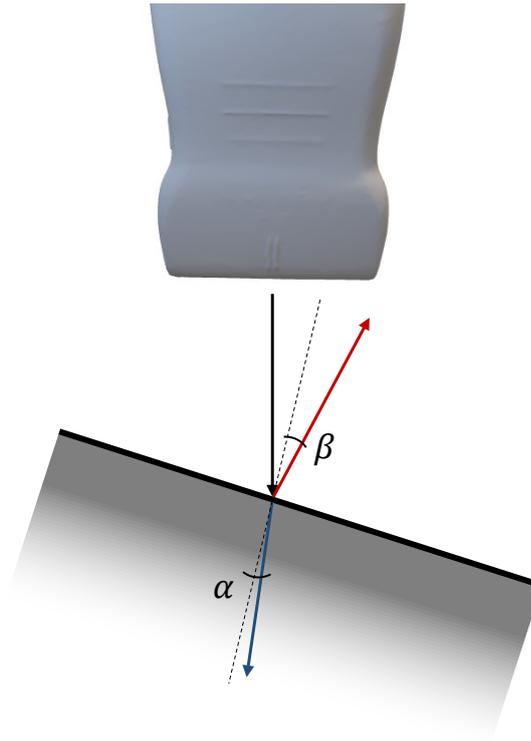


Fig. 5.12: The interaction between an ultrasound beam with a surface. Part of the beam (blue arrow) propagates in the surface with refraction angles α , and part (red arrow) is reflected back with a reflection angle β equal to the angle between the incoming beam (black arrow) and the normal to the surface (dashed line)

Note that all that we have described at a microscopic scale still holds true. They are the (microscopic) cause underneath these phenomena that we observe at a macroscopic level. That is, reflection and refraction are the macroscopic results of the contribution of interaction between each point in the surface and the ultrasound waves.

5.0.2 Ultrasound Imaging - Noise and Artifacts

In the previous subsection, we have seen how ultrasound waves are utilized to generate focused beams, and how echoes from reflective surfaces are exploited to

gather a spatial reconstruction of the environment. In this process, we indeed had to make some assumptions. As an example, we had to assume that, when electronically generating a focused beam, waves fronts outside of the scanline would undergo destructive interference, thus not contributing to the recorded signals. Moreover, we assumed that the propagation speed is constant and that the beam never deviates from the scanline. Although reasonable, all these assumptions do not always fully hold true. Excitations of regions outside of the focused beam might occur. Moreover, the speed of sound is not constant and changes in different tissues. Different tissues in turn can have different refraction indexes, thus causing beams to be refracted from their desired path. These and other phenomena commonly occur during ultrasound data recording and reconstruction, affect the overall signal-to-noise ratio, and in general lead to the generation of different noises and artifacts in the reconstructed images. In the next paragraphs, we will give an introduction to the most common noise and artifacts on ultrasound images, together with their physical explanation in relation to what was discussed in the previous section.

1. Signal-to-Noise ratio and contrast

In ultrasound b-mode compounding, it is generally assumed that ultrasound waves will interfere constructively along the scanlines and cancel each other out elsewhere due to destructive interference. However, for a finite number of piezoelectric elements, this is not always the case. Even outside the scanline, the constructive and destructive interference between waves can still generate specific patterns and produce spurious signals that can reach the transducer elements during recording, reducing the overall signal-to-noise ratio. Additionally, tissue absorption can significantly reduce the intensity of the ultrasound beam in regions far from the transducer, causing decreased contrast and visibility of different anatomical structures.

1. Speckle Noise

Speckle noise is a common source of noise in ultrasound images. It is caused by the scattering of ultrasound waves by molecules in tissues, which reflects the waves in random directions. The scattered waves from different particles combine to create an interference pattern that is recorded by the ultrasound transducer, resulting in a characteristic scattered texture in the image. While the presence of speckles can sometimes be useful for characterizing tissue in ultrasound images, they can also reduce the visibility of certain anatomical structures and make data interpretation more difficult.

2. Mirroring

The mirroring effect occurs when the ultrasound ray is reflected off a surface and then bounces off a second structure before returning to the transducer (Fig. 5.13). This second reflective structure causes the ultrasound ray to be redirected toward the transducer, resulting in the creation of a mirrored image.

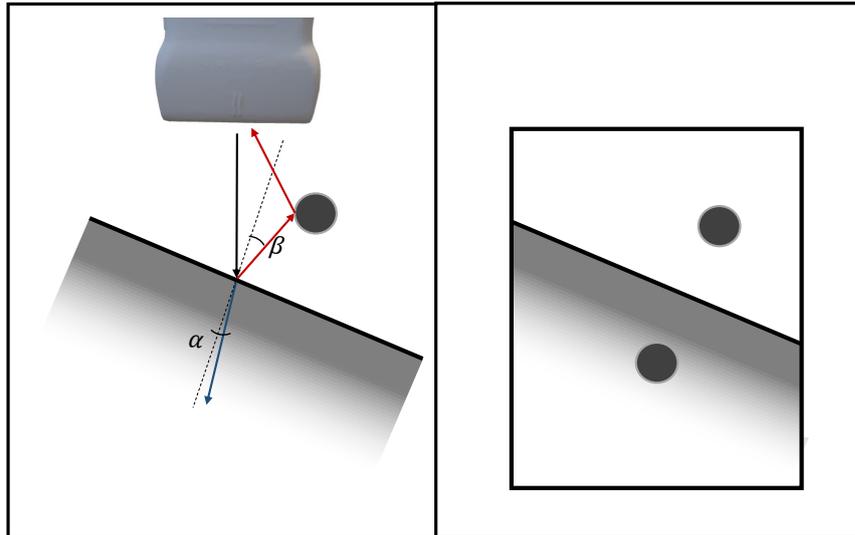


Fig. 5.13: Illustration of the mirroring effect in ultrasound imaging, where the original ultrasound ray is reflected off a surface and then bounces off a second structure before returning to the transducer, resulting in a mirrored image of the second reflective object.

Multiple reflection and Reverberation

Previously, we discussed how an echo wave can be generated by the interaction between an ultrasound pulse emitted by a transducer and a reflective surface. But what happens when the echo wave reaches the transducer surface? Part of the wave energy is first converted into mechanical energy, causing the vibration of the piezoelectric element, and then into electrical energy due to the piezoelectric effect. However, some of the waves may also be reflected back towards the tissues. The reflected echo will then travel back towards the reflective structure and be reflected again, resulting in an echo at $t_{echo} = 4d/v$, where d is the distance between the transducer and the reflective structure and v is the sound propagation velocity in the medium. This process can create the appearance of multiple reflections of the same structure at multiples of the actual distance between the object and the transducer (Fig. 5.14).

Reverberation can also be caused by multiple reflections between reflective surfaces. In this case, reverberation is generated by multiple reflections occurring between parallel surfaces located close to each other.

Acoustic Shadowing

As previously discussed, ultrasound imaging relies on the reflection of ultrasound waves at tissue interfaces. The amount of energy reflected depends on the tissue properties, specifically the reflectivity. Tissues with high reflectivity will reflect a

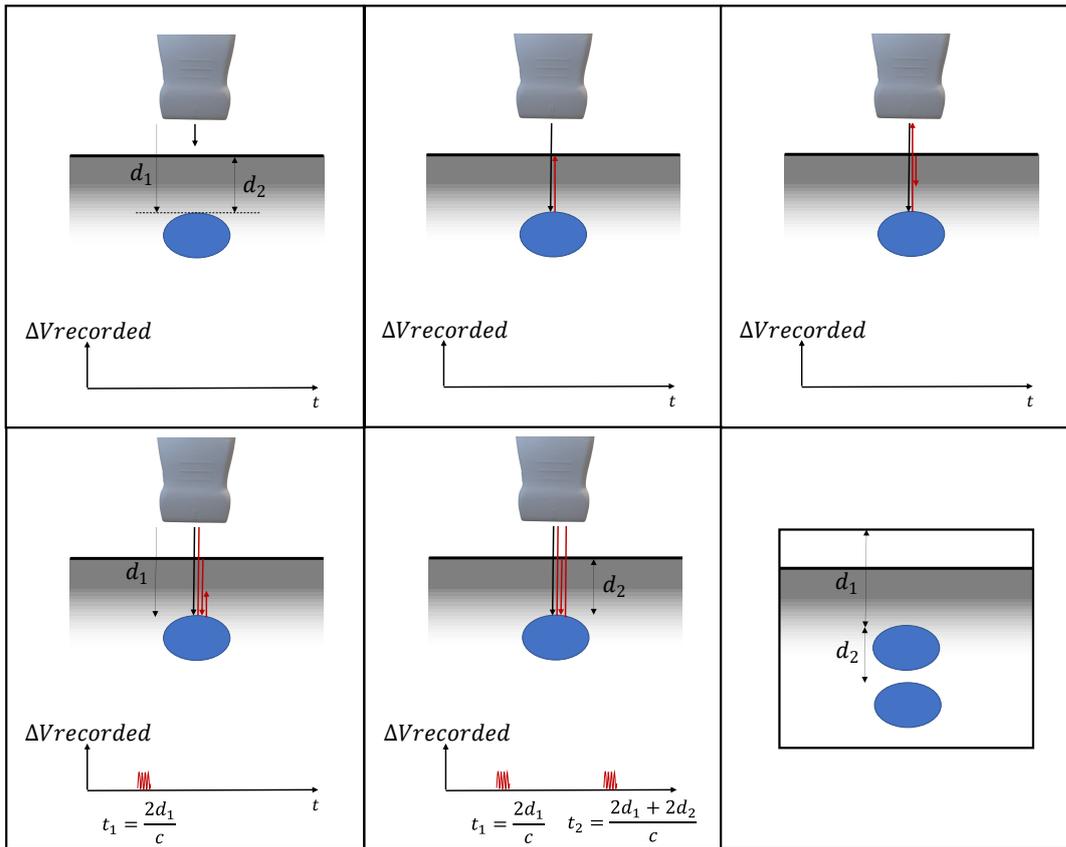


Fig. 5.14: Representation of multiple reflections against the ultrasound transducer, showing the reflection of the echo from the blue structure back to the transducer and then back to the structure, resulting in the creation of multiple echoes at different timestamps

significant portion of the incoming energy, causing the energy of the wave passing over the surface to be highly attenuated. As a result, structures beneath highly reflective surfaces will be reached by a weak ultrasound wave, resulting in a low-energy echo signal and the creation of dark regions in the reconstructed ultrasound image (Fig. 5.15).

Speed Displacement

Earlier, when deriving the relationship between the location of a point in space and the time at which its echo signal is recorded, we made the assumption that the propagation speed of ultrasound waves is constant. While this is a reasonable approximation, the speed of sound can vary among different tissues. If we assume a constant speed of sound, v_c , then we would expect a structure at a distance d from the transducer to generate an echo at $t_{echo}^{ideal} = 2dv_c$. However, if the actual speed of sound in the tissue is higher than the assumed value, v_{real} , then the reflected echo will occur earlier than expected, at $t_{echo}^{real} = 2dv_{real} < t_{echo}^{ideal}$. As a result, the echo received from the structure at a distance d from the transducer will be incorrectly associated with the

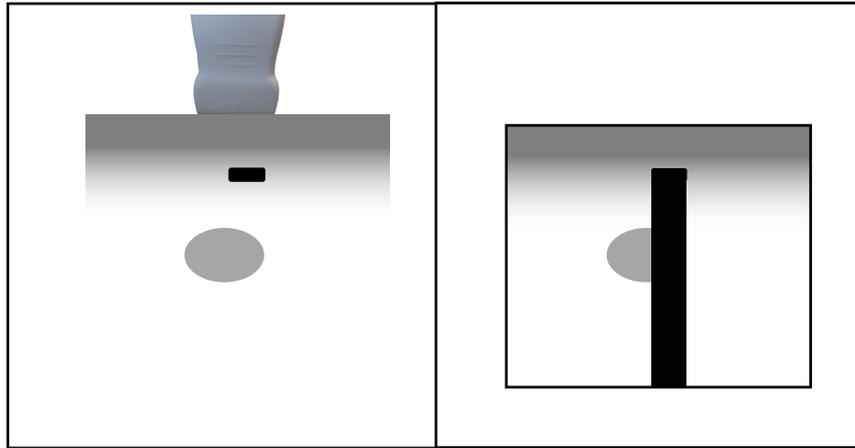


Fig. 5.15: Illustration of the acoustic shadow created by a highly reflective structure (black rectangle) partially obscuring structures beneath it

point $d = t_{echo}^{real}/2v_c$, rather than $d = t_{echo}^{real}/2v_{real}$. This leads to a vertical shift of the imaged structure compared to its actual position (Fig. 5.16).

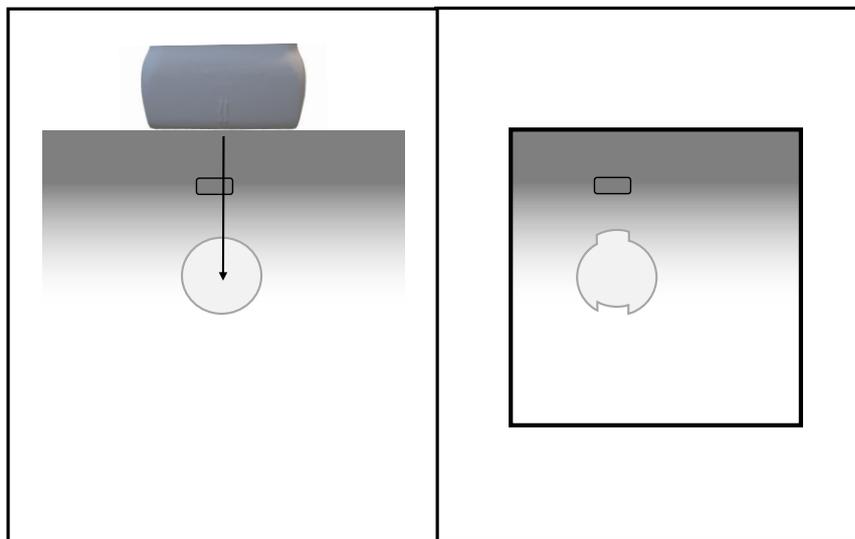


Fig. 5.16: Image showing the deformation of a structure caused by the varying propagation speed of the ultrasound wave along different scanlines

5.0.3 Ultrasound Imaging and Deep Learning - Data Augmentation

Ultrasound is a valuable imaging modality that allows for the acquisition of body images without exposing the patient or operator to ionizing radiation. However, as we have seen, ultrasound images can be affected by various artifacts, noise, and potentially low signal-to-noise ratios. To address these issues, researchers have attempted to use deep learning networks to automatically process and de-noise

ultrasound data. Ultrasound data are often presented as B-mode images, which are pixelized matrices where each spatial location represents the reflective property of the surface point at that location in the body. Due to the similarity with photographic images, it is common in the literature to transfer techniques used for photographic image analysis to ultrasound data. An example of this is data augmentation, which involves introducing random transformations to the input data during the training of the neural network. These transformations are intended to artificially increase the input data distribution and introduce new transformations that will help the network generalize better to real data. Common augmentations used in photographic imaging include rotation, scaling, translation, and in some cases, random affine or deformable transformations. The main scientific question addressed in this thesis is: is it reasonable to use such transformation techniques for ultrasound data, or should we focus on identifying specific transformations that better represent the data distribution of ultrasound B-mode images? The answer to this question is somewhat complex. When we augment data during training, we impact the network in multiple ways: (i) we force the layers of the network to correctly classify features even in the presence of transformations that do not occur in the training dataset but may occur in the test set, and (ii) we force the network to focus on and base its classification on transformation invariant features. This means that, for example, the effect of using Gaussian noise for data augmentation will be noticeable regardless of whether Gaussian noise is present in the actual data distribution. When trained with Gaussian noise augmentation, the network will not only learn to classify images in the presence of Gaussian noise, but it will also be forced to focus on features that are invariant to the injected transformation. In simpler terms, any form of data augmentation should ideally force the network to "isolate" features that are invariant to these transformations and use them for classification. Therefore, augmentation in general is advisable regardless of the overall data distribution of the training/testing population. On the other hand, in cases where the training dataset is very small, as is often the case with medical data, the subset of training data may not be representative of the actual data distribution and may not fully reflect its variance. In these cases, introducing data-specific augmentation may be beneficial in providing the network with additional training data to learn features that are not present in the original dataset or to ensure that the network learns features that are invariant to application-specific transformations. Therefore, it is recommended to use both standard and modality-specific data augmentation for network training. We expect standard augmentation techniques to improve the training as they can help the network learn transformation-invariant features, which are likely to be characterizing features of the structure being classified. Modality-specific transformations will help the network learn any transformation or feature that is not present in the training dataset, thus giving cues toward relevant features which are not possible in general but are specific and likely to occur in that specific modality.

5.1 Ultrasound 3D Volume Compounding

Ultrasound volume compounding consists in combining a set of tracked ultrasound images into a volumetric object. There are three main techniques to perform volume compounding: voxel-based and pixel-based methods [19]

5.1.1 Voxel-based methods

In voxel-based volume compounding techniques, the values of the voxels in the target volume are determined by considering the nearest pixels among the acquired ultrasound images. These pixels are often weighted according to their proximity to the voxel, and the value assigned to the voxel is calculated as the weighted average of these pixels. (Fig. 5.17).

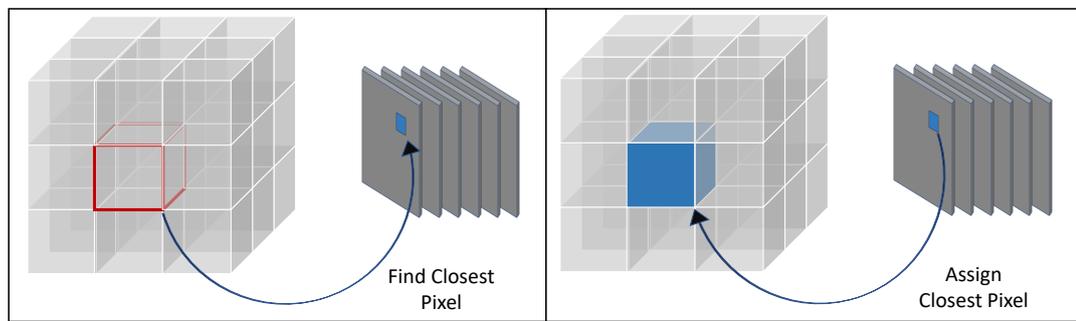


Fig. 5.17: Schematic representation of the voxel assignment in the case of voxel-based methods for volume compounding

This approach ensures that all voxels in the target volume are filled, however, it should be noted that in areas of the volume where there is insufficient coverage from the ultrasound images, the values assigned to the voxels may not accurately reflect the true values, potentially leading to artifacts in the resulting compounded volume.

5.1.2 Pixel-based methods

In pixel-based image composition techniques, the values of the voxels in the target volume are determined by considering the corresponding pixels in the ultrasound images. The value of each pixel is assigned to the closest voxel in the target volume. When multiple pixels are assigned to the same voxel, the values can be interpolated or the value of the closest pixel can be retained (Fig. 5.18).

It should be noted that this approach may result in empty holes in the compounded volume if no voxel is found to be closest to a pixel in the source ultrasound data. These holes can be filled by interpolating the values of nearby voxels. However, it should be

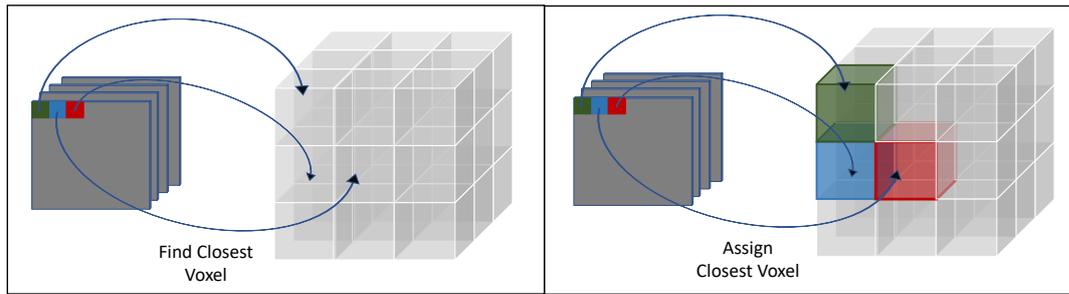


Fig. 5.18: Schematic representation of the voxel assignment in the case of pixel-based methods for volume compounding

noted that, as for the previous method, in areas of the target volume where there is insufficient coverage from the source images, the values assigned to the voxels may not accurately reflect the true values, potentially leading to artifacts in the resulting composite volume.

Part II

Contributions

Visual-Tactile Fusion for Autonomous Robotic Scanning

The material presented in this chapter was published in [91]:

Maria Tirindelli^{1,*}, Maria Victorova^{2,*}, Javier Esteban¹, Seong Tae Kim¹, David Navarro-Alarcon², Yong Ping Zheng² and Nassir Navab^{1,3}, "Force-ultrasound fusion: Bringing spine robotic-us to the next "level" ", *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5661–5668, 2020

¹ Technical University of Munich

² The Hong Kong Polytechnic University

³ Johns Hopkins University

* Equal contribution

<https://ieeexplore.ieee.org/abstract/document/9140314>

6.1 Introduction and Motivation

In this chapter, we introduce a novel approach for integrating visual feedback derived from ultrasound data and haptic feedback from force sensors for the purpose of actuating a robot in autonomous scanning tasks. Within this context, the specific application of spinal injections is examined. Spinal injections are commonly utilized in neurosurgical procedures for the release of anesthetics in pain treatment, such as epidural injections, facet joint injections, and nerve blocking [66], [77]. The precise placement of the needle is of vital importance to prevent unwanted side effects. Due to the periodic shape of the spine, injections at incorrect vertebral levels are a potential hazard. To mitigate this risk, procedures are typically guided by either palpation or X-ray guidance. Palpation, while safe as it does not expose the patient or surgeon to ionizing radiation, can be error-prone as it relies heavily on the expertise of the surgeon and the identification of specific anatomical landmarks in the spine. On the other hand, X-ray guidance offers a more accurate alternative to palpation [7], but exposes both the patient and surgeon to ionizing radiation, making it inadvisable for certain populations of patients. As a safe and accurate alternative to X-ray guidance, ultrasound guidance has been proposed in the literature. Compared to palpation, ultrasound guidance provides a more comprehensive, real-time representation of the patient's anatomy, increasing the overall accuracy of the procedure [20]. However, compared to x-ray imaging, ultrasound imaging presents challenges such as image interpretation due to noise and artifacts. Noise and artifacts in ultrasound images of the spine are primarily caused by the highly reflective nature of the bones, leading to multiple reflections and mirroring artifacts. Additionally, the complex organization and shape of vertebral bones, in conjunction with their angle relative to the ultrasound probe, can result in artifacts arising from reflections in undesired directions. To alleviate these problems, several authors have proposed utilizing deep learning techniques to automatically process ultrasound images, providing the user with higher-level information such as the current vertebral location or the injection point location, without the need for manual interpretation of the data.

6.1.1 Current State of Automatic Vertebral Classification and Robotic Ultrasound for Spinal Injections

The identification of the correct vertebral level is crucial for spinal injections to ensure a proper injection site. Some authors propose utilizing a technique known as panorama stitching to "stitch" together ultrasound data in order to obtain a full lateral section representation of the spine. This approach, as described in studies such as [22] and [41], can be used to identify vertebral levels by extracting the periodic, hyper-echoic regions in the ultrasound data. These hyperechoic regions can be further

processed using techniques such as enhancing filters to extract the bony structures and finding local minima on the extracted pattern [22] or template matching to extract specific landmark locations [41].

Alternatively, the ultrasound probe can be moved along the spine with a transverse orientation, and images of the spine can be acquired and classified as the probe moves upwards. In the study by [55], a Convolutional Neural Network (CNN) was used to classify the ultrasound data acquired along the spine as either vertebrae or intervertebral gaps. The classification was then passed to a state machine, which processed the data in time, and outputted a prediction of the current vertebral number.

The advancement of technology in the field of medicine has led to the development of robotic ultrasound systems as a means of addressing the challenges faced by surgeons in performing spinal injections. In a study by [62], a semi-automatic system was designed and tested to address these challenges. In the proposed pipeline, the surgeon initiates the scan, which is then performed autonomously by the robot. The surgeon then identifies the injection location on the acquired volume and the robot navigates to the corresponding position on the patient's body, performing the injection using a calibrated needle guide rigidly attached to the ultrasound probe holder. Though promising results were demonstrated in clinical practice, further improvements are needed to reduce the reliance on manual input from clinicians for navigation.

6.1.2 Fusing Haptic and Visual Feedback for Robotic-Ultrasound Scanning

This chapter presents a novel technique for combining visual and haptic feedback in order to guide a robot in the task of scanning and identifying vertebral levels for spinal injections. As described in [91], this work addresses the limitations of current methods by taking inspiration from expert sonographers, who rely on both visual and haptic feedback to identify vertebral levels. In our proposed system, we utilize force sensors to gather information on the outer profile of the patient's anatomy, and ultrasound data to acquire information on the inner anatomy. By fusing this data, we are able to provide a robust and accurate localization of vertebral levels. Compared to existing methods, our proposed system integrates automatic classification and analysis of sensor data for robotic actuated spinal ultrasound acquisitions. To the best of our knowledge, this is the first system to rely on both haptic and visual feedback for vertebra localization, mimicking the visual and palpation feedback of standard, manually performed procedures.

6.2 Fusing Haptics and Vision for Robot Control - Method

In this section, we will discuss the methodology we used for the system design, data acquisition, and processing pipeline. Before doing so, we will give a brief introduction to relevant anatomical structures and notation, which we will refer to throughout the chapter.

6.2.1 Useful Notation and Anatomy

The spine consists of 33 vertebrae. In this work, we will focus on the 5 lumbar ones, where facet joint injections are typically performed. The lumbar vertebrae are labeled from L1 to L5 moving from the sacrum location upwards (Fig. 6.1).

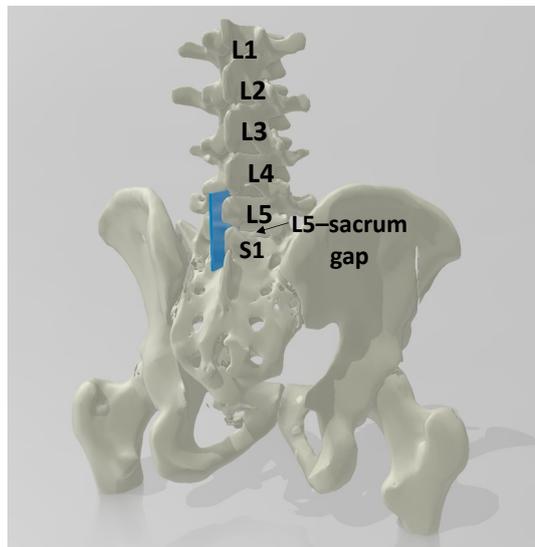
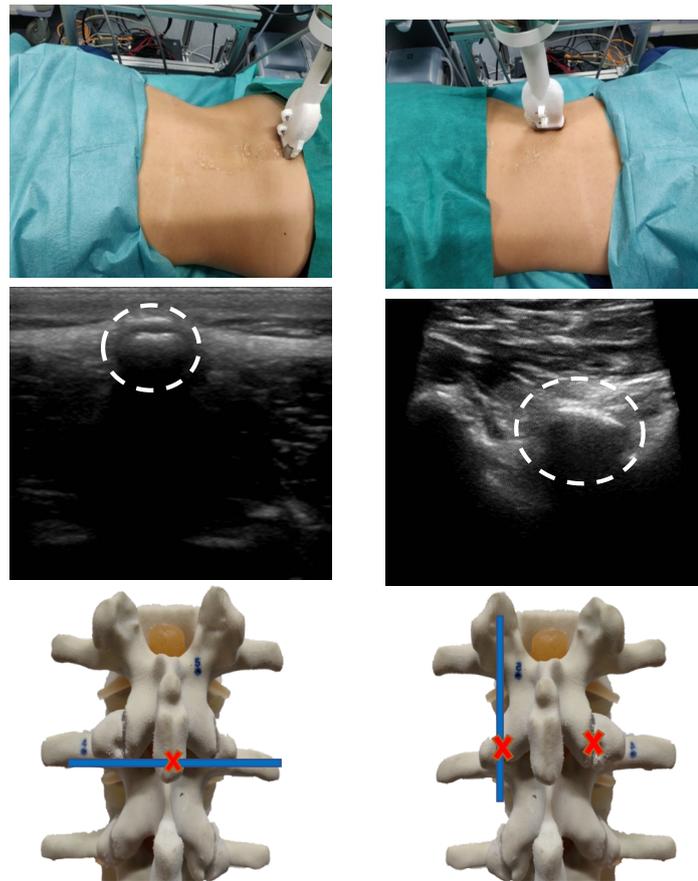


Fig. 6.1: The lumbar vertebrae in the human spine

The most prominent part of each vertebra, located at its axis of symmetry, is called **spinous process**. The articular joint connecting subsequent vertebrae are called **facet joint**. We call the **transverse probe orientation** the orientation of the probe perpendicular to the spine direction and **paramedian-sagittal probe orientation** the orientation of the probe parallel to the spinal cord 6.2b.

6.2.2 System Design

The utilized robot ultrasound system consists of the following elements:



(a) Data acquisition with a probe in transverse orientation and the respective ultrasound image of the spinous process.

(b) Data acquisition with the probe in paramedian sagittal orientation and the respective ultrasound image of the facet joint.

Fig. 6.2: The probe orientation with respective ultrasound views ©2020 IEEE

- A KUKA LBR iiwa 7 R800 robotic arm, equipped with 6 joints and links and certified for human interactions.
- A FTD-GAMMA (SCHUNK GmbH & Co. KG) 6-axis force-torque sensor.
- A custom-designed, 3D printed ultrasound probe holder
- A zonare L8-3 linear probe with purely linear and trapezoidal imaging modes.
- An Epiphan DVI2USB 3.0 frame-grabber (Epiphan Systems Inc. Palo Alto, California, USA). The frame grabber was set to use a 800x600 resolution and 30 fps sampling rate.)
- A workstation equipped with an NVIDIA Titan V GPU with 12 GB of graphic memory.

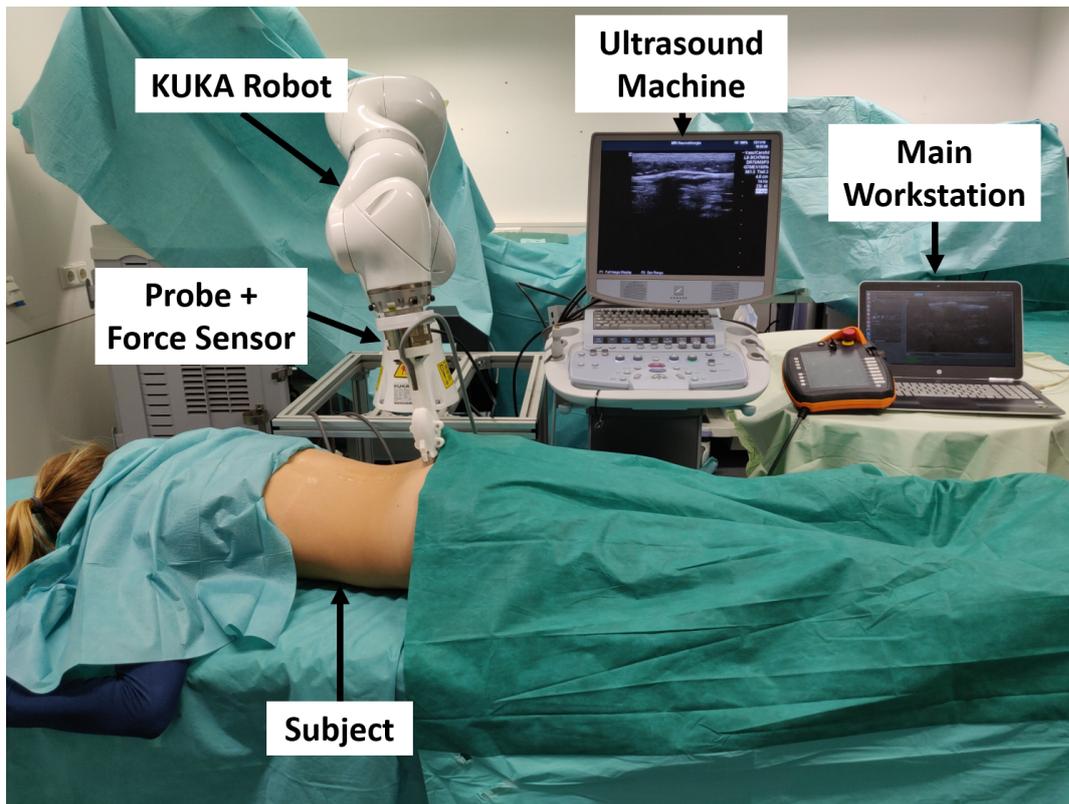


Fig. 6.3: The system setup ©2020 IEEE

6.2.3 Autonomous Scanning Pipeline

During the autonomous scanning, the patient lies on the treatment bed in a prone position. The robot is manually placed at the level of the sacrum location with the probe in a transverse orientation. The robot autonomously moves upwards along the spine direction. During the acquisition, ultrasound data and force data are acquired simultaneously. The acquisition lasts for around 10 seconds and is performed in breath-hold. Upon identification of the vertebral level, the ultrasound probe takes a further scan of the desired vertebra with a paramedian-sagittal orientation and automatically identifies the facet plane. An enhanced image of the facet can be additionally computed and provided to the sonographers to ease the injection.

6.2.4 Data Acquisition for Parameters Tuning and Model Training

To fine-tune the parameters of the acquisition and to train and validate the deep-learning classification models, three offline datasets were acquired (Table 6.1).

Dataset I

The dataset I was acquired on 19 subjects and consists of ultrasound data only. During the acquisition, the probe was robotically navigated from the sacrum location up the spine, in order to fully cover the whole lumbar section of the spine. The probe was oriented with a transverse orientation and used with a pure linear imaging modality. The force applied by the robot was kept constant at 2 Newtons and the speed to 20 mm/s. From this dataset, we manually labeled each frame as either being a vertebra or a vertebral gap.

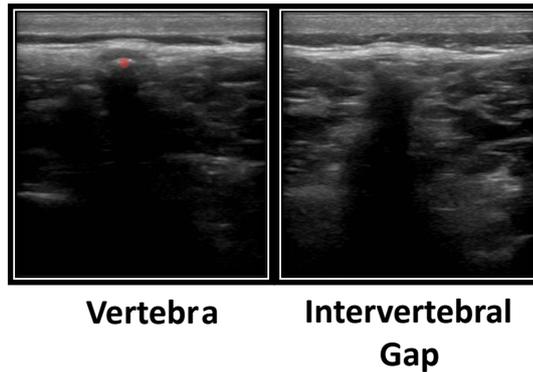


Fig. 6.4: Vertebra Classification as it was done for Dataset I

Dataset II

Dataset II was acquired on 14 subjects and consists of ultrasound data and force data. As for dataset I, the probe was robotically navigated from the sacrum location up the spine, in order to fully cover the whole lumbar section of the spine and oriented with a transverse orientation and used with a pure linear imaging modality. The acquisition was performed with 3 different force values, i.e., 2, 10, and 15 N, and with 3 different robot speeds, i.e. 12, 20, 40 mm/s. Before each acquisition, we manually counted the vertebrae using both the ultrasound probe and palpation and labeled each ultrasound data with classification labels defining vertebrae and non vertebrae frames. The force data were synchronized with the ultrasound data, such that the classification was valid for both data.

Dataset III Dataset III was acquired on 19 subjects and consists of ultrasound data only. The probe was robotically navigated over the whole spine with the probe in a paramedian-sagittal orientation and the probe used in a trapezoidal, steered modality. The force applied by the robot was constant and set to 2 N and the speed to 5 mm/s. The data are both classified as containing a facet or non-containing a facet, similar to what was done in dataset I for the spinous process. Additionally, we manually segmented the bones on the frame where they are visible (Fig. 6.5).

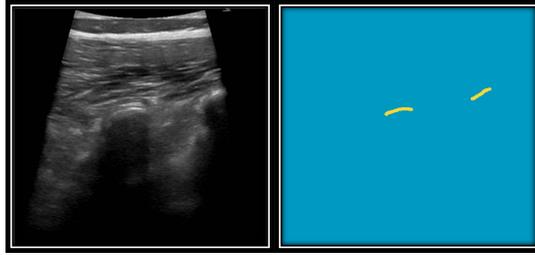


Fig. 6.5: Vertebral Bone Labelling in Dataset III

Tab. 6.1: DATASET TABLE WITH CORRESPONDENT SIZE, DATA AND SENSOR SETTINGS
©2020 IEEE

Dataset	N. Subjects	Acquired Data	Probe Orientation	Applied Force [N]	Robot Speed [mm/s]
Dataset 1	19	B-Mode Linear US	Transverse	2	20
Dataset 2	14	B-Mode Linear US Force Data	Transverse	[2, 10, 15]	[12, 20, 40]
Dataset 3	19	B-Mode Convex US	Paramedian-Sagittal	2	5

6.2.5 Force Data Extraction

When performing the robotic-actuated acquisition, we utilize a hybrid force-impedance controller with the force-controlled direction being the z-axis of the robot, perpendicular to the patient body. This means that we expect the force applied along the z direction to always be constant. Therefore, we can model the interaction between the robot and the robotically actuated ultrasound probe as follows.

If we assume that there are no vertebrae below the ultrasound probe, the tissue can be modeled as a homogeneous tissue, and we can say that in this case the probe purely moves along the y direction, and not along z. That is, the reaction forces from the tissues balance the force applied by the robot. If a vertebra is present below the ultrasound probe, the local direction of the surface along which the probe moves suddenly changes. If we consider the slow motions of the probe, at a given timestamp t_0 we will still have the reaction forces balance the applied force along the z-axis. However, due to the inclination of the surface, we now also have the development of a force along the y direction (see Fig. 6.6a). This force can be measured via a force sensor and results in a signal like the one depicted in Fig. 6.6b. The description given above holds true for the ideal case of a robot moving at a constant (slow) speed along the y-axis, ensuring that the y-component is pure to be associated with the reaction force with the vertebral body. However, the robot experiences an initial acceleration and final deceleration during its motion, causing a low-frequency drift

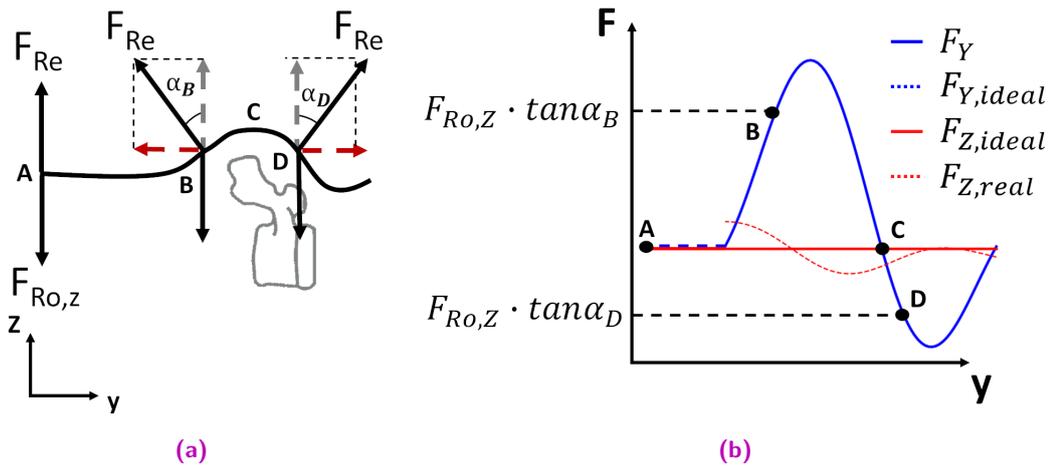


Fig. 6.6: (a) The modeled interaction between robot and patient back during the robotic scanning procedure. (b) Z component (red) and Y component (blue) of the force signal recorded over a single vertebra. ©2020 IEEE

in the signal. To correct this, we first compute the low-frequency drift by applying a low-pass second-order Butterworth filter with a cutoff frequency of 0.05 Hz. We then subtract the computed drift from the original signal. Finally, we low-pass filter the resulting signal with a second-order Butterworth filter with a cutoff frequency at 0.3 Hz, we normalize the signal between 0 and 1 and we resample such that samples are equally spaced in space (instead of time) (Fig. 6.7).

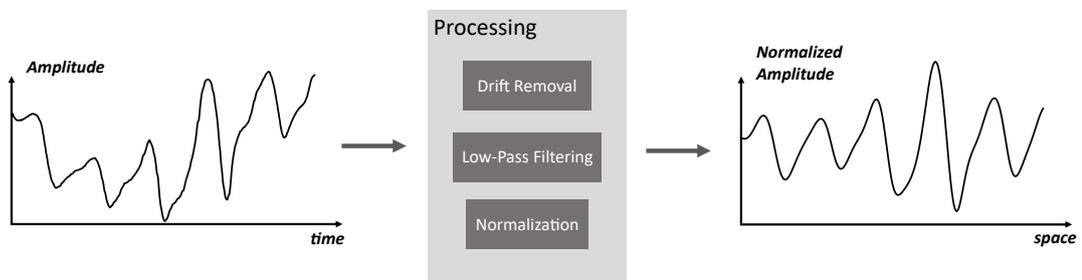


Fig. 6.7: The pre and post-processed force signal.

Within this work, we studied the effect of the subject BMI and desired force along the z-axis on the recorded force signal on the axis y. Expectedly, we noticed that for subjects with lower BMI, the vertebrae profile is well visible in the recorded signal, independent on the BMI value. For subjects with higher BMI, vertebral levels are better visible with higher forces (of 10 to 15 N). A profile of the force signal measured along the y direction, with respect to the force applied along the z direction, is reported in Figure 6.8 for two subjects with BMI values above and below 23.

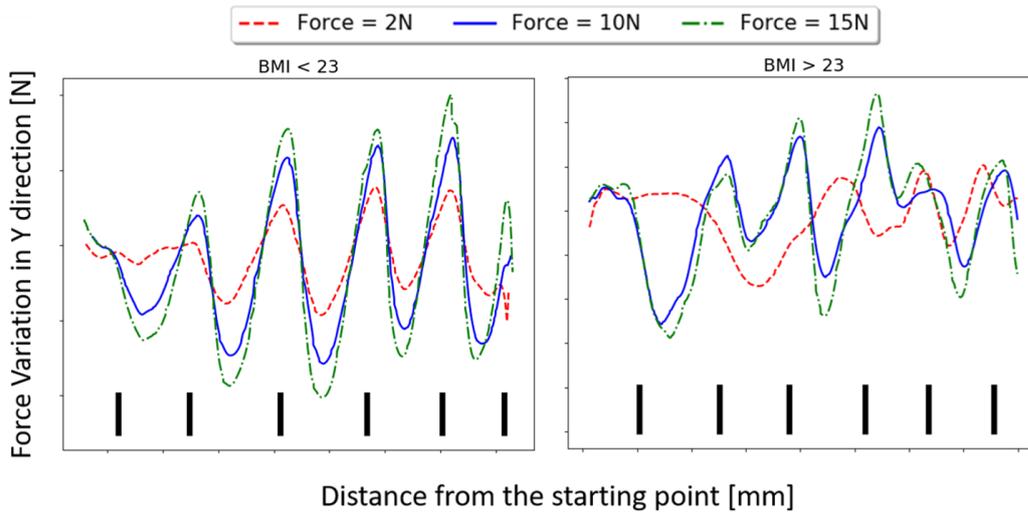


Fig. 6.8: The force signal recorded in the y-axis with 3 different values (2, 10 and 15 N) of the z-force applied by the robot for subjects with $BMI < 23$ (left) and for subjects with $BMI > 23$ (right). ©2020 IEEE

In light of these findings, we empirically set the acquisition force to 10 N for subjects with $BMI < 23$ and to 15 N for subjects with $BMI > 23$.

6.2.6 Ultrasound Data Extraction and Processing

What we obtain from the ultrasound scanning is a sequence of B-mode images each of which is associated with a tracking matrix. The signal we are interested in is a 1-D signal providing us information on the location of vertebral level along the spine. To obtain a such signal, we, therefore, need to process the tracked ultrasound data to reduce them into a 1D vector. For such purpose, we utilize a Convolutional Neural Network trained for the task of binary classification, where for each frame we extract its probability of containing a vertebra. For the network training, we utilize Dataset 1, where we manually labeled the ultrasound data as either "vertebra" or "intervertebral gap". We then fuse the classification value with the corresponding timestamp to get a classification in space on the location of the vertebrae. For the classification, we evaluated 3 different states of art classification neural networks: ResNet18 [38], DenseNet121 [45], VGG11 with batch normalization [40] and we compared their classification performance in the following cases: a) Using random initial weights b) Using pre-trained weights obtained from the network pre-trained on ImageNet [24] as an initialization for the weights c) Using pre-trained weights obtained from the network pre-trained on ImageNet [24], and freezing them during training except for the last layer (i.e. fine-tuning the last layer only).

The ground truth labels are binary values where 1 corresponds to the presence of a vertebra in the analyzed image and 0 corresponds to no vertebra in the image. For training, we used Cross entropy as the loss function, Adam optimizer, and initial learning rate equal to 0.0005

During the procedure, the b-mode images belonging to the ultrasound sweep acquired along the spine are processed using the trained Convolutional Networks. Therefore, we obtained a set of classification predictions for each image in the sweep. Each of these classification predictions is a value between 0 and 1 corresponding to the probability of that given bmode image containing a vertebra. We then concatenate these values spatially along the y-direction utilizing the tracking data coming from the robot. To remove high-frequency noise coming from misclassification, we low-pass the obtained signal using a II order Butterworth filter with 0.3 Hz cutoff frequency, and resample the obtained signal such that the samples have equal mutual distance in space.

6.2.7 Haptic-Visual Data Fusion

The rationale behind fusion haptic and visual data is to increase the robustness of the acquisition procedure. Similarly, as sonographers do, we leverage haptics and visual data, in a way to have a better understanding of the patient's anatomy. In fact, it can be the case sometimes that vertebral levels are well detectable via palpation but not visible on ultrasound data due to noise and artifacts. This can be appointed to multiple reflection artifacts due to the reflectivity of bones and spine geometry, as well as shadowing artifact and suboptimal probe-skin coupling which is likely to occur for thin patients, where the spinous process is prominent and doesn't allow for a flat contact area. In some other cases, vertebrae might not be easily detectable via palpation but can be detected on the acquired ultrasound data. This can be the case for subjects with particular spine anatomies or high BMIs. An example of this is reported in Fig. 6.9a.

To account for these problems, we introduce a novel method to fuse force and ultrasound data. To this end, we utilize a Temporal Convolutional Network.

Temporal Convolutional Network

A Temporal Convolutional Network, or TCN, is a specialized type of Convolutional Neural Network designed to handle temporal data, such as time series or videos. These networks are built using a series of 1D convolutional filters, which can be applied in both a causal and acausal manner to extract relevant temporal patterns from the input signal. By stacking multiple layers of these convolutional filters, TCNs

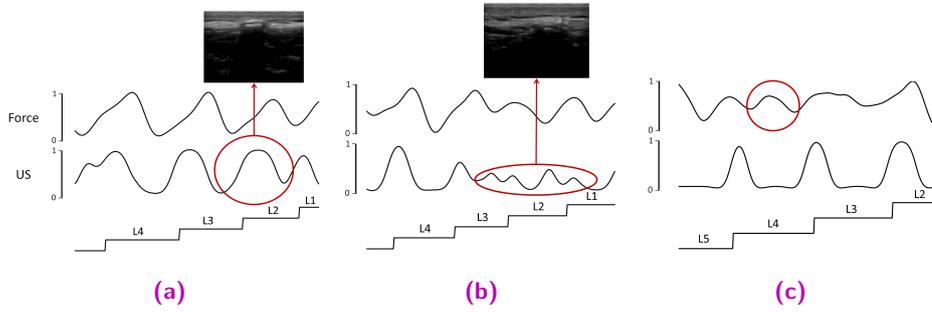


Fig. 6.9: (a) Force signal, Ultrasound signal, and labels in the presence of non-corrupted force and ultrasound data. (b) Force signal, Ultrasound signal, and labels in the presence of noisy ultrasound data. (c) Force signal, Ultrasound signal, and labels in presence of noisy force signal. ©2020 IEEE

can learn increasingly complex temporal patterns in the data, similar to how 2D convolutional networks learn increasingly complex spatial patterns at different scales. Compared to LSTMs, TCNs offer several advantages, such as larger temporal receptive fields and less susceptibility to issues like vanishing gradients. Additionally, the use of dilated convolution can further expand the capture range of TCNs, allowing them to effectively handle long-range dependencies in the data. The dilated convolution between an input signal F and a kernel k is defined as:

$$(F *_l k)_t = \sum_{s+lt=p} F(s)k(t) \quad (6.1)$$

where l is the dilation factor. It can be noticed that with increasing values of the dilation factor, the receptive field of each output sample gets wider.

In this work, we utilized a temporal convolutional network, which architecture is devised from [72]. The architecture contains three different stages, each one directly trained to regress the vertebral level associated with each input sample. Subsequent stages have the purpose to refine the output of the previous one. The input data consists of a 2-channel signal, where one channel is the signal coming from the force sensor, processed as described in Seq. 7.2.1 and the second channel consists of the 1D signal extracted from the ultrasound data as described in Seq.7.2.2. We call N the number of samples in the input signal. The architecture of each stage is the same (Fig. 6.10).

Architecture of the Single Stage: The architecture of the network is the following: *i* A 1×1 convolutional layer containing 32 filters, which resizes the input signal into a $32 \times N$ sequence. *ii* 9 $1 \times D$ dilated convolutional layer with increasing dilation sizes and kernel size 3.

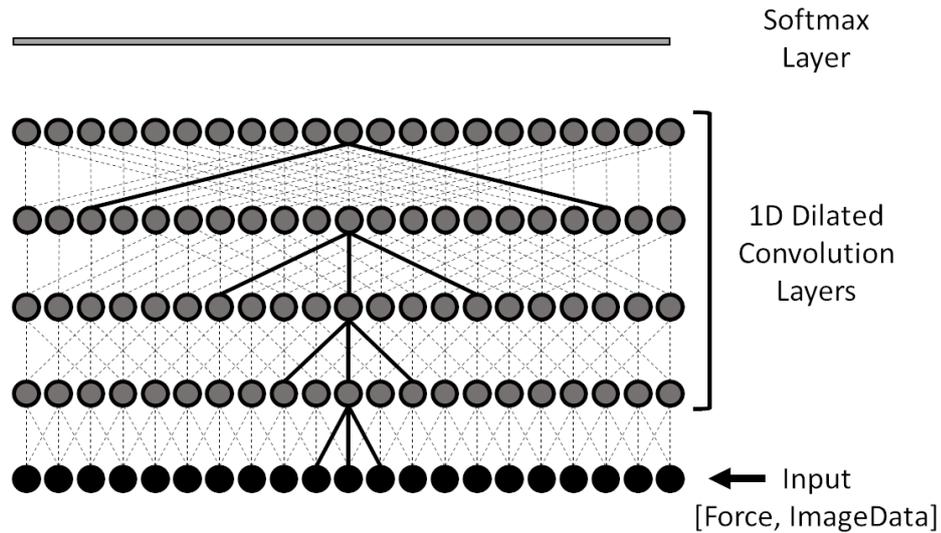


Fig. 6.10: The architecture of the single stage of the 1D convolutional network for data fusion.
©2020 IEEE

Similarly to [72] we used cross-entropy for network training, batch size 1 Adam optimizer, the learning rate of 0.0005 and we trained for 110 epochs. We used 34 input sequences acquired over 9 subjects for training (7 for training, 2 for validation), and 4 subjects for test. For the test subjects, we set the acquisition force to 10N or 15N depending on the subject BMI and the robot velocity to 20 mm/s.

6.2.8 Facet Plane identification

In order to localize the facet joints in an ultrasound image, the plane that contains the joints is identified using a ResNet18 CNN trained with Adam optimization, cross-entropy loss, and a learning rate of 0.0005 that decays by 0.1 every 5 epochs for 30 epochs. The training data set consists of 15 subjects, 12 for training and 3 for validation, and the test set consisted of 4 subjects. Next, the position of the facet joints on the identified plane is localized using a FCN-ResNet101 CNN trained to regress heatmaps. This method has been shown to produce state-of-the-art results for landmark position detection. The input to the network is a 3-channel image which includes the original ultrasound image, the Bone Probability Map, and the image gradient along the y-direction. The network was trained for 50 epochs with an initial learning rate of 0.001 on 13 subjects, validated on 3 subjects, and tested on 4 subjects, with 7 vertebral scans per subject.

6.3 Experiments

Tab. 6.2: RESULTS OF 5-FOLDS CROSS-VALIDATION STUDY FOR VARIOUS MODELS WITH DIFFERENT TRAINING MODES. ©2020 IEEE

	ResNet18	DenseNet121	VGG11
Case a	0.817 ± 0.118	0.878 ± 0.047	0.635 ± 0.15
Case b	0.929 ± 0.006	0.89 ± 0.014	0.878 ± 0.055
Case c	0.6 ± 0.02	0.577 ± 0.006	0.63 ± 0.03

6.3.1 Ultrasound Signal Processing

In table 6.2 we report the accuracy results for the binary classification networks trained on ultrasound data for vertebrae presence recognition. We can see that the best accuracy is obtained by utilizing pre-trained weights obtained from pretraining the network on ImageNet, yielding an accuracy of 0.929 ± 0.006 in 5 cross-fold validation experiments. We then tested the network yielding the best performance among the 5 trained for the 5 cross-fold validation tests and tested on the test set, yielding an accuracy of 0.938, showing that the training distribution well represents test one.

In table 6.3 we reported the confusion matrix evaluated on the test data, where we normalized the values by the total number of frames. The values expressed in a number of frames are reported in parentheses.

Tab. 6.3: CONFUSION MATRIX FOR THE BEST MODEL PERFORMANCE EVALUATED ON THE TEST SET OF 4 SUBJECTS. ©2020 IEEE

		Predicted	
		Vertebra	Intervertebral Gap
Actual	Vertebra	True Positive 0.459 (n = 640)	False Negative 0.04 (n = 56)
	Intervertebral Gap	False Positive 0.02 (n = 30)	True Negative 0.478 (n = 666)

6.3.2 Force-Ultrasound Data Fusion

We evaluated the capability of the Temporal Convolutional Network to correctly label the vertebral levels. We tested the network on a group of 5 unseen subjects, thus on 25 vertebral levels. We utilized both the TCN and a conventional peak detector (CPD) for the task of vertebral level counting. We set the parameters of the peak

Tab. 6.4: THE CLASSIFICATION PERFORMANCES AND DISTANCE FROM THE GROUND TRUTH VERTEBRAE POSITION FOR ALL TESTED METHODS. FOR THE TCN METHODS THE RESULTS ARE REPORTED AS MEAN (STD) FOR THE 5-FOLD CROSS VALIDATION. ©2020 IEEE

	Correctly Classified Levels [num/total]				Distance from Ground Truth Label [mm]				
	Below	Average	Above	Overall	Below	Average	Above	Overall	
CPD	Image	0.4	0.73	1.0	0.72	27.359 (26.9)	9.85 (14.0)	3.079 (1.96)	9.74 (15.95)
	Force	0.2	0	0	0.04	20.01 (9.779)	37.58 (7.10)	30.7 (3.70)	32.09 (10.3)
	Fusion	1.0	0.933	1.0	0.96	2.495 (3.2)	2.357 (1.8)	2.386 (2.196)	2.39 (2.23)
TCN	Image	0.48 (0.09)	1.0 (0.0)	0.68 (0.097)	0.832 (0.03)	10.93 (0.90)	3.7 (1.17)	8.224 (1.20)	6.05 (0.735)
	Force	0.439 (0.079)	0.92 (0.06)	0.72 (0.16)	0.784 (0.04)	14.74 (2.8)	6.18 (1.79)	8.88 (1.46)	8.43 (1.02)
	Fusion	0.439 (0.149)	1.0 (0.0)	0.6 (0.0)	0.808 (0.03)	12.64 (1.6)	3.76 (0.99)	8.72 (0.70)	6.52 (0.5)

detector empirically to 0.5 amplitude threshold and 10 samples of the minimum distance between peaks. We evaluated the performance of the proposed pipeline in the following scenarios, for both the TCN and the peak detector:

1. Using the pure image signal in the input
2. Using the pure force signal in the input
3. Using the fusion of force and image signals in the input

We obtained the fusion signal for the peak detector as the sum of the force and ultrasound signals. For the peak detector, each detected peak leads to a transition to the next vertebral level.

In our analysis, we considered the performance of the networks with respect to the height of the subjects. To this end, we divided the test subjects into 3 groups of heights: *i*) height < 163cm (below average), *ii*) height between 163cm and 183cm (average) and *iii*) height > 183cm (above average). The height threshold was selected based on the height distribution in the training set, where the mean is equal to 173cm and the standard deviation to 10cm. For each height group, we analyzed the performance of the TCN and peak detector when using the three different input signals, both in terms of classification accuracy and distance to the corresponding peak in the ground truth labels. We define success in classifying vertebral levels as having an overlap of more than 0.5 between the labels and predictions, which is in line with previous research in the field (as seen in [55]). The result of the analysis is reported in table 6.4.

We can see that the Temporal Convolutional Network yields better results for subjects with average height compared to the peak detector, in the case of a pure image, pure force, and fusion signals as input. The TCN's ability to learn anatomical priors during the training process is a possible explanation for its ability to classify vertebral levels so accurately for average spinal lengths (i.e. average heights). This means that the TCN is able to identify vertebral levels not just based on the presence or absence of peaks in the input signals, but also by taking into account what it has learned about the structure of the spine. This is particularly evident when looking at results using the pure force signal only as the input. In these cases, L5 is often not detectable as it is less prominent compared to upper lumbar vertebrae and this often results in a missing peak at the beginning of the force signal. Additionally, the robot's initial acceleration can also add noise in this region in the force signal (as seen in Fig. 6.11).

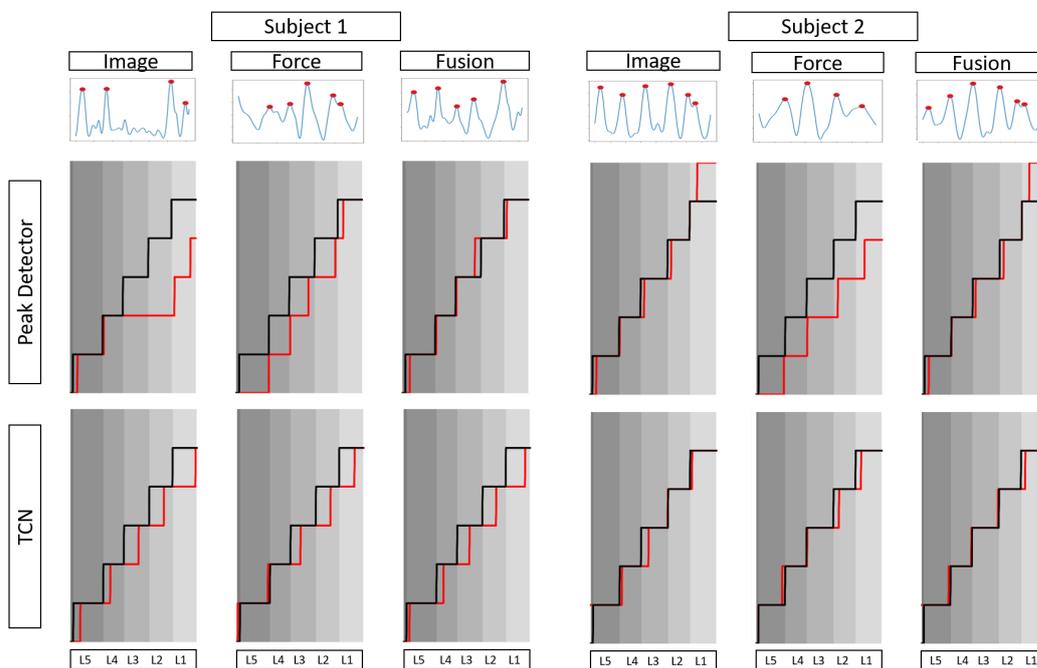


Fig. 6.11: The predicted (red line) and ground-truth (black line) vertebral levels for pure force-based, pure ultrasound-based, and force-ultrasound fusion both when using a peak detector and a tcn. Subject 1 with anatomical characteristics non-well-represented in the tcn training set (Subject Gender: Male, BMI: 30, Height: 186 cm). Subject 2 with anatomical characteristics well-represented in the tcn training set (Subject Gender: Female, BMI: 22, Height: 172 cm). ©2020 IEEE

When dealing with subjects with heights far from the training distribution, the peak detector method yields better performance compared to the TCN. This suggests that the distribution of the training dataset might not be sufficiently representative of these categories. It is our belief that retraining the network with a larger and more representative training dataset would be beneficial for the TCN performance. We can notice that when using the CPD for below and above-average subjects, the fusion method yields better performance compared to the pure image and pure force

methods, suggesting that fusion haptic and visual data has an impact on the vertebral classification accuracy.

For the TCN, we can notice that the pure force and pure signal methods have similar performances which are slightly higher than the pure force method. It is relevant to notice that the ground truth was obtained by manually labeling the ultrasound data, and was not done considering the force signal. This way of labeling the input data has an impact on the performance of the TCN. Specifically, it is biasing the results towards the pure-image-based TCN, meaning that, given that the network correctly classifies a vertebral level for all input data (i.e. force, image, fusion), it is likely that the distance from the ground truth peak is lower for the pure image data, given that the ground truth was defined on images. Therefore, the pure image method can be considered an upper bound in terms of distance to the ground-truth data.

6.3.3 Facet Plane Identification

The spatial discrepancy between the planes identified by the model and the labeled planes was evaluated using a sample of 4 test subjects from Dataset 3, which included 20 vertebrae sweeps and a total of 40 facet joints (two joints per sweep and five sweeps per subject). Out of the 40 joints evaluated, 37 had a mean distance error of 2.08 ± 2.63 mm between the detected and manually labeled planes. This error is within the range of 5 mm or less, as noted in [12], which is considered sufficient for successful facet joint injections. However, for the remaining 3 joints, the error was found to be 8.43 ± 8.98 mm. An example image of the enhanced facet is presented in figure 6.12.

6.4 Conclusions

In conclusion, we proposed a novel method to automatize spinal injections by means of robotic-ultrasound acquisition. In this work, we proposed a pipeline for automatic vertebral level counting and facet plane detection. Our proposed method leverages the visual information coming from ultrasound data and the haptic information coming from force sensors to obtain robust results on the vertebrae's location along the spine. To this end, we proposed an interaction model between the ultrasound probe and the patient's back, and we derived from it an interpretation of the data acquired via the force sensor. We further proposed a processing pipeline for reducing the drifts in the force signal and reducing the ultrasound data into a 1-D signal consisting of the probability of each location containing a vertebral level. We proposed two methods for the fusion of haptic and visual data. A standard CPD, which showed robust performance independent on the training set size, for different heights of the

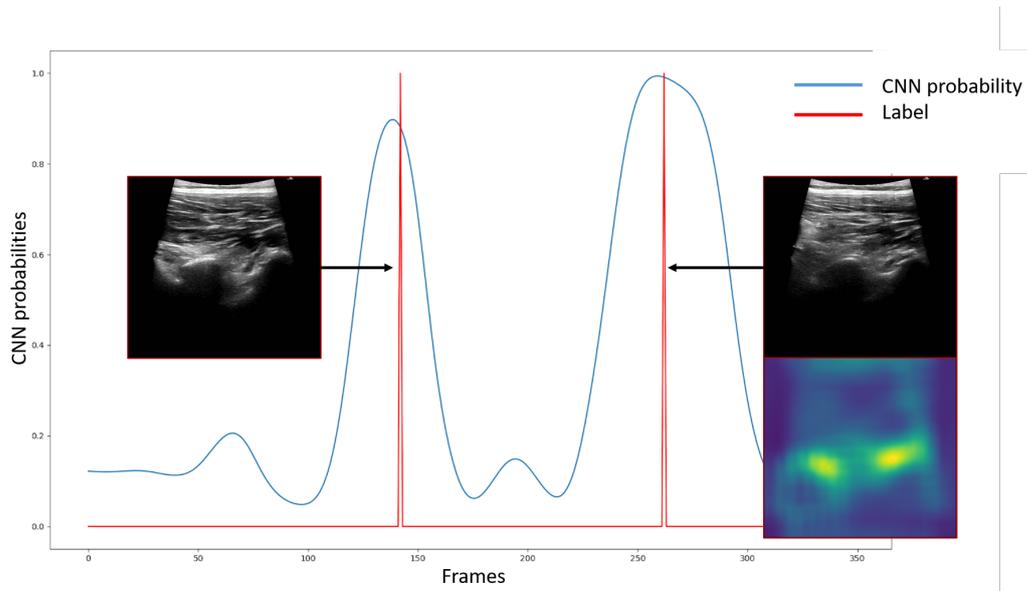


Fig. 6.12: The result of the facet detection and facet enhancement for one sample test sweep

test subjects and a TCN method. Expectedly, the TCN showed lower performance on data far from the training distribution but showed promising results in average-height subjects suggesting its potential for the procedure in presence of larger training datasets.

Physics-Inspired Augmentation

The material presented in this chapter was published in [115]:

Maria Tirindelli^{1,*}, Christine Eilers^{1,*}, Walter Simson¹, Magdalini Paschali¹, Mohammad Farid Azampour^{1,2} and Nassir Navab^{1,3}, *Rethinking ultrasound augmentation: A physics-inspired approach*, Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. Springer International Publishing, 2021.

¹ Technical University of Munich

² Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

³ Johns Hopkins University

* Equal contribution

https://link.springer.com/chapter/10.1007/978-3-030-87237-3_66

7.1 Introduction and Motivation

This chapter introduces ultrasound physics-inspired data augmentation approaches for improving learning algorithms in ultrasound imaging tasks such as classification and segmentation. An improved accomplishment of these imaging tasks paves the way toward autonomous robotic ultrasound acquisitions.

Deep learning methods have thus far revolutionized machine perception of the visual world. And recently, these approaches are rising up to improve common ultrasound imaging tasks such as classification, segmentation, and registration **empty citation**. However, several challenges inherent in ultrasound data hinder the progress and success of these algorithms in the ultrasound imaging domain. The main culprits are low signal-to-noise ratio (SNR) and the presence of artifacts arising from reverberation, phase aberration, and scattering. The issues, in addition to inter-operator dependency, render the interpretation of ultrasound data difficult for deep learning models and humans. The deep learning models face an additional challenge; the ultrasound image datasets are restricted due to the difficulties inherent in image acquisition and annotation. Moreover, the ultrasound datasets must reflect the significant variability within acquired data resulting from different operators and ultrasound systems for the deep learning models to be generalizable. Data augmentation is a remedy for the aforementioned challenges arising from limited data and lack of variability.

7.1.1 Current State of Augmentation Approaches in Ultrasound Imaging

Data Augmentation in machine learning refers to augmenting the dataset for the purpose of improving the generalizability of models. For instance, one common approach in the visual domain is to apply transformations to the data points, such as rotation and flipping. These transformations keep the semantics of the data points intact and provide the model with more data and more cues toward the predictive features within the dataset. Intuitively, augmentation approaches that modify the data assist the model in learning features that are invariant to the modifications and, depending on the choice of modification approach (e.g., rotation transformation), can lead to model generalizability. Data augmentation becomes specifically effective in the regime of limited data. And as discussed earlier, ultrasound image datasets are limited in magnitude. Moreover, due to the difficulty in understanding the features within the image and the various features introduced by artifacts, augmentation can be additionally helpful for guiding the models toward learning generalizable features.

In the domain of ultrasound image analysis, we can categorize the data augmentation approaches into two groups: image synthesis approaches and image modification approaches. In the first category, we can generate new data points and augment the dataset by training a generative model on the dataset [70], [79], [90], [95], [106], [110], [111]. The generative models can be optimized without any need for data annotation and thus have access to larger datasets. For instance, Zaman et al. [95] leverage generative adversarial networks (GANs) to generate synthetic ultrasound images that simulate variable roll and pitch motions, depths, and time-gain compensations. While a promising approach, generative-based approaches require large datasets. Moreover, they can generate unrealistic data points depending on the resulting generative model and thus misguide the models trained on them towards features that are not generalizable. The second category applies modifications to the images. The common modifications are random translations, rotations, scaling, and adding noise (e.g., gaussian) [51], [60], [64], [67], [71], [78], [80], [83], [87], [89], [92]. Specifically, [71], [89], [92] apply random flipping, [64], [71], [80], [83], [87], [89], [92] apply random shifting, [64], [67], [80], [87], [90], [92] apply rotation transformation, and [80], [83] modify the brightness of the images. An interesting image modification approach is that of Baka et al. [51] for the task of ultrasound bone segmentation in ultrasound spine images. In addition to mirroring, [51] leverages localized free-form deformations.

7.1.2 Intelligent Augmentation by Leveraging Physics of Ultrasound

Albeit being helpful by guiding the model toward learning features invariant to these modifications, these modifications generate unrealistic samples that risk the generalizability of the models since these unrealistic samples do not represent the variability in the test data. Some of these modifications, such as horizontal flipping 7.2 are reasonable as it is equivalent to Π rotation of the probe. However, a vertical flipping (7.2) does not abide by the physics of attenuation as it leads to a shadowing region between the probe and a reflector. Translation and rotation transformations also generate unrealistic points as the transducer will be displaced to meaningless positions, resulting in voids between the wave source and tissues in the acquisition. The unreal-generated augmented samples motivate our work in this chapter. We leverage the physical principles governing ultrasound imaging to apply realistic modifications to ultrasound images. Specifically, we leverage the mechanics of tissue deformation while applying the transducer to subjects. We introduce realistic artifacts by considering reverberation and generate scans with various signal-to-noise ratios. The augmented data thus has a higher likelihood to match the distribution of unseen ultrasound images on test datasets and real scenarios.

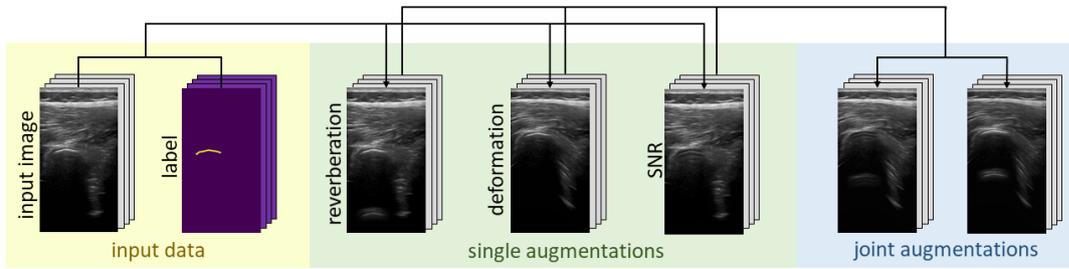


Fig. 7.1: The proposed physics-inspired data augmentation framework: using ultrasound scans and their corresponding bone annotations (yellow), we generate augmentations leveraging the physical properties of ultrasound scanning (green) and can also merge multiple augmentations (blue).

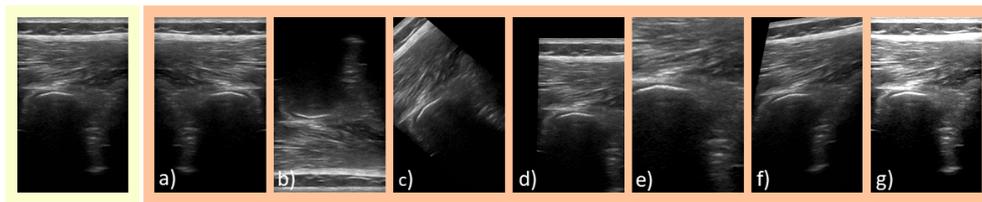


Fig. 7.2: Examples of applying classical augmentation approaches to ultrasound scans: a) horizontal flipping, b) vertical flipping, c) rotation, d) translation, e) scaling, f) shearing, g) brightness adjustment.

7.1.3 Contributions

Inspired by the physical principles governing ultrasound imaging, we propose a set of data augmentation methods that augment the dataset with samples more likely to appear in test scenarios, thus improving the generalizability of the models. The approaches utilize the mechanics of tissue deformation, the physical effects resulting in reverberations, and the engineering of signal-to-noise ratio. The variability introduced by these augmentation techniques is physically and anatomically consistent with ultrasound scans. The presented methods are thoroughly evaluated on principal ultrasound image analysis tasks, bone segmentation, and classification. The improvement on these tasks translates to the downstream task of intelligent robotic ultrasound imaging. In addition to the methodological contributions, we introduce a spine ultrasound dataset that can foster future research efforts in robotic ultrasound imaging.

7.2 Physics-inspired Data Augmentation

7.2.1 Augmentation by Deformation

During ultrasound acquisition, the operator (robot or sonographer) applies a variable force on the scanned subject. Applying force by placing the transducer on the scanned subject induces deformation in the scanned tissues and the resulting ultrasound image. The deformation caused by the variability in applied force can thus be a major factor in the variability existing in ultrasound images across datasets. Inspired by this phenomenon, we propose simulating this deformation on the ultrasound images to perform data augmentation.

To simulate the deformation on a specific ultrasound image arising from a probe displacement d_{probe} , we need to compute the probe displacement's corresponding deformation field (DF). We denote the deformation at each point (x, y) by $\Sigma(x, y)$. The deformation field can be characterized using the physical laws governing the applied force to the area (tensile stress $\sigma = \frac{F}{EA}$) and the compression (ϵ). The relationship is modulated by Young's modulus, $Y = \frac{\sigma}{\epsilon}$. To simulate the deformation field arising from the probe displacement on an ultrasound scan, we impose a few assumptions for simplification. We assume that the probe is applied vertically on the medium (along the axial axis) and that the soft tissues are homogenous and isotropic. The bone is also considered a body without deformation. We also consider the lateral deformation negligible and assume that the compression happens only on the tissues between the transducer and the static body (the bone). In principle, we can remove this assumption and compute the compression for the lateral tissues. However in this work, we resort to simulating the compression of the tissues in between the probe and the bone and simulate the lateral compression using a gaussian smoothing term. Please refer to Fig. 7.3 for an overview of the method.

Following the assumptions and considering the coordinate system placed on the bone (Fig. 7.3), the deformation for the lateral tissues will be constant and equal to the probe displacement d_{probe} . For the tissues in between the transducer and the static body, the compression is computed by $\epsilon = \frac{F}{EA}$, where A denotes the probe area in contact with the medium and E denotes Young's modulus. The relationship can be reformulated as $F = EA \frac{d_{\text{probe}}}{y_{\text{probe}}}$, where y_{probe} denotes the vertical coordinate of the probe as seen from the bone coordinate system. The compression characterizes the deformation at each point arising from the transducer's force on the area in contact. Thus, the deformation at each point $(\Sigma(x, y))$ is

$$\Sigma(x, y) = \int \epsilon_{yy} dy \Big|_{\Sigma(x,0)=0} = -H(-y) \frac{F}{EA} y = -H(-y) \frac{d_{\text{probe}}}{y_{\text{probe}}} y \quad (7.1)$$

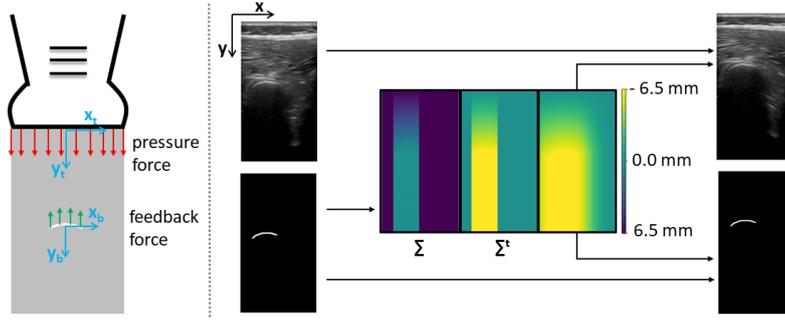


Fig. 7.3: Augmentation by Deformation Pipeline: Left: The image displays relevant forces and coordinate systems. The force applied by the transducer is represented with red arrows, and the feedback force from the bone is with green arrows. The coordinate system placed on the bone is represented by (x_b, y_b) and the transducer centroid coordinate space by (x_p, y_p) . Right: The B-mode ultrasound scan and its corresponding bone annotation. The bone annotation is used in the pipeline for computing the deformation field (Σ) induced by the transducer force. The deformation field is displayed in the transducer coordinate system to avoid generating a gap on top of the ultrasound scan. We apply a gaussian smoothing on (Σ^t) to account for and approximate the lateral forces. The resulting deformation field is then applied to the original scan and the annotation.

The deformation field computed by Eq. 7.1 is in the coordinate system placed on the bone. Applying it to the ultrasound image will result in generating voids under the transducer. Therefore, we compute the deformation field in the coordinate system of the probe so that the applied transformation does not introduce the void to the ultrasound image. The new deformation field can be computed by adding an offset equal to d_{probe} . The deformation field in the probe's coordinate system is denoted by Σ^t , and y_{bone} denotes the vertical position of the bone. The deformation field Σ^t is applied to the tissues in between the probe and the bone. The lateral deformations are approximated by a gaussian smoothing term. The methodology is described as a pseudo algorithm in Algorithm 1.

Algorithm 1 Deformation Field Computation (Σ^t)

```

for  $i, j = 1 : \text{Width, Height}$  do
   $\Sigma(i, j)^t = -d_{probe}$  if  $(i, j)$  is bone or below bone
   $\Sigma(i, j)^t = -d_{probe}/y_{bone} \cdot j$  if  $(i, j)$  is above bone
   $\Sigma(i, j)^t = 0$  else
end for
smooth( $\Sigma^t(i, j)$ )

```

To augment the dataset with deformed samples, we compute the deformation for multiple randomly selected probe displacement d_{probe} values.

7.2.2 Augmentation by Reverberation

A commonly occurring artifact in ultrasound data is due to the reverberation phenomenon. Knowing the physics of this phenomenon, we can simulate the artifact in images and augment the datasets realistically by introducing the reverberation artifact. The phenomenon is due to the sound wave being echoed multiple times on a highly reflective matter. Assuming a constant speed of sound c in a medium, having bounced from a reflective tissue at Δ , the wave generates an echo in $t_1 = 2\Delta/c$. If the reflected wave is bounced back from a highly reflective surface, the sound wave can reflect back from the transducer-tissue interface and generate another echo at $t_2 = 4\Delta/c$ and can give the impression that another reflective surface exists at a distance Δ_r :

$$\Delta_r = 2\Delta = t_1 c = \frac{t_2 c}{2} \quad (7.2)$$

As the wave can bounce several times and generate multiple echoes, the phenomenon is known as reverberation. The reverberation artifact can then show several times in multiples of Δ within the ultrasound scan. Knowing the principles behind the reverberation artifact within ultrasound images, we can simulate them and augment the dataset by reverberation artifact. To accomplish this, we first identify the reflective surface within the image that leads to the reverberation effect. In our use case, the reflective surface is the bone; thus, we first identify the bone centroid (x_c, y_c) , and the centroid is represented in the coordinate probe space. Based on the previous statement regarding reverberation, the location of the artifact will then be in $2y_c$. Therefore we select a patch around the bone centroid and attempt to merge the patch into the image in location $2y_c$ such that it shows up as a reverberation artifact. For merging the path with the image, we perform a weighted sum of the two. We denote the weighting factor by w_r and consist of a Gaussian filter and a scaling factor r_i . We refer to this scaling factor as reverberation intensity, as it controls the intensity value of the reverberation artifact. For augmenting the dataset, we sample the reverberation intensity r_i randomly. During the merge step, the original image is weighted by $1 - w_r$.

7.2.3 Augmentation by SNR Modification

In this section, we augment the dataset by tampering with the signal-to-noise ratio (SNR). The signal in the B-mode ultrasound scans refers to coherent structures such as bones, while the noise refers to speckle patterns. In essence, our method can enhance or diminish the intensity values of the signal within the image. SNR variability is a common property of ultrasound scans from different sources. Thus augmenting

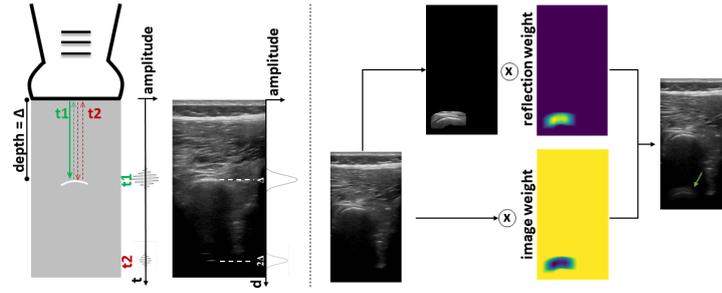


Fig. 7.4: Augmentation by Reverberation Pipeline. Left: A schematic representation of the transmitted wave echoing on the bone at depth Δ and time t_1 and the reverberation echo happening at time t_2 . In the following time-amplitude plot, the signals arising from the first echo t_1 and the reverberation echo t_2 are represented. The echo effect at depth Δ and the reverberation effect at depth 2Δ are shown on the following B-mode image. Right: Synthetic generation of reverberation artifact on B-mode scans for the purpose of data augmentation. The bone is extracted from the original ultrasound scan and repeated at the reverberation depth and is finally added to the original B-mode scan.

the dataset via SNR modification increases the likelihood of matching the training distribution with the testing and real-world ultrasound data distribution.

To modify the SNR, we need to first identify the signal and distinguish it from the noise. For instance, we need to distinguish between the bone structures and the background. Subsequently, we can enhance the signal, i.e., rendering the bones more visible. We leverage local energy (LE) maps **Bridge** for identifying the signal within B-mode ultrasound images. LE is a methodology that computes the energy of the signal present at different pixel locations in the image using the monogenic signal components derived from an ultrasound image. The monogenic components are $f_e(x)$, $f_{o1}(x)$ and $f_{o2}(x)$, and the local energy at each location x in the image is computed as:

$$LE(x) = f_e(x)^2 + f_{o1}(x)^2 + f_{o2}(x)^2 \quad (7.3)$$

Having normalized the images by local energy values, we multiply the local energy values with different scaling factors for bone and non-bone pixels. The local energy of bone pixels is scaled by i_b , and the non-bone pixels are scaled by i_{bg} . These factors modulate the energy ratio between bone and background. Specifically, whenever $i_b > i_{bg}$, we enhance the bone energy values, and whenever $i_b < i_{bg}$, the background pixel energy values are enhanced. I.e., we are enhancing bone visibility versus background and vice versa. In the end, we modify the B-mode ultrasound scan by the updated local energy values, thus modifying the SNR of the scan. For generating various samples for augmentation, we randomly select values for i_b , and i_{bg} .

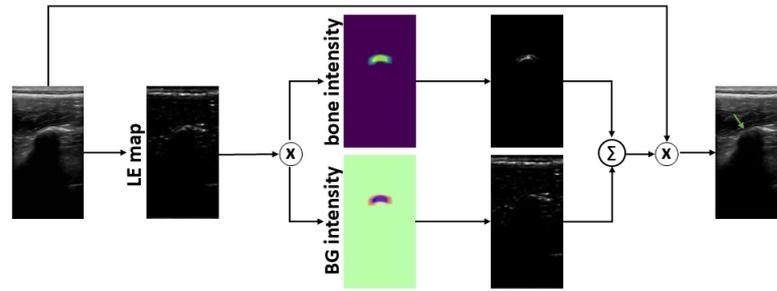


Fig. 7.5: Augmentation by SNR modification pipeline: First, we compute the local energy (LE) map using the input US scan. We then multiply the computed LE map with the bone intensity and background intensity values and apply weighting to obtain tuned LE maps. Subsequently, we sum the tuned maps and multiply them with the original scan to compute the SNR-modified scan.

7.3 Experiments

The augmentation approaches in this chapter are introduced to improve ultrasound image analysis tasks that facilitate intelligent autonomous robotic ultrasound acquisitions. Two prominent tasks are ultrasound image classification and segmentation:

7.3.1 Classification

To carry out the task of bone classification in ultrasound images, we procured a dataset comprised of 5,692 frames retrieved from 22 individuals. The frames were divided into two disjoint classes depending on their content. The first class contains at least one bone per frame, while the other class contains no bones in each frame. To ensure the generalizability of the trained models, we employ a subject-level approach in splitting the dataset into three distinct subsets: The training set contains 3821 frames from 16 subjects, the validation set contains 1037 frames from 4 subjects, and the testing set contains 834 frames from 2 subjects. This approach makes the model less prone to overfitting to a specific group of subjects and more probable to generalize to new subjects.

7.3.2 Segmentation

For segmenting bones in ultrasound images, a dataset consisting of 5,284 frames that contain bones is procured. The frames are annotated in a pixel-wise fashion; that is, each individual pixel is labeled as either a bone or non-bone. To improve the generalizability of the models, a subject-level approach is leveraged in dividing the dataset into three distinct subsets similar to the classification tasks: The training set

Tab. 7.1: The range of parameter values for the augmentation approaches (classical and physics-inspired) in our experiments. For each augmentation algorithm, its parameter is uniformly selected within the range.

Augmentation	Parameter	Value Range
Classical	rotation in degree	-10 - 10
	translation (horizontal and vertical)	0.2, 0.2
	scaling in both axis	1, 1
	shearing in both axis	1, 1
	brightness	0.2
Proposed	deformation - d_{probe}	30 - 100
	reverb - r_i	0.50 - 0.9
	SNR - i_b, i_{bg}	0.70 - 1.40

(3,972 frames from 17 subjects), validation set (782 frames from 5 subjects), and testing set (500 frames from 2 subjects).

7.3.3 Dataset Acquisition

We trained the models on a dataset of 10,656 ultrasound scans acquired from 24 healthy individuals with a Body Mass Index (BMI) range of 20 to 25. The ultrasound images were gathered using a Zonare z.one ultra sp Convertible Ultrasound System (Zonare Medical Systems Inc.) and an L8-3 linear probe. The parameters of the system were set to the specific following configuration; with the image depth set to 7cm, the gain set to 92%, the frequency set to 10 MHz, and the sampling rate set to 13 Hz.

7.3.4 Implementation Details

The neural network models are implemented using the PyTorch 1.7.1 framework and trained on an NVIDIA Titan V 12 GB HBM2 GPU utilizing the Polyaxon platform. The source code for our proposed approaches, experiments and the procured dataset is publicly available to facilitate reproducibility and future research. In order to assess the effect of each of our proposed augmentation approaches, we incorporated them individually in network training on the segmentation and classification tasks and compared them against existing data augmentation techniques. The chosen parameters for augmentation techniques are detailed in Table 7.1. Our proposed augmentation techniques were only applied during the training step for both the segmentation and classification tasks.

Tab. 7.2: Results of applying augmentation approaches to the two tasks of segmentation (UNet) and classification (DenseNet). The results are the average and the standard deviation (\pm) of 5-fold cross-validation

	Segmentation		Classification	
	DSC	HDF	ACC	F1
None	0.589 \pm 0.07	20.72 \pm 3.84	0.876 \pm 0.06	0.770 \pm 0.15
Classical	0.625 \pm 0.03	17.76 \pm 3.17	0.883 \pm 0.04	0.780 \pm 0.09
Reverb	0.604 \pm 0.03	19.71 \pm 2.20	0.883 \pm 0.03	0.802 \pm 0.04
Deformation	0.626 \pm 0.01	19.06 \pm 3.63	0.865 \pm 0.04	0.759 \pm 0.11
SNR	0.626 \pm 0.02	17.24 \pm 1.83	0.877 \pm 0.06	0.764 \pm 0.16
All	0.600 \pm 0.02	17.32 \pm 2.97	0.834 \pm 0.02	0.742 \pm 0.04

7.3.5 Neural Network Architectures and Training

We leverage a U-Net neural network architecture **Ronneberger** for the bone segmentation tasks. We utilize a DenseNet121 **Huang** for the classification task. In terms of training, a learning rate of 0.01, Adam optimizer, and Binary Cross Entropy (BCE) loss function were used as the training parameters. The utilized architectures and training parameters are standard in the literature and have repeatedly been shown to be effective in many studies.

7.3.6 Evaluation Metrics

To assess the performance of the models, we used 5-fold cross-validation for both segmentation and classification experiments. To evaluate the segmentation model, we used two mainstream metrics: Dice Score (DSC) and the Hausdorff Distance (HDF). We also reported the standard deviation of these two metrics, which indicate to what degree the results vary between the folds. For the classification task, we utilized average accuracy (ACC) and F1-score (F1), which are widely used in the literature to evaluate the classification models' performance. We also reported the standard deviation of the results between folds.

7.4 Results and Discussion

The results for both segmentation and classification experiments are reported in Table 7.2 for various augmentation techniques: classical augmentation, augmentation by reverberation, deformation, and SNR modification. We also present the results for the case of applying all proposed augmentations (presented as "All").

Classification

We see the result of applying each augmentation technique during the training of the bone classification model in Table 7.2 on the right column. The results demonstrate that the model trained using reverberation-based augmentation outperforms other augmentation approaches by 2.7% in the F1-Score metric. This highlights the potential of reverberation-based augmentation in improving model performance. In line with our findings on the segmentation task, it can be observed that using a combination of all augmentation techniques does not provide any additional benefit in the classification task.

Segmentation

In the context of segmentation performance, we observe that (in Table 7.2) augmentation approaches are significantly effective, as evidenced by the increase of DSC (Dice Similarity Coefficient) by 2.4% and the improvements in HDF (Hausdorff Distance). The segmentation task is more challenging than the classification, and we observe that the data augmentations' effectiveness becomes more evident. Augmentation by deformation and SNR modification outperform classical augmentation methods by 1%. Moreover, augmentation by SNR modification results in the lowest HDF distance among all methods. Furthermore, we observe that utilizing the proposed augmentations demonstrates the least amount of variation among the folds, thus indicating the consistency of their performance of improvement. By combining all of the proposed augmentations, it is observed that there is a decline of 2% in performance. The decline may be caused by the absence of the combined cases in the unseen test sets.

Summary of Findings

The findings suggest that the incorporation of augmentation by deformation and SNR modifications demonstrate a positive impact on the task of bone segmentation, while augmentation by reverberation is found to have a positive impact on the bone classification task. Future investigations could aim at investigating more realistic and anatomically-consistent ultrasound image modifications for training more generalizable neural networks. The improved models will, in turn, lead to more intelligent robotic ultrasound imaging systems.

Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark

Viviana Sutedjo^{1,*}, Maria Tirindelli^{1,*}, Christine Eilers^{1,*}, Walter Simson¹, Benjamin Busam¹ and Nassir Navab^{1,2}, Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark., *IEEE Robotics and Automation Letters*. 2022 Jan 10;7(2):1808-15.

¹ Technical University of Munich

² Johns Hopkins University

* Equal contribution

<https://ieeexplore.ieee.org/abstract/document/9140314>

8.1 Robotic Path Optimization

8.1.1 Introduction and Motivation

It has previously been established that the utilization of robotics-ultrasound systems presents the advantage of being able to acquire tracked ultrasound data. This is achieved through the synchronization of ultrasound image data with the 3D pose of the robot. By utilizing information on the geometry of the image, such as the spacing and 3D position, the position of each pixel in space can be inferred and a 3D volume can be reconstructed, according to what is described in section 5.1.

There are several methods for determining the optimal trajectory of the robot to acquire the ultrasound scan. One straightforward approach involves manually defining a region of interest above the anatomy to be scanned and performing a parallel ultrasound scan. This method is simple and does not require additional hardware, but it does not account for the presence of ultrasound artifacts or suboptimal coupling between the ultrasound probe and skin. To address these issues, an initial trajectory can be estimated based on the surface to be scanned. For example, authors have proposed using an RGB-D camera to acquire the mesh of the patient's external surface and design a trajectory to ensure the probe moves perpendicular to the surface on a region of interest [43], [44]. Others have proposed adjusting the ultrasound probe trajectory in real time based on the acquired ultrasound data to maximize image quality. As an example, [35] proposes to compute the ultrasound image entropy and adjust the position and orientation of the probe to maximize such value. Other authors propose using confidence maps to navigate the robot and maximize image confidence [37], [52].

Although these approaches have shown promising results, they also have some limitations:

1. Despite efforts to acquire a trajectory perpendicular to the patient's surface, non-flat surfaces can still result in incomplete coverage of the anatomy of interest.
2. The visibility of the structure of interest can be negatively impacted by the presence of shadowing structures.
3. Even when navigating the robot to maximize confidence/entropy, some areas of the volume of interest might remain uncovered.

In this chapter, we present a novel methodology for computing robot trajectories. Our inspiration for this methodology comes from the field of outdoor area scanning, where small motorized robots gather information about the environment by traveling

around unseen areas [84]. In such a field, the purpose of the robot is *i* to explore as many unseen areas as possible in the region of interest, *ii* try to reach unseen areas by avoiding known occlusions, *iii* be aware of seen regions and avoid re-scanning. Analogously, in the context of ultrasound data acquisition, our objective is to thoroughly examine the anatomy of interest while also striving to gather as much information as possible about regions that may be impacted by artifacts such as acoustic shadowing. To achieve this, it is necessary to *i* Establish a trajectory that ensures full coverage of the area of interest, *ii* Determine areas that are not visible (i.e., accurately classify regions affected by artifacts), *iii* Identify alternative trajectories to access areas impacted by corruption.

To accomplish this objective, we introduce a novel approach for determining the robot's trajectory as follows:

1. We establish a first optimizer to guarantee complete coverage of the area being scanned. The first optimizer is agnostic to the presence and location of shadowing structures and only operates based on the defined region of interest, regardless of its shape. A first ultrasound scan is performed based on the result of this first optimization process
2. A second optimizer analyzes the data collected during the first acquisition and performs two tasks: a. It identifies areas affected by shadowing and b. It suggests alternative trajectories that would allow approaching the corrupted areas from different angles.

The proposed method is modular and offers multiple advantages: *i*. It provides a general optimizer for volume coverage that is independent of the shape of the scanned area and can be modified with additional regularised according to the specific requirements of the application. *ii*. It allows for calculating pixel-wise occlusion occurrence in the acquired images and employs a second optimizer to refine the initial trajectory in order to *i*. avoid rescanning of already known regions, and *ii*. explore the corrupted areas from angles that may minimize the occlusion.

8.2 Methodology

In this chapter, we present the two optimization techniques aimed at determining the optimal trajectory for covering a designated volume of interest, as well as computing a refined trajectory that minimizes the occurrence of acoustic shadowing during ultrasound scanning. We call these two optimizers *Volume Coverage Optimizer* and *Shadow Reduction Optimizer*

8.2.1 Volume Coverage Optimizer

The objective of this optimizer is to guarantee comprehensive coverage of the specified volume of interest. The probe's pose is represented by its position (x, y, z) and orientation $(\Omega_x, \Omega_y, \Omega_z)$. We define the space W of all possible probe poses that the robot can assume. The optimizer's task is to sort these poses in order of priority, with the output being a set of weights assigned to each pose such that higher weights are given to poses that provide better coverage of the target region of interest (ROI).

To formulate the optimization function, we first establish a voxel grid centered around the volume of interest. When the probe scans the volume, we say that a given probe pose **covers** a given voxel if the ultrasound plane at that pose intersects with the voxel. To account for reduced resolution away from the transducer's focal point, we adjust the coverage value by the voxel's distance from the probe's focal point. Given a probe pose $t(x, y, z, \Omega_x, \Omega_y, \Omega_z)$ and a voxel v in the defined grid, we can formulate the voxel coverage $d(t, v)$ between the voxel v and the pose t as follows:

$$\text{Voxel Coverage} = d(t, v) = \begin{cases} f(t, v), & \text{if } v \text{ is intersected by } t \\ 0, & \text{otherwise} \end{cases} \quad (8.1)$$

where f is a function that linearly decreases with increasing distances from the focal point.

The objective of the coverage optimizer is to determine a set of N probe poses that optimize the coverage of the specified volume of interest. An initial derivation of the optimization function is represented by the following equation:

$$E_{VC} = c_1 \cdot \exp\left(-\sum_{t,v} d(t, v) \cdot w(t)\right) \quad (8.2)$$

In this equation, the coverage of the volume is computed as the sum of the coverage over all voxels in the defined grid. However, this formulation has the drawback of being prone to falling into local minima, where the optimizer may select poses that only cover limited areas in repetitive manners. To mitigate this limitation, a regularization term is added to the energy function. This way, the optimizer not only maximizes the overall coverage but also ensures that the coverage is evenly distributed among the voxels.

$$E_{VC} = c_1 \cdot \exp\left(-\sum_{t,v} d(t,v) \cdot w(t)\right) + c_2 \cdot \sum_v \left(\sum_t (d(t,v) \cdot w(t) - \overline{cov_v})\right) \quad (8.3)$$

where $\overline{cov_v}$ is the average voxel coverage. We can see that the added term forces the variance of the voxel-wise coverage to be low, ensuring all voxels are equally covered.

Finally, to prevent the assigned weights from diverging to large values and ensure stability in the optimization process, we include an additional regularization term that minimizes the sum of the frame weights:

$$E_{VC} = c_1 \cdot \exp\left(-\sum_{t,v} d(t,v) \cdot w(t)\right) + c_2 \cdot \sum_v \left(\sum_t (d(t,v) \cdot w(t) - \overline{cov_v})\right) + c_3 \cdot \sum_t w(t) \quad (8.4)$$

The coefficient c_1 , c_2 , and c_3 determine the effect of each of the terms on the overall energy function and are selected empirically. The number of poses to be selected is chosen when a predefined desired average pose coverage is reached.

8.2.2 Shadow Reduction Optimizer

The objective of the shadow reduction optimizer is to determine an alternative trajectory that reduces the occurrence of voxels being occluded in comparison to a previous trajectory. To this end, we first need to determine which areas in the acquired volume are occluded, i.e. to classify occluded voxels. For this purpose, we utilize confidence maps [52]. Confidence maps are computed on B-mode ultrasound images and defined for a given pixel as the probability of a random walk reaching a transducer element starting from that pixel. We can therefore say that the confidence value at each pixel is correlated with the presence of shadowing artifacts, with low confidence values being likely associated with shadowed pixels and high confidence values with non-shadowed pixels. Since a given voxel v in the defined grid can be intersected by different ultrasound planes, thus by different bmode images, we classify v as non-occluded if its confidence value is above-threshold in at least one of the ultrasound images acquired during the first acquisition. After classifying each voxel in terms of occlusion, we formulate a new optimization function aimed at recalculating new pose weights in order to maximize the coverage of the occluded voxels, while ensuring that poses that have previously caused occlusions are not repeated, and new poses reaching occluded voxels are explored. To summarize, such energy function should ensure the following:

1. Poses that cover unseen or occluded voxels are prioritized
2. When trying to explore occluded voxels, new poses should be explored, possibly far from the ones affected by occlusions in the first acquisition.

In mathematical terms, we can formulate the energy function as:

$$E_{ASR} = E_{VC_{occ}} + c_4 \cdot \sum_t [w(t) \cdot g_{occ}(t)] + c_5 \cdot \sum_t \left(w(t) \cdot \sum_{i,j} \left[\sum_{v \in V} d(t, v) \cdot covered(v) \right] \right) \quad (8.5)$$

where $E_{VC_{occ}}$ is the volume coverage function as defined in section 9.2.1, computed considering occluded voxels only and $g_{occ}(t)$ is a multivariate gaussian function defined as:

$$g_{occ}(t) = \left[\prod_{t_o \in O} \mathcal{N}(t_o, \sigma^2) \right] (t_x, t_y) \quad (8.6)$$

forcing the optimizer to choose poses lying as far as possible from the one where voxel v results to be occluded. The last term is a regularizer penalizing poses that reacquire already seen voxels.

8.3 Experiments and Results

In order to assess the proposed methodology, we made the following assumptions. Firstly, the volume of interest was assumed to be square-shaped. Secondly, only 1 degree of freedom in rotation, specifically the tilting angle, was considered for simplicity. As for the translational component of the probe pose, only the x, y directions were taken into account, as the z direction is regulated by a force controller, as outlined in section 4.2.4.

To evaluate the proposed method, we performed extensive simulation experiments in a controlled environment. We further evaluate the potential of the application using a real robotic-ultrasound setup and a custom-designed phantom.

8.3.1 Simulation Experiments

To generate the simulated acquisition environment, we utilized the ImFusion software. For the simulated acquisitions, we defined four different phantoms, depicted in Figure 8.1

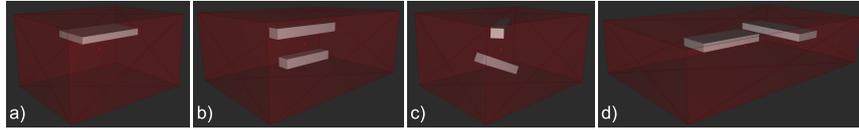


Fig. 8.1: The phantom utilized for the simulation experiments ©2022 IEEE

Each phantom was assigned highly reflective properties, comparable with one of the bone tissues in humans. We simulated two different media where the phantoms were inserted, the first simulating water-like material and the second soft-tissue-like material. For the simulation, we utilized a ray-tracing model, employing a generative model to simulate speckles properties. The method was implemented in the ImFusion suite. The acoustic impedance, speed of sound, and attenuation coefficient chosen for the phantom and media characterization are reported in Table 8.1.

Tab. 8.1: MATERIAL PROPERTIES OF SIMULATED TISSUES

	Acoustic Impedance (<i>MRayl</i>)	Speed of Sound (<i>m/s</i>)	Attenuation Coefficient (<i>Mhz/cm</i>)
Phantom	0.612	3600	7.9
Water-like medium	0.149	1492	0.019
Soft Tissue-like medium	0.163	1540	0.54

We used a linear probe with a 32.5 mm width, a central frequency of $2Mhz$, a depth of $10cm$, and a focal distance of $2.5cm$. For the optimizers, the volume of interest was divided into $8 \times 10 \times 6$ voxels, and the values of the coefficients c_1, c_2, c_3, c_4, c_5 were set empirically to 1, 1, 1, 2, 500. We measured the effect of the volume coverage optimizer in terms of covered pixels, where we used the definition of coverage defined in 8.1. To evaluate the performance of the shadowing optimizer, we measure the intensity level of the areas lying below the shadowing structures, as we expect that shadowed areas appear darker.

As a baseline, we employ parallel and random robot trajectories. The impact of the volume coverage optimizer on parallel and random scans is assessed, and the combination of the occlusion reduction module with parallel, random, and volume coverage optimized scans are evaluated. The results are presented in figure 8.2. The

quantitative results for the described methods and baseline are reported in Table 8.2.

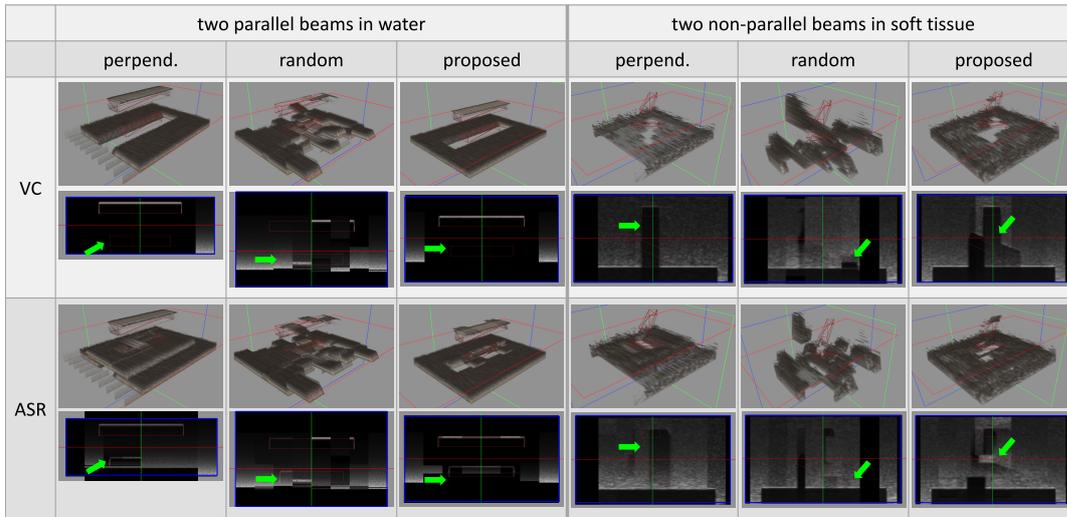


Fig. 8.2: The results of the compounding for the proposed and baseline methods ©2022 IEEE

Tab. 8.2: EVALUATION RESULTS. ASR IS SHADOW REDUCTION, VC IS VOLUME COVERAGE. RANDOM SCANS ARE AVERAGED OVER 10 RUNS. ©2022 IEEE

phantom	method	soft tissue					water		
		avg. coverage	avg. confidence	avg. intensity	combined intensity	avg. number of poses	avg. coverage	avg. confidence	avg. number of poses
Square	perpend.	0.921	0.442	21.520	21.323	80	0.750	0.352	80
	perpend. + ASR	0.958	0.460	22.187	28.021	138	0.942	0.441	191
	random	0.598	0.287	31.267	29.655	80	0.585	0.255	80
	random + ASR	0.754	0.362	25.007	29.014	162	0.741	0.332	167.7
Two plates	VC	0.983	0.472	9.963	9.712	448	0.954	0.454	448
	VC + ASR	0.992	0.476	5.280	10.695	468	0.954	0.454	505
	perpend.	0.894	0.429	10.033	9.020	80	0.833	0.400	80
	perpend. + ASR	0.975	0.468	14.831	20.750	106	0.900	0.432	174
Two beams, parallel	random	0.441	0.212	27.914	26.775	80	0.428	0.188	80
	random + ASR	0.514	0.247	16.465	23.841	113	0.525	0.232	129.5
	VC	0.973	0.467	10.794	10.518	448	0.883	0.424	448
	VC + ASR	0.981	0.471	6.067	11.460	474	0.900	0.432	609
Two beams, non-parallel	perpend.	0.915	0.439	4.825	3.822	80	0.850	0.404	80
	perpend. + ASR	0.975	0.468	8.757	14.664	116	0.917	0.436	166
	random	0.339	0.163	28.930	27.735	80	0.579	0.236	80
	random + ASR	0.499	0.240	22.354	28.722	138.4	0.740	0.300	146.6
Two beams, non-parallel	VC	0.992	0.476	8.268	7.266	448	0.950	0.456	448
	VC + ASR	0.992	0.476	3.688	8.941	460	0.965	0.460	499
	perpend.	0.890	0.427	15.517	14.466	80	0.840	0.390	80
	perpend. + ASR	0.904	0.434	13.021	20.924	211	0.917	0.436	166
Two beams, non-parallel	random	0.567	0.272	15.106	13.701	80	0.628	0.260	80
	random + ASR	0.757	0.363	12.760	16.058	195.7	0.773	0.325	155.3
	VC	0.910	0.437	12.935	11.827	448	0.950	0.440	448
	VC + ASR	0.910	0.437	10.849	16.484	577	0.970	0.450	490

The findings indicate that the implementation of the proposed shadow-reduction optimizer leads to an improvement in both the overall coverage and the pixel visibility, irrespective of the type of phantom employed. However, in some cases, the use of

this optimizer results in lower intensities below the phantom surface as compared to compounding the images acquired during the first scan alone. This is likely to occur when the optimizer encounters new occlusions for voxels that were successfully covered during the initial iteration. In such scenarios, the utilization of weighted compounding may alleviate this effect.

Moreover, the optimal outcome is obtained by combining the volume coverage optimizer with the shadow reduction optimizer, albeit at the cost of lengthier scanning times (more poses to be acquired). Therefore, for procedures where scanning time is a critical factor (such as acquisitions performed in breath-hold), it is recommended to employ parallel scanning in conjunction with the shadow-reduction optimizer, which produces marginally lower performance but faster scanning times.

8.3.2 Real Robot Experiment

To evaluate the potential of the proposed methodology in a more realistic environment, we conducted experiments using a real robot scanning a custom 3D-printed rectangular phantom placed in water. For this purpose, we used a KUKA LBR iiwa 14 R830 robot, with a Siemens Acuson ultrasound device equipped with a linear probe. Results of the robot experiments are reported in figure 8.3 and table 8.3 and show consistency with the simulation experiments.

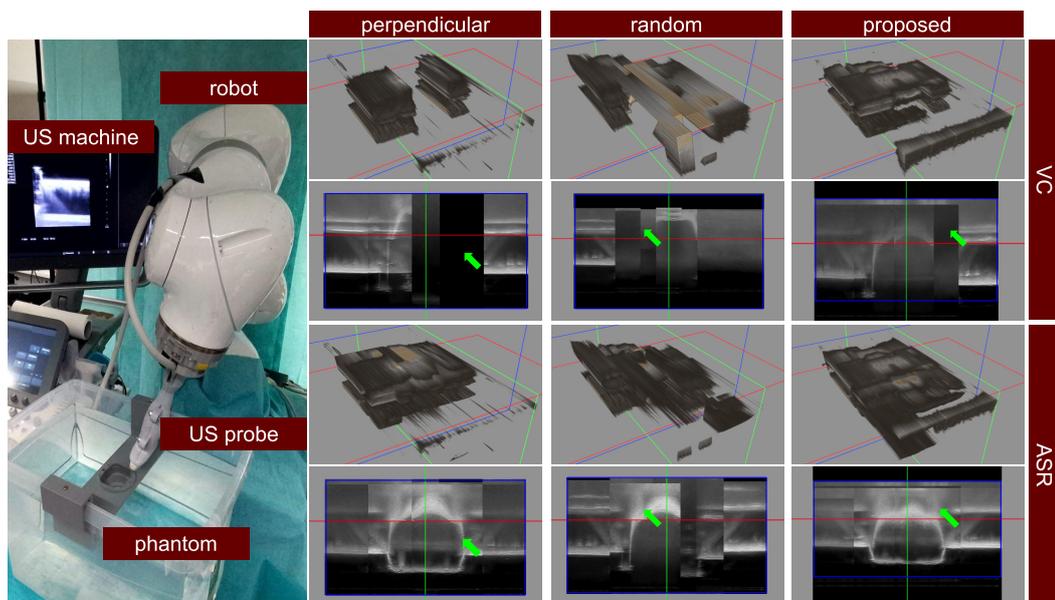


Fig. 8.3: Compounding results obtained with the real robot ©2022 IEEE

Tab. 8.3: RESULTS FOR THE PROPOSED OPTIMIZERS AND BASELINE OBTAINED WITH THE REAL ROBOT-ULTRASOUND SETUP ©2022 IEEE

method	avg. coverage	avg. confidence	number of poses	scanning time (s)
perpend.	0.425	0.052	40	65
perpend. + ASR	0.433	0.058	71	137
random	0.213	0.024	37	97
random + ASR	0.279	0.036	71	186
VC	0.4375	0.064	176	88
VC + ASR	0.438	0.064	211	175

8.4 Conclusion

In conclusion, this study presents novel optimization methods for robot acquisition optimization. The effectiveness of the proposed optimizer was demonstrated through both simulation experiments and real robot experiments. However, this work made certain assumptions, such as the flat patient surface assumption and the reduced degrees of freedom of the probe pose. Future research should focus on extending the proposed method to all degrees of freedom and more complex geometries of the working volumes. Moreover, conducting experiments on humans or animals would allow for the evaluation of the impact of motion artifacts on the effectiveness of the optimizers. Implementing the methods in real-time, such as modifying the probe pose in real time to cover unexplored or shadowed pixels, presents an exciting direction for exploration, which has the potential to mitigate motion artifacts in the acquisition.

Part III

Conclusions

Conclusion and Future Directions

In conclusion, this dissertation explores ways to design intelligent robotic ultrasound systems by leveraging multi-sensory information with anatomical and physical priors.

In the first proposed study, we developed a robotic-ultrasound method for vertebral level detection and counting for use in clinical spine injection procedures [91]. This system, which integrates visual and force feedback for vertebral level classification, demonstrated that fusing force and ultrasound data yields improved results compared to using either type of data alone. The use of a temporal convolutional network (TCN) was also evaluated and found to improve performance compared to a simple peak detector, suggesting that the TCN can effectively learn anatomical priors on vertebral level positions. The method was tested on a diverse group of healthy volunteers and showed promising results. Future works in this direction include further clinical studies to evaluate its accuracy in pathological patients. Further automation may also be explored, such as automating the initial probe placement and enabling spine tracking.

To further explore the challenges and possible solutions of training deep learning networks in presence of limited datasets, we further investigate the role and potential of data augmentation for ultrasound data. In our research, we focused on determining the appropriate use of data augmentation in the context of ultrasound images, which differ in their generation principles from photographic data. To this end, we conducted a second study in which we introduced and evaluated a novel augmentation method for B-mode ultrasound images on a dataset of spine ultrasound images for the task of vertebral level detection [115]. The results showed that the proposed augmentations improved model performance and generalization ability. However, further research is needed to investigate the robustness of the method in other US imaging tasks and to determine the optimal combination of augmentations.

In this research, we finally examined the impact of incorporating visual and anatomical information derived from anatomical and physical priors, as well as ultrasound visual feedback, on the optimization of robot trajectory generation for ultrasound scanning [121]. To this end, we proposed a novel method for improving the performance of robotic ultrasound scanning by implementing multiple optimization modules in a

chain and combining the resulting sweeps for the final compounding. This method included a volume coverage module designed to efficiently cover a desired volume of interest, as well as a shadow reduction optimization module that optimizes visibility in presence of acoustic shadowing. The shadow reduction module leverages the system's ability to identify pixels affected by acoustic shadowing in the image and adjust the robot trajectory accordingly.

The automation of medical robotics systems, has been an active area of research in recent years, due in part to the development of deep learning techniques that have been effective in automatically analyzing imaging and temporal data. However, there are still several challenges that need to be addressed in order to create truly intelligent systems. One of these challenges is the issue of data variability and limited data availability in medical datasets, particularly in the case of ultrasound data, which is influenced by a range of factors such as tissue attenuation, ultrasound carrier frequency, and artifacts. This makes it difficult to create a standardized dataset that can represent the full range of variability across patients, scanners, anatomy, and operators. To address this, researchers are exploring ways to effectively learn relevant features and generalize over new, unseen data, through techniques such as data augmentation.

Another challenge in the automation of medical robots is the integration of multiple sensory feedbacks into the robot's control strategies. There is an ongoing debate in the research community about whether it is best to try to mimic human behavior or to develop machine intelligence in a more autonomous way when it comes to building a robotic system that should take over tasks currently performed by humans. Some argue that machine models should be treated as "black boxes" that learn to process data in the most effective way, while others believe that certain behaviors should be enforced based on our existing understanding of the world, such as physical laws and properties.

In this dissertation, we focus on integrating anatomical and physical priors into the machine and deep learning models for robot-enabled task automation. This approach is not meant to diminish the potential of deep models to autonomously learn tasks, but rather reflects the belief that we still have a long way to go before achieving fully intelligent systems, and its feasibility is still under debate. In the meantime, we can use our understanding of physical and anatomical laws to augment the capabilities of machine models in analyzing data quickly and effectively.

Bibliography

- [1] F. W. Kremkau and K. Taylor, “Artifacts in ultrasound imaging.”, *Journal of ultrasound in medicine*, vol. 5, no. 4, pp. 227–237, 1986.
- [2] G. Finet, E. Maurincomme, A. Tabib, *et al.*, “Artifacts in intravascular ultrasound imaging: Analyses and implications”, *Ultrasound in medicine & biology*, vol. 19, no. 7, pp. 533–547, 1993.
- [3] R. Rohling, A. Gee, and L. Berman, “Three-dimensional spatial compounding of ultrasound images”, *Medical Image Analysis*, vol. 1, no. 3, pp. 177–193, 1997.
- [4] R. M. Comeau, A. Fenster, and T. M. Peters, “Integrated mr and ultrasound imaging for improved image guidance in neurosurgery”, in *Medical Imaging 1998: Image Processing*, SPIE, vol. 3338, 1998, pp. 747–754.
- [5] R. N. Rohling, A. H. Gee, and L. Berman, “Automatic registration of 3-d ultrasound images”, *Ultrasound in medicine & biology*, vol. 24, no. 6, pp. 841–854, 1998.
- [6] S. E. Salcudean, G. Bell, S. Bachmann, W.-H. Zhu, P. Abolmaesumi, and P. D. Lawrence, “Robot-assisted diagnostic ultrasound—design and feasibility experiments”, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 1999, pp. 1062–1071.
- [7] M. Y. Stitz and H. M. Sommer, “Accuracy of blind versus fluoroscopically guided caudal epidural injection”, *Spine*, vol. 24, no. 13, p. 1371, 1999.
- [8] S. E. Salcudean, W. H. Zhu, P. Abolmaesumi, S. Bachmann, and P. D. Lawrence, “A robot system for medical ultrasound”, in *Robotics Research*, J. M. Hollerbach and D. E. Koditschek, Eds., London: Springer London, 2000, pp. 195–202, ISBN: 978-1-4471-0765-1.
- [9] W. Zhu, S. Salcudean, S. Bachmann, and P. Abolmaesumi, “Motion/force/image control of a diagnostic ultrasound robot”, in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, vol. 2, 2000, 1580–1585 vol.2. DOI: 10.1109/ROBOT.2000.844822.

- [10] P. Abolmaesumi, S. Salcudean, W. Zhu, S. DiMaio, and M. Sirouspour, “A user interface for robot-assisted diagnostic ultrasound”, in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, vol. 2, 2001, 1549–1554 vol.2. DOI: 10.1109/ROBOT.2001.932831.
- [11] P. Abolmaesumi, S. Salcudean, W.-H. Zhu, M. Sirouspour, and S. DiMaio, “Image-guided control of a robot for medical ultrasound”, *IEEE Transactions on Robotics and Automation*, vol. 18, no. 1, pp. 11–23, 2002. DOI: 10.1109/70.988970.
- [12] M. Greher Manfred, M. Scharbert Gisela, M. Kamolz Lars P., *et al.*, “Ultrasound-guided lumbar facet nerve block: A sonoanatomic study of a new methodologic approach”, *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 100, no. 5, pp. 1242–1248, May 2004, ISSN: 0003-3022. eprint: jasa/content_public/journal/jasa/931196/0000542-200405000-00028.pdf.
- [13] J. Hanc, E. F. Taylor, and S. Tuleja, “Deriving lagrange’s equations using elementary calculus”, *American Journal of Physics*, vol. 72, no. 4, pp. 510–513, 2004.
- [14] P. Marhofer, M. Greher, and S. Kapral, “Ultrasound guidance in regional anaesthesia”, *British journal of anaesthesia*, vol. 94, no. 1, pp. 7–17, 2005.
- [15] W. A. Berg, J. D. Blume, J. B. Cormack, and E. B. Mendelson, “Operator dependence of physician-performed whole-breast us: Lesion detection and characterization”, *Radiology*, vol. 241, no. 2, pp. 355–365, 2006.
- [16] “Fundamentals of piezoelectricity”, in *Piezoelectric Transducers for Vibration Control and Damping*. London: Springer London, 2006, pp. 9–35, ISBN: 978-1-84628-332-1. DOI: 10.1007/1-84628-332-9_2.
- [17] P. Soler, G. Delso, N. Villain, E. Angelini, and I. Bloch, “Superresolution spatial compounding techniques with application to 3d breast ultrasound imaging”, in *Medical Imaging 2006: Ultrasonic Imaging and Signal Processing*, SPIE, vol. 6147, 2006, pp. 281–292.
- [18] R. H. Taylor, “A perspective on medical robotics”, *Proceedings of the IEEE*, vol. 94, no. 9, pp. 1652–1664, 2006.
- [19] O. V. Solberg, F. Lindseth, H. Torp, R. E. Blake, and T. A. N. Hernes, “Freehand 3d ultrasound reconstruction algorithms—a review”, *Ultrasound in medicine & biology*, vol. 33, no. 7, pp. 991–1009, 2007.
- [20] J. C. A. Carvalho, “Ultrasound-facilitated epidurals and spinals in obstetrics”, *Anesthesiology clinics*, vol. 26, no. 1, pp. 145–158, 2008.
- [21] M.-A. Janvier, L.-G. Durand, M.-H. R. Cardinal, *et al.*, “Performance evaluation of a medical robotic 3d-ultrasound imaging system”, *Medical image analysis*, vol. 12, no. 3, pp. 275–290, 2008.

- [22] B. Kerby, R. Rohling, V. Nair, and P. Abolmaesumi, “Automatic identification of lumbar level with ultrasound”, *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 08 - Personalized Healthcare through Technology*, pp. 2980–2983, 2008. DOI: 10.1109/iembs.2008.4649829.
- [23] R. Mebarki, A. Krupa, and F. Chaumette, “Image moments-based ultrasound visual servoing”, in *2008 IEEE International Conference on Robotics and Automation*, 2008, pp. 113–119. DOI: 10.1109/ROBOT.2008.4543195.
- [24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database”, in *CVPR09*, 2009.
- [25] A. Krupa, G. Fichtinger, and G. D. Hager, “Real-time motion stabilization with b-mode ultrasound using image speckle information and visual servoing”, *The International Journal of Robotics Research*, vol. 28, no. 10, pp. 1334–1354, 2009.
- [26] J. C. Rippey and A. G. Royse, “Ultrasound in trauma”, *Best Practice & Research Clinical Anaesthesiology*, vol. 23, no. 3, pp. 343–362, 2009.
- [27] B. Ihnatsenka and A. P. Boezaart, “Ultrasound: Basic understanding and learning the language”, *International journal of shoulder surgery*, vol. 4, no. 3, p. 55, 2010.
- [28] R. Nakadate, J. Solis, A. Takanishi, E. Minagawa, M. Sugawara, and K. Niki, “Out-of-plane visual servoing method for tracking the carotid artery with a robot-assisted ultrasound diagnostic system”, in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 5267–5272. DOI: 10.1109/ICRA.2011.5979594.
- [29] N. Hungr, M. Baumann, J.-A. Long, and J. Troccaz, “A 3-d ultrasound robotic prostate brachytherapy system with prostate motion tracking”, *IEEE Transactions on Robotics*, vol. 28, no. 6, pp. 1382–1397, 2012.
- [30] K. Ertan, C. Linsler, A. Di Liberto, M. F. Ong, E. Solomayer, and J. Endrikat, “Axillary ultrasound for breast cancer staging: An attempt to identify clinical/histopathological factors impacting diagnostic performance”, *Breast Cancer: Basic and Clinical Research*, vol. 7, BCBCR–S11215, 2013.
- [31] F. Lindseth, T. Langø, T. Selbekk, *et al.*, “Ultrasound-based guidance and therapy”, in *Advancements and breakthroughs in ultrasound imaging*, IntechOpen, 2013.
- [32] A. S. B. Mustafa, T. Ishii, Y. Matsunaga, *et al.*, “Development of robotic system for autonomous liver screening using ultrasound scanning device”, in *2013 IEEE international conference on robotics and biomimetics (ROBIO)*, IEEE, 2013, pp. 804–809.

- [33] C. Nadeau and A. Krupa, “Intensity-based ultrasound visual servoing: Modeling and validation with 2-d and 3-d probes”, *IEEE Transactions on Robotics*, vol. 29, no. 4, pp. 1003–1015, 2013. DOI: 10.1109/TRO.2013.2256690.
- [34] A. M. Priester, S. Natarajan, and M. O. Culjat, “Robotic ultrasound systems in medicine”, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 60, no. 3, pp. 507–523, 2013.
- [35] I. Kuhlemann, R. Bruder, F. Ernst, and A. Schweikard, “We-g-brf-09: Force-and image-adaptive strategies for robotised placement of 4d ultrasound probes”, *Medical Physics*, vol. 41, no. 6Part30, pp. 523–523, 2014.
- [36] D. Nicholls, L. Sweet, and J. Hyett, “Psychomotor skills in medical ultrasound imaging: An analysis of the core skill set”, *Journal of Ultrasound in Medicine*, vol. 33, no. 8, pp. 1349–1352, 2014.
- [37] P. Chatelain, A. Krupa, and N. Navab, “Optimization of ultrasound image quality via visual servoing”, in *2015 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2015, pp. 5997–6002.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. arXiv: 1512.03385 [cs.CV].
- [39] S. Merouche, L. Allard, E. Montagnon, G. Soulez, P. Bigras, and G. Cloutier, “A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of b-mode images”, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 1, pp. 35–46, 2015.
- [40] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, in *International Conference on Learning Representations*, 2015.
- [41] S. Yu, K. K. Tan, B. L. Sng, S. Li, and A. T. H. Sia, “Real-time automatic spinal level identification with ultrasound image processing”, in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2015, pp. 243–246. DOI: 10.1109/ISBI.2015.7163859.
- [42] L.-A. Duflot, A. Krupa, B. Tamadazte, and N. Andreff, “Towards ultrasound-based visual servoing using shearlet coefficients”, in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 3420–3425. DOI: 10.1109/ICRA.2016.7487519.
- [43] C. Graumann, B. Fuerst, C. Hennersperger, F. Bork, and N. Navab, “Robotic ultrasound trajectory planning for volume of interest coverage”, in *2016 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2016, pp. 736–741.
- [44] C. Hennersperger, B. Fuerst, S. Virga, *et al.*, “Towards mri-based autonomous robotic us acquisitions: A first feasibility study”, *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 538–548, 2016.

- [45] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, *Densely connected convolutional networks*, 2016. arXiv: 1608.06993 [cs.CV].
- [46] C. Nadeau, A. Krupa, J. Petr, and C. Barillot, “Moments-based ultrasound visual servoing: From a mono- to multiplane approach”, *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1558–1564, 2016. DOI: 10.1109/TRO.2016.2604482.
- [47] P. A. Patlan-Rosales and A. Krupa, “Automatic palpation for quantitative ultrasound elastography by visual servoing and force control”, in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 2357–2362. DOI: 10.1109/IROS.2016.7759367.
- [48] S. Virga, O. Zettinig, M. Esposito, *et al.*, “Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms”, in *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 2016, pp. 508–513.
- [49] S. Wang, D. Singh, D. Johnson, K. Althoefer, K. Rhode, and R. J. Housden, “Robotic ultrasound: View planning, tracking, and automatic acquisition of transesophageal echocardiography”, *IEEE Robotics & Automation Magazine*, vol. 23, no. 4, pp. 118–127, 2016. DOI: 10.1109/MRA.2016.2580478.
- [50] O. Zettinig, B. Fuerst, R. Kojcev, *et al.*, “Toward real-time 3D ultrasound registration-based visual servoing for interventional navigation”, pp. 945–950, 2016.
- [51] N. Baka, S. Leenstra, and T. van Walsum, “Ultrasound aided vertebral level localization for lumbar surgery”, *IEEE transactions on medical imaging*, vol. 36, no. 10, pp. 2138–2147, 2017.
- [52] P. Chatelain, A. Krupa, and N. Navab, “Confidence-driven control of an ultrasound probe”, *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1410–1424, 2017.
- [53] R. Elek, T. D. Nagy, D. A. Nagy, *et al.*, “Robotic platforms for ultrasound diagnostics and treatment”, in *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, IEEE, 2017, pp. 1752–1757.
- [54] S. Gerlach, I. Kuhlemann, P. Jauer, *et al.*, “Robotic ultrasound-guided sbrr of the prostate: Feasibility with respect to plan quality”, *International journal of computer assisted radiology and surgery*, vol. 12, no. 1, pp. 149–159, 2017.
- [55] J. Hetherington, V. Lessoway, V. Gunka, P. Abolmaesumi, and R. Rohling, “SLIDE: automatic spine level identification system using a deep convolutional neural network”, in *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 7, pp. 1189–1198, Jul. 2017, ISSN: 1861-6410, 1861-6429.

- [56] R. Kojcev, A. Khakzar, B. Fuerst, *et al.*, “On the reproducibility of expert-operated and robotic ultrasound acquisitions”, *International journal of computer assisted radiology and surgery*, vol. 12, no. 6, pp. 1003–1011, 2017.
- [57] P. A. Patlan-Rosales and A. Krupa, “Strain estimation of moving tissue based on automatic motion compensation by ultrasound visual servoing”, in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 2941–2946. DOI: 10.1109/IROS.2017.8206128.
- [58] L. Schwentner, G. Helms, V. Nekljudova, *et al.*, “Using ultrasound and palpation for predicting axillary lymph node status following neoadjuvant chemotherapy—results from the multi-center sentina trial”, *The Breast*, vol. 31, pp. 202–207, 2017.
- [59] G.-Z. Yang, J. Cambias, K. Cleary, *et al.*, *Medical robotics-regulatory, ethical, and legal considerations for increasing levels of autonomy*, 2017.
- [60] Y. Zhu, Z. Fu, and J. Fei, “An image augmentation method using convolutional network for thyroid nodule classification by transfer learning”, in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, IEEE, 2017, pp. 1819–1823.
- [61] A. Denasi, F. Khan, K. J. Boskma, *et al.*, “An observer-based fusion method using multicore optical shape sensors and ultrasound images for magnetically-actuated catheters”, in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018, pp. 50–57.
- [62] J. Esteban, W. Simson, S. R. Witzig, *et al.*, “Robotic ultrasound-guided facet joint insertion”, *International journal of computer assisted radiology and surgery*, vol. 13, no. 6, pp. 895–904, 2018.
- [63] Q. Huang, J. Lan, and X. Li, “Robotic arm based automatic ultrasound scanning for three-dimensional imaging”, *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1173–1182, 2018.
- [64] X. Qi, N. Vora, L. Riera, *et al.*, “Automatic scan plane identification from 2d ultrasound for pedicle screw guidance”, *CAOS*, vol. 2, pp. 168–174, 2018.
- [65] S. Virga, R. Göbl, M. Baust, N. Navab, and C. Hennersperger, “Use the force: Deformation correction in robotic 3d ultrasound”, *International journal of computer assisted radiology and surgery*, vol. 13, no. 5, pp. 619–627, 2018.
- [66] C. E. Alexander and M. Varacallo, “Lumbosacral facet syndrome”, in *StatPearls [Internet]*, StatPearls Publishing, 2019.
- [67] A. Alsinan, M. Vives, V. Patel, and I. Hacihaliloglu, “Spine surface segmentation from ultrasound using multi-feature guided cnn”, *CAOS*, vol. 3, pp. 6–10, 2019.

- [68] M. Antico, F. Sasazawa, L. Wu, *et al.*, “Ultrasound guidance in minimally invasive robotic procedures”, *Medical image analysis*, vol. 54, pp. 149–167, 2019.
- [69] J. Carriere, J. Fong, T. Meyer, *et al.*, “An admittance-controlled robotic assistant for semi-autonomous breast ultrasound scanning”, in *2019 international symposium on medical robotics (ISMR)*, IEEE, 2019, pp. 1–7.
- [70] W. Al-Dhabyani, M. Gomaa, H. Khaled, and F. Aly, “Deep learning approaches for data augmentation and classification of breast masses using ultrasound images”, *Int. J. Adv. Comput. Sci. Appl*, vol. 10, no. 5, pp. 1–11, 2019.
- [71] D. Q. Duong, K.-C. T. Nguyen, N. R. Kaipatur, *et al.*, “Fully automated segmentation of alveolar bone using deep convolutional neural networks from intraoral ultrasound images”, in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 6632–6635.
- [72] Y. A. Farha and J. Gall, “Ms-tcn: Multi-stage temporal convolutional network for action segmentation”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3575–3584.
- [73] F. Langsch, S. Virga, J. Esteban, R. Göbl, and N. Navab, “Robotic ultrasound for catheter navigation in endovascular procedures”, in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2019, pp. 5404–5410.
- [74] S. Liu, Y. Wang, X. Yang, *et al.*, “Deep learning in medical ultrasound analysis: A review”, *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.
- [75] A. Østvik, L. E. Bø, and E. Smistad, “Echobot: An open-source robotic ultrasound system”, *Proc. IPCAI*, pp. 1–4, 2019.
- [76] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning”, *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [77] I. M. Skaribas, J. L. Erian, D. Reynolds, and E. E. Skaribas, “Lumbar interlaminar epidural injection”, in *Deer’s Treatment of Pain*, Springer, 2019, pp. 405–412.
- [78] H. Tanaka, S.-W. Chiu, T. Watanabe, S. Kaoku, and T. Yamaguchi, “Computer-aided diagnosis system for breast ultrasound images using deep learning”, *Physics in Medicine & Biology*, vol. 64, no. 23, p. 235 013, 2019.
- [79] A. Z. Alsinan, C. Rule, M. Vives, V. M. Patel, and I. Hacihaliloglu, “Gan-based realistic bone ultrasound image and label synthesis for improved segmentation”, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 795–804.

- [80] B. Benjdira, K. Ouni, M. M. Al Rahhal, A. Albakr, A. Al-Habib, and E. Mahrous, “Spinal cord segmentation in ultrasound medical imagery”, *Applied Sciences*, vol. 10, no. 4, p. 1370, 2020.
- [81] O. Boubaker, “Medical robotics”, *Control Theory in Biomedical Engineering*, pp. 153–204, 2020.
- [82] H. Hase, M. F. Azampour, M. Tirindelli, *et al.*, “Ultrasound-guided robotic navigation with deep reinforcement learning”, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2020, Las Vegas, NV, USA, October 24, 2020 - January 24, 2021*, IEEE, 2020, pp. 5534–5541.
- [83] B. Hohlmann, J. Glanz, and K. Radermacher, “Segmentation of the distal femur in ultrasound images”, *Current Directions in Biomedical Engineering*, vol. 6, no. 1, 2020.
- [84] P. Huang, L. Lin, K. Xu, and H. Huang, “Autonomous outdoor scanning via on-line topological and geometric path optimization”, *ArXiv*, vol. abs/2012.12642, 2020.
- [85] Z. Jiang, M. Grimm, M. Zhou, *et al.*, “Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement”, *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1342–1349, 2020.
- [86] Z. Jiang, M. Grimm, M. Zhou, *et al.*, “Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement”, *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1342–1349, 2020.
- [87] K. Luan, Z. Li, and J. Li, “An efficient end-to-end cnn for segmentation of bone surfaces from ultrasound”, *Computerized Medical Imaging and Graphics*, vol. 84, p. 101 766, 2020.
- [88] M. Mischi, M. A. L. Bell, R. J. Van Sloun, and Y. C. Eldar, “Deep learning in medical ultrasound—from image formation to image analysis”, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2477–2480, 2020.
- [89] K.-C. T. Nguyen, D. Q. Duong, F. T. Almeida, *et al.*, “Alveolar bone segmentation in intraoral ultrasonographs with machine learning”, *Journal of Dental Research*, vol. 99, no. 9, pp. 1054–1061, 2020.
- [90] H. Patel and I. Hacihaliloglu, “Improved automatic bone segmentation using large-scale simulated ultrasound data to segment real ultrasound bone surface data”, in *2020 IEEE 20th International Conference on Bioinformatics and Biengineering (BIBE)*, IEEE, 2020, pp. 288–294.

- [91] M. Tirindelli, M. Victorova, J. Esteban, *et al.*, “Force-ultrasound fusion: Bringing spine robotic-us to the next “level””, *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5661–5668, 2020.
- [92] T. Ungi, H. Greer, K. R. Sunderland, *et al.*, “Automatic spine ultrasound segmentation for scoliosis visualization and measurement”, *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3234–3241, 2020.
- [93] Z. Wang, “Deep learning in medical ultrasound image segmentation: A review”, *arXiv preprint arXiv:2002.07703*, 2020.
- [94] M. Welleweerd, A. de Groot, S. de Looijer, F. Siepel, and S. Stramigioli, “Automated robotic breast ultrasound acquisition using ultrasound feedback”, in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 9946–9952. DOI: 10.1109/ICRA40945.2020.9196736.
- [95] A. Zaman, S. H. Park, H. Bang, C.-w. Park, I. Park, and S. Joung, “Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images”, *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 6, pp. 931–941, 2020.
- [96] Y. Zhang, Z. Tian, Y. Lei, *et al.*, “Automatic multi-needle localization in ultrasound images using large margin mask rcnn for ultrasound-guided prostate brachytherapy”, *Physics in Medicine & Biology*, vol. 65, no. 20, p. 205 003, 2020.
- [97] A. P. Adedigba and S. A. Adeshina, “Deep learning-based classification of covid-19 lung ultrasound for tele-operative robot-assisted diagnosis”, in *2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS)*, IEEE, 2021, pp. 1–6.
- [98] M. Akbari, J. Carriere, T. Meyer, *et al.*, “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*, vol. 8, p. 645 424, 2021.
- [99] P. E. Dupont, B. J. Nelson, M. Goldfarb, *et al.*, “A decade retrospective of medical robotics research from 2010 to 2020”, *Science Robotics*, vol. 6, no. 60, eabi8017, 2021.
- [100] F. von Haxthausen, S. Böttger, D. Wulff, J. Hagenah, V. García-Vázquez, and S. Ipsen, “Medical robotics for ultrasound imaging: Current systems and future trends”, *Current Robotics Reports*, vol. 2, no. 1, pp. 55–71, 2021.
- [101] A. L. Y. Hung and J. Galeotti, “Good and bad boundaries in ultrasound compounding: Preserving anatomic boundaries while suppressing artifacts”, *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 11, pp. 1957–1968, 2021.

- [102] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab, “Automatic force-based probe positioning for precise robotic ultrasound acquisition”, *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 11 200–11 211, 2021. DOI: 10.1109/TIE.2020.3036215.
- [103] Z. Jiang, H. Wang, Z. Li, *et al.*, “Motion-aware robotic 3d ultrasound”, in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 12 494–12 500.
- [104] Z. Jiang, Y. Zhou, Y. Bi, M. Zhou, T. Wendler, and N. Navab, “Deformation-aware robotic 3d ultrasound”, *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7675–7682, 2021. DOI: 10.1109/LRA.2021.3099080.
- [105] M. Le Lous, F. Despinoy, M. Klein, E. Fustec, V. Lavoué, and P. Jannin, “Impact of physician expertise on probe trajectory during obstetric ultrasound: A quantitative approach for skill assessment”, *Simulation in Healthcare*, vol. 16, no. 1, pp. 67–72, 2021.
- [106] L. H. Lee, Y. Gao, and J. A. Noble, “Principled ultrasound data augmentation for classification of standard planes”, in *International Conference on Information Processing in Medical Imaging*, Springer, 2021, pp. 729–741.
- [107] K. Li, J. Wang, Y. Xu, *et al.*, “Autonomous navigation of an ultrasound probe towards standard scan planes with deep reinforcement learning”, in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 8302–8308.
- [108] K. Li, Y. Xu, and M. Q.-H. Meng, “An overview of systems and techniques for autonomous robotic ultrasound acquisitions”, *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 2, pp. 510–524, 2021. DOI: 10.1109/TMRB.2021.3072190.
- [109] K. Li, Y. Xu, J. Wang, D. Ni, L. Liu, and M. Q.-H. Meng, “Image-guided navigation of a robotic ultrasound probe for autonomous spinal sonography using a shadow-aware dual-agent framework”, *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 1, pp. 130–144, 2021.
- [110] J. Liang and J. Chen, “Data augmentation of thyroid ultrasound images using generative adversarial network”, in *2021 IEEE International Ultrasonics Symposium (IUS)*, IEEE, 2021, pp. 1–4.
- [111] Y.-T. Liao, C.-H. Lee, K.-S. Chen, C.-P. Chen, and T.-W. Pai, “Data augmentation based on generative adversarial networks to improve stage classification of chronic kidney disease”, *Applied Sciences*, vol. 12, no. 1, p. 352, 2021.
- [112] G. Ning, X. Zhang, and H. Liao, “Autonomic robotic ultrasound imaging system based on reinforcement learning”, *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 9, pp. 2787–2797, 2021. DOI: 10.1109/TBME.2021.3054413.

- [113] S. H. Park, “Artificial intelligence for ultrasonography: Unique opportunities and challenges”, *Ultrasonography*, vol. 40, no. 1, p. 3, 2021.
- [114] F. Suligoj, C. M. Heunis, J. Sikorski, and S. Misra, “Robust-an autonomous robotic ultrasound system for medical imaging”, *IEEE Access*, vol. 9, pp. 67 456–67 465, 2021. DOI: 10.1109/ACCESS.2021.3077037.
- [115] M. Tirindelli, C. Eilers, W. Simson, M. Paschali, M. F. Azampour, and N. Navab, “Rethinking ultrasound augmentation: A physics-inspired approach”, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 690–700.
- [116] Y. Bi, Z. Jiang, Y. Gao, T. Wendler, A. Karlas, and N. Navab, “Vesnet-rl: Simulation-based reinforcement learning for real-world us probe navigation”, *IEEE Robotics and Automation Letters*, 2022.
- [117] S. R. Brandigampala, A. F. Al-Battal, and T. Q. Nguyen, “Data augmentation methods for object detection and segmentation in ultrasound scans: An empirical comparative study”, in *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2022, pp. 288–291.
- [118] P.-C. Li and P.-H. Tsui, *Advances in ultrasound imaging for diagnostic and therapeutic purposes*, 2022.
- [119] M. Ronchetti, J. Rackerseder, M. Tirindelli, *et al.*, “Pro-tip: Phantom for robust automatic ultrasound calibration by tip detection”, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 84–93.
- [120] Y. Song, Z. Zhong, B. Zhao, *et al.*, “Medical ultrasound image quality assessment for autonomous robotic screening”, *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6290–6296, 2022.
- [121] V. Sutedjo, M. Tirindelli, C. Eilers, W. Simson, B. Busam, and N. Navab, “Acoustic shadowing aware robotic ultrasound: Lighting up the dark”, *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1808–1815, 2022.
- [122] M. Victorova, H. H. T. Lau, T. T.-Y. Lee, D. Navarro-Alarcon, and Y. Zheng, “Comparison of ultrasound scanning for scoliosis assessment: Robotic versus manual”, *The International Journal of Medical Robotics and Computer Assisted Surgery*, e2468, 2022.
- [123] J. Zielke, C. Eilers, B. Busam, W. Weber, N. Navab, and T. Wendler, “Rsv: Robotic sonography for thyroid volumetry”, *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3342–3348, 2022.

List of Figures

4.1	Behavior of a robot under stiffness control, where the robot behaves as if a spring with stiffness K_m were attached to its end-effector	22
4.2	Behavior of a robot under impedance control, where the robot behaves as if a spring with stiffness K_m and a damper with damping coefficient D_{m-m} were attached to its end-effector	22
4.3	Behavior of a robot under hybrid impedance -force control, where impedance control is used along the horizontal axis while force control is used along the vertical axis.	24
4.4	Schematic representation of the incorporation of visual servoing into the robot control scheme.	25
5.1	Top: The areas of air compression and rarefaction produced by a sound wave. Bottom: The sinusoidal representation of the acoustic wave, where p_0 is the wave amplitude and λ is the wavelength	27
5.2	The electrical field generated by the presence of a potential difference ΔV . Negative charges (red circles) cannot travel towards positive charges (green circles) due to the presence of insulating material in between them.	28
5.3	(left) a non-polar molecule (right) a polar molecule. It can be appreciated how in the non-polar molecule the negative and positive charge center coincides, due the molecule geometry. The elongated geometry in the non-polar molecule causes the centers of positive and negative charges to differ, causing a non-null dipole moment.	29
5.4	A non-polar dielectric in absence of an electric field. (right) A non-polar dipole undergoing polarization due to the presence of an electric field. It can be noticed how the presence of an electric field causes a rearrangement of the charges within the molecule, thus the split of positive and negative charge centers and consequently the generation of a non-null dipole moment.	30
5.5	Molecules in a dielectric cannot freely move in the solid. However, their structure allows for minor motions as changes in their orientation align to an applied electrical field. This figure shows the example of a polar dielectric, which undergoes polarization due to the re-orientation of its electrical dipoles.	30

5.6	Polarization occurring in a polar piezoelectric material due to the application of a mechanical strain ϵ . The polarization leads to the generation of a net electric field within the dipole, thus of a voltage difference ΔV which can be measured at its side	31
5.7	Polarization occurring in a polar piezoelectric material due to the application of a mechanical strain ϵ . The polarization leads to the generation of a net electric field within the dipole, thus of a voltage difference ΔV which can be measured at its side	31
5.8	The generation of an ultrasound wave and the generation of an echo from a point P reached by the ultrasound wave	32
5.9	Left: generation and recording of simultaneous echoes generated from point masses located at equal distances from the ultrasound source. Right: generation and recording of echoes undergoing destructive interference	33
5.10	The transmit delay pattern to achieve focusing (i.e. constructive interference) at a focus point	33
5.11	The delayed pattern applied to the transducer elements to synchronize them to the reception of the echoes coming from point P	34
5.12	The integration between an ultrasound beam with a surface. Part of the beam (blue arrow) propagates in the surface with refraction angles α , and part (red arrow) is reflected back with a reflection angle β equal to the angle between the incoming beam (black arrow) and the normal to the surface (dashed line)	35
5.13	Illustration of the mirroring effect in ultrasound imaging, where the original ultrasound ray is reflected off a surface and then bounces off a second structure before returning to the transducer, resulting in a mirrored image of the second reflective object.	37
5.14	Representation of multiple reflections against the ultrasound transducer, showing the reflection of the echo from the blue structure back to the transducer and then back to the structure, resulting in the creation of multiple echoes at different timestamps	38
5.15	Illustration of the acoustic shadow created by a highly reflective structure (black rectangle) partially obscuring structures beneath it	39
5.16	Image showing the deformation of a structure caused by the varying propagation speed of the ultrasound wave along different scanlines	39
5.17	Schematic representation of the voxel assignment in the case of voxel-based methods for volume compounding	41
5.18	Schematic representation of the voxel assignment in the case of pixel-based methods for volume compounding	42
6.1	The lumbar vertebrae in the human spine	48
6.2	The probe orientation with respective ultrasound views ©2020 IEEE	49
6.3	The system setup ©2020 IEEE	50

6.4	Vertebra Classification as it was done for Dataset I	51
6.5	Vertebral Bone Labelling in Dataset III	52
6.6	(a) The modeled interaction between robot and patient back during the robotic scanning procedure. (b) Z component (red) and Y component (blue) of the force signal recorded over a single vertebra. ©2020 IEEE	53
6.7	The pre and post-processed force signal.	53
6.8	The force signal recorded in the y-axis with 3 different values (2, 10 and 15 N) of the z-force applied by the robot for subjects with $BMI < 23$ (left) and for subjects with $BMI > 23$ (right). ©2020 IEEE	54
6.9	(a) Force signal, Ultrasound signal, and labels in the presence of non-corrupted force and ultrasound data. (b) Force signal, Ultrasound signal, and labels in the presence of noisy ultrasound data. (c) Force signal, Ultrasound signal, and labels in presence of noisy force signal. ©2020 IEEE	56
6.10	The architecture of the single stage of the 1D convolutional network for data fusion. ©2020 IEEE	57
6.11	The predicted (red line) and ground-truth (black line) vertebral levels for pure force-based, pure ultrasound-based, and force-ultrasound fusion both when using a peak detector and a tcn. Subject 1 with anatomical characteristics non-well-represented in the tcn training set (Subject Gender: Male, BMI: 30, Height: 186 cm). Subject 2 with anatomical characteristics well-represented in the tcn training set (Subject Gender: Female, BMI: 22, Height: 172 cm). ©2020 IEEE	60
6.12	The result of the facet detection and facet enhancement for one sample test sweep	62
7.1	The proposed physics-inspired data augmentation framework: using ultrasound scans and their corresponding bone annotations (yellow), we generate augmentations leveraging the physical properties of ultrasound scanning (green) and can also merge multiple augmentations (blue).	66
7.2	Examples of applying classical augmentation approaches to ultrasound scans: a) horizontal flipping, b) vertical flipping, c) rotation, d) translation, e) scaling, f) shearing, g) brightness adjustment.	66

7.3	Augmentation by Deformation Pipeline: Left: The image displays relevant forces and coordinate systems. The force applied by the transducer is represented with red arrows, and the feedback force from the bone is with green arrows. The coordinate system placed on the bone is represented by (x_b, y_b) and the transducer centroid coordinate space by (x_p, y_p) . Right: The B-mode ultrasound scan and its corresponding bone annotation. The bone annotation is used in the pipeline for computing the deformation field (Σ) induced by the transducer force. The deformation field is displayed in the transducer coordinate system to avoid generating a gap on top of the ultrasound scan. We apply a gaussian smoothing on (Σ^t) to account for and approximate the lateral forces. The resulting deformation field is then applied to the original scan and the annotation.	68
7.4	Augmentation by Reverberation Pipeline. Left: A schematic representation of the transmitted wave echoing on the bone at depth Δ and time t_1 and the reverberation echo happening at time t_2 . In the following time-amplitude plot, the signals arising from the first echo t_1 and the reverberation echo t_2 are represented. The echo effect at depth Δ and the reverberation effect at depth 2Δ are shown on the following B-mode image. Right: Synthetic generation of reverberation artifact on B-mode scans for the purpose of data augmentation. The bone is extracted from the original ultrasound scan and repeated at the reverberation depth and is finally added to the original B-mode scan.	70
7.5	Augmentation by SNR modification pipeline: First, we compute the local energy (LE) map using the input US scan. We then multiply the computed LE map with the bone intensity and background intensity values and apply weighting to obtain tuned LE maps. Subsequently, we sum the tuned maps and multiply them with the original scan to compute the SNR-modified scan.	71
8.1	The phantom utilized for the simulation experiments ©2022 IEEE	81
8.2	The results of the compounding for the proposed and baseline methods ©2022 IEEE	82
8.3	Compounding results obtained with the real robot ©2022 IEEE	83

List of Tables

6.1	DATASET TABLE WITH CORRESPONDENT SIZE, DATA AND SENSOR SETTINGS ©2020 IEEE	52
6.2	RESULTS OF 5-FOLDS CROSS-VALIDATION STUDY FOR VARIOUS MODELS WITH DIFFERENT TRAINING MODES. ©2020 IEEE	58
6.3	CONFUSION MATRIX FOR THE BEST MODEL PERFORMANCE EVALUATED ON THE TEST SET OF 4 SUBJECTS. ©2020 IEEE	58
6.4	THE CLASSIFICATION PERFORMANCES AND DISTANCE FROM THE GROUND TRUTH VERTEBRAE POSITION FOR ALL TESTED METHODS. FOR THE TCN METHODS THE RESULTS ARE REPORTED AS MEAN (STD) FOR THE 5-FOLD CROSS VALIDATION. ©2020 IEEE	59
7.1	The range of parameter values for the augmentation approaches (classical and physics-inspired) in our experiments. For each augmentation algorithm, its parameter is uniformly selected within the range.	72
7.2	Results of applying augmentation approaches to the two tasks of segmentation (UNet) and classification (DenseNet). The results are the average and the standard deviation (\pm) of 5-fold cross-validation	73
8.1	MATERIAL PROPERTIES OF SIMULATED TISSUES	81
8.2	EVALUATION RESULTS. ASR IS SHADOW REDUCTION, VC IS VOLUME COVERAGE. RANDOM SCANS ARE AVERAGED OVER 10 RUNS. ©2022 IEEE	82
8.3	RESULTS FOR THE PROPOSED OPTIMIZERS AND BASELINE OBTAINED WITH THE REAL ROBOT-ULTRASOUND SETUP ©2022 IEEE	84

