

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Haptic Feedback in Robot-Assisted Minimally Invasive Surgery

Tomás Pereira Curralo Cruz



MSc in Bioengineering

Supervisor: Prof. Dr. António Pedro Rodrigues Aguiar

Collaborator: Daniel Corona Oliveira Costa

September 8, 2025

Haptic Feedback in Robot-Assisted Minimally Invasive Surgery

Tomás Pereira Curralo Cruz

MSc in Bioengineering

September 8, 2025

Resumo

Feedback de força é um aspeto crucial dos sentidos humanos e, no contexto de cirurgia, permite que o praticante clínico evite aplicar força excessiva e provocar dano ao paciente. Apesar de todas as vantagens de cirurgia minimamente invasiva e, especialmente cirurgia minimamente invasiva com assistência robótica, este tipo de *feedback* é maioritariamente perdido, contribuindo para uma perda de segurança que pode resultar de utilização de força excessiva no contacto com o paciente. Para abordar este problema, *feedback* háptico tem de ser implementado nesses sistemas, para restaurar parte dessa informação sensorial ao cirurgião e melhorar o seu desempenho.

Esta dissertação apresenta uma prova-de-conceito para um sistema de *feedback* háptico para cirurgia minimamente invasiva com assistência robótica, desenvolvendo uma solução externa, *wireless* e *wearable*, baseada numa modalidade híbrida de *feedback* visual e vibrotátil, juntamente com um algoritmo de estimação de forças sem sensores.

O algoritmo de estimação de forças usa uma *recurrent neural network* com um *encoder* de *Vision Transformer* para processar imagens de vídeo monoculares e um *decoder* de LSTM para processar o *output* do *encoder*, com dados do estado do manipulador. Duas abordagens foram avaliadas para o *encoder* de imagem, construindo sobre trabalho anterior da literatura: um *Vision Transformer* pré-treinado no *dataset ImageNet* e um *Tokens-to-Tokens Vision Transformer* treinado de raiz. Os modelos alcançaram *Root Mean Squared Errors* de 0.4158 N e 0.4411 N, respetivamente, com frequências de inferência de 66.31 Hz e 105.58 Hz. Apesar de a exatidão ainda não ser suficiente para uma aplicação clínica segura, dado que os modelos podem desorientar o operador, estes resultados são promissores e atingem tempos de inferência mais rápidos comparados com a literatura existente. Trabalho futuro inclui a aplicação de *transfer learning* à abordagem de *Tokens-to-Tokens* e o uso de uma métrica de *loss* que tem em conta tanto os erros de estimação como as diferenças dos gradientes, permitindo a modelação de variações de alta frequência.

O sistema de *feedback* háptico, construído com componentes comerciais, fornece alertas vibrotáteis baseados em limites customizáveis de força e suplementa-os com um gráfico visual do vetor de força 3-D em tempo real para compensar pelos graus de liberdade limitados em sistemas de *feedback* de vibração. O sistema foi concebido para alta autonomia e mobilidade, com um *design wireless* em duas braceletes, com baixo consumo elétrico. Mesmo que o sistema não tenha sido criticamente avaliado no seu conforto e desempenho, demonstra com sucesso a capacidade desta abordagem vibrotátil. Trabalho futuro nesta abordagem envolveria a análise da discernibilidade dos modos de vibração aplicados, assim como a ergonomia para integração com plataformas de cirurgia minimamente invasiva com assistência robótica.

Em conclusão, este trabalho contribui para a área de robótica cirúrgica desenvolvendo técnicas de estimação de força sem sensores e criando uma base para uma solução *feedback* háptico externa e *wireless*. Ao abordar os desafios identificados, este sistema pode auxiliar na redução de danos nos tecidos de pacientes e contribuir para a criação de um ambiente cirúrgico mais seguro.

Palavras-chave: Cirurgia minimamente invasiva com assistência robótica, Estimação de forças sem sensores, *Feedback* Háptico, *Vision Transformer*

Abstract

Force feedback is a crucial aspect of human senses and, in the context of surgery, allows the clinical practitioner to avoid applying excessive force and provoking damage to the patient. Despite all the advantages of minimally invasive surgery and, especially robot-assisted minimally invasive surgery, such as reduced hospitalization time and internal bleeding, this type of feedback is mostly lost, contributing to a loss of safety that may result from excessive force usage when interacting with the patient. To tackle this issue, haptic feedback must be implemented in such systems, to restore some of that sensory information to the surgeon and enhance their performance.

This dissertation presents a proof-of-concept for a haptic feedback system for robot-assisted minimally invasive surgery by developing an external, wireless, and wearable solution, based on a hybrid visual-vibrotactile feedback modality, coupled with a sensorless force estimation algorithm.

The force estimation algorithm uses a recurrent neural network with a Vision Transformer encoder to process monocular video frames and an LSTM decoder to process the output from the encoder, along with manipulator state data. Two primary approaches were evaluated for the image encoder, building on previous work from the literature: a Vision Transformer pre-trained on the ImageNet dataset and a Tokens-to-Tokens Vision Transformer trained from scratch. The models achieved Root Mean Squared Errors of 0.4158 N and 0.4411 N, respectively, with inference rates of 66.31 Hz and 105.58 Hz. The accuracy achieved is not yet sufficient for safe clinical application, as the models can misguide operators, particularly when presented with high-frequency force variations. Nonetheless, these results are promising and offer faster inference times compared to existing literature. Future improvements could include applying transfer learning to the Tokens-to-Tokens approach and using a mixed loss metric that accounts for both force estimation errors and gradient differences, to allow for high-frequency variation modeling.

The haptic feedback system, built with off-the-shelf components, provides vibrotactile alerts based on customizable force thresholds and supplements this with a real-time visual plot of the 3-D force vector to compensate for the limited degrees of freedom in vibration-based feedback. The system is designed for high autonomy and mobility, with a wireless double-bracelet design, with low power consumption. Although the system was not critically evaluated for its comfort and performance, it successfully demonstrates the feasibility of this vibrotactile approach. Future work for this approach would involve the assessment of the distinctness of the applied vibration modes, along with overall ergonomics for seamless integration with an RMIS platform.

In conclusion, this work contributes to the field of surgical robotics by advancing sensorless force estimation techniques and creating a foundation for an external, wireless haptic feedback solution. By addressing the identified challenges, this system can help reduce patient tissue damage and contribute to the creation of a safer surgical environment.

Keywords: Robot-assisted Minimally Invasive Surgery, Sensorless Force Estimation, Haptic Feedback, Vision Transformer

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework to achieve a better and more sustainable future for all. It includes 17 goals to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice. This dissertation relates to the UN Sustainable Development Goal (SDG) 3, which focuses on "Good Health and Well-Being". The work contributes to this goal by addressing challenges in robot-assisted minimally invasive surgery (RMIS), aiming to improve patient safety and surgical outcomes. By researching and looking to improve these technologies, this work also contributes to SDG 9, which focuses on "Infrastructure, Industrialization and Innovation".

The specific Sustainable Development Goals mentioned have the following names:

SDG 3 Ensure healthy lives and promote well-being for all at all ages

SDG9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

SDG	Target	Contribution	Performance Indicators and Metrics
3	3.4	The project serves as a proof-of-concept for a haptic feedback system to prevent tissue damage from excessive force, which has the potential to contribute to reducing mortality and improving patient quality of life.	Successful demonstration of sensorless force estimation to be implemented in robot-assisted minimally invasive surgery, along with vibrotactile-based feedback alerts based on customizable force thresholds.
9	9.5	This dissertation presents a novel, wireless haptic feedback system for robot-assisted surgery, showcasing advancements in sensorless force estimation using a recurrent neural network, thereby promoting innovation and a new generation of surgical robotics.	Development of a novel force estimation algorithm. Successful integration of a vision-based force estimation algorithm with a haptic feedback system.

Acknowledgements

Primeiro que tudo, gostaria de agradecer ao Professor Pedro e ao Daniel pela oportunidade de trabalhar neste tema, e por todo o apoio, disponibilidade e paciência que demonstraram ao longo do trabalho. As suas indicações e conhecimento foram cruciais durante todo o processo, e por isso agradeço.

Ainda na veia académica, queria dar um agradecimento especial ao professor Joaquim Gabriel por todo o apoio que me deu ao longo do curso e por ser uma figura de destaque no corpo docente pelo seu interesse e amizade para com os alunos.

De seguida, não podia deixar de agradecer ao meu pai e à minha mãe por sempre terem sido uma presença tão importante em todo o meu desenvolvimento, e pelos sussurros e a mímica do "estás em reunião..?", com tudo o que fizeram para que eu me pudesse focar a 100% no trabalho. À Margarida por ser uma constante inspiração para tudo o que faço e por me mostrar o quanto podemos puxar por nós mesmos e aventurar-nos por aí fora. Ao Martim por tirar tempo das suas férias para me facilitar a vida e ser um momento de leveza pelo meio de todas as complicações da tese.

Ao Farol por serem a rede em que me suporto, é com vocês que processo muita da minha vida e não escolheria de outra forma para uma etapa tão importante como o mestrado. Sabem bem o quão importantes são para mim, portanto não nos foquemos demasiado em dizê-lo, ou guardemo-lo para a Bemposta.

A todos os bastardinhos, por terem feito estes 5 anos mais suportáveis, por entre os guerreiros no matlab, os projetos de sites para a máfia ou o trailer do eegaming, até ao São João, Fão e as queimas, momentos esses em que realmente percebemos o quanto precisamos e gostamos uns dos outros. Todos vocês me marcaram, sem exceção.

Aos the boyz pelas noites passadas a rir de como o Félix calçou o cogumelo explosivo e nos mandou todos pelo ar, ou como o Special sozinho conseguiu trazer-nos todos de volta à nave. Todos esses momentos de relaxamento foram uma parte importante das memórias que levo comigo do nosso tempo cá.

Queria também agradecer ao NEB e ao pessoal que lá conheci, foi uma jornada curta, mas marcante neste percurso, e sei que o núcleo continuará a aumentar a fasquia de todos os anos. Estarei atento e interessado durante o desenvolvimento de todos os projetos em que participámos.

Adicionalmente, queria agradecer à Catarina por ter sido uma agradável surpresa e mais agradável ainda companhia desde o Erasmus para cá.

Por último, queria agradecer a toda a comunidade da FEUP por ter fomentado um ótimo ensino e ter contribuído para a fantástica experiência durante os 5 anos em que cá estive.

Tomás

“Practice any art... no matter how well or badly, not to get money and fame, but to experience becoming, to find out what’s inside you, to make your soul grow.”

Kurt Vonnegut

Contents

1	Introduction	1
1.1	Context	1
1.2	Objectives	2
1.3	Structure	3
2	Literature Review	4
2.1	Force Sensing	4
2.1.1	Direct Force Sensing	4
2.1.2	Indirect Force Sensing	7
2.2	Haptic Feedback	15
2.2.1	Force Feedback	15
2.2.2	Multimodal Force Feedback	16
2.2.3	Sensory Substitution	17
3	Haptic Feedback Development	20
3.1	Datasets	20
3.2	Force estimation algorithm	23
3.2.1	Pre-processing	23
3.2.2	Algorithm Architectures	23
3.3	Haptic Feedback Integration	25
3.3.1	Wearable Haptic Components	26
3.3.2	Full live haptic feedback system	27
4	Results and Discussion	29
4.1	Force estimation	29
4.2	Haptic system	38
5	Conclusions	43
	References	45
A	Table of Vibrational modes for DRV2605L driver	54

List of Figures

1.1	The daVinci surgical robot, adapted from [11]	2
2.1	Examples of direct force sensing technologies. 1) Strain gauge sensor proposed by Trejos et al. [22]; 2) Fiber Bragg Grating-based sensor proposed by Lv et al. [25]; 3) Capacitive sensor proposed by Kim et al. [36], 4) MEMS sensor proposed by Liu et al. [30]	6
2.2	Reaction of the developed lattice when force is applied, Reproduced with permission from [52] © 2014 IEEE	10
2.3	Elementary LSTM Block, reproduced with permission from Springer Nature [54]	10
2.4	Vision Transformer architecture, as explained in the original source. Reproduced with permission from [61]	13
2.5	Token-to-token module and sequential image embedding instead of rigid patch tokenization, adapted from [65]	14
2.6	Moving platform operation schematic, with differential inflation of A, B, C and D for shear force feedback and E for normal force feedback, from [73]	17
2.7	Video sequence of knot-tying trial, with translucent 3-color circles for visual feedback [78]	18
3.1	Examples from both Dvrk and DaFoEs dataset. On the left is Chua’s configuration [46] to collect the images for the dataset, while on the right is an example from Reyzaal’s [37] dataset.	20
3.2	Schematic for the full pipeline of the developed system. Note that due to working with pre-recorded images, the operator to manipulator interaction is removed, even if it would be present in a final application of this system.	21
3.3	A) Architecture for the T2T approach. B) Architecture for the Transfer Learning approach	24
3.4	Full wearable vibrotactile system developed. The bracelets have a dual strapping mechanism, allowing to protect the electronic components by integrating them between both straps	27
4.1	Training and validation loss evolution over epochs for the first badly-performing T2T networks. Due to showing no meaningful learning, this model’s training was interrupted	30
4.2	Training and validation loss evolution over epochs for the best performing network in the T2T approach	31
4.3	Training and validation loss evolution over epochs for the Transfer learning approach. On the left, the results for the first trained network and on the right, the results for the second best performing network of this approach.	31

4.4	Training and validation loss evolution over epochs for both approaches with hidden size 128 for the decoder. On the left, the results for the T2T model and on the right, the results for the Transfer learning model. Due to lack of significant development for the training of the T2T approach, it was interrupted. The Transfer learning training also was interrupted, but only when overfitting began.	32
4.5	Segment taken from figure 4.7, for Fz estimation, for interval from 14 to 17 seconds	34
4.6	A) Best performing model for the T2T approach. B) Best performing model for the Transfer Learning approach	35
4.7	Results obtained for the comparison of the model's predictions to the ground truth over an unseen test run.	36
4.8	Prototyping of the DRV2605L driver and the haptic motor	38
4.9	Components integrated in the wearable bracelets. On top, the back bracelet, with the Beetle and battery (not shown). On the bottom, the fore bracelet with the driver and haptic motor	39
4.10	Photograph of the stitching on the back of the fore bracelet, for stabilization of the DRV2605L haptic driver	40
4.11	Demonstration of the sewn clothstrap, stabilizing the battery in the bracelet. . . .	40
4.12	Visual feedback system developed for the live implementation. On the left side is represented the current image being processed and on the right side, the 3-dimensional force vector plot corresponding to the model's estimation.	41
4.13	Snapshot from mobility demonstration, available at [90]	42

List of Tables

2.1	Mechanical properties of biological tissues. Adapted from [45]	8
3.1	Current Consumption Estimate	26
3.2	Vibration modes applied per Danger level	28
4.1	Comparison of Transfer and T2T Models for RMSE and Mean and Standard Deviation of the Absolute deviation for directions Fx, Fy and Fz	37
A.1	DRV2605L Haptic Driver Waveforms	54

List of Acronyms

2D	Two dimensional
3D	Three dimensional
BLE	Bluetooth Low Energy
BLSTM	Bilateral Long Short-Term Memory
CNN	Convolutional Neural Network
FBG	Fiber Bragg Grating
FEA	Finite Element Analysis
GDL	Gradient Difference loss
GRU	Gated Recurrent Unit
IMU	Inertial Measurement Unit
LIM	Light Intensity Modulation
LSTM	Long Short-Term Memory
LSTM-CIFG	Long Short-Term Memory with Coupled Input-Forget Gates
MEMS	Micro-Electrical Mechanical Systems
MIS	Minimally-Invasive Surgery
MSE	Mean Squared Error
NLP	Natural Language Processing
NMES	Neuromuscular Electrical Stimulation
PSM	Patient-Side Manipulator
QTC	Quantum Tunneling Composites
RGB	Red, Green and Blue
RGBD	Red, Green, Blue and Depth
RMIS	Robot-assisted Minimally Invasive Surgery
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SNR	Signal to Noise Ratio
T2T-ViT	Tokens-to-Tokens Vision Transformer
UKF	Unscented Kalman Filter
UTF-8	Unicode-Transformation Format - 8 bit
ViT	Vision Transformer

Chapter 1

Introduction

1.1 Context

Minimally Invasive Surgery (MIS) refers to a set of surgical techniques whose main focus is to reduce the number of incisions and openings necessary to perform an operation, while maintaining the results obtained by open surgery. This approach aims to reduce intraoperative blood loss and, consequently, reduce discomfort and danger for the patient, along with reducing hospitalization time [1, 2]. These benefits are present across various surgical fields, including oncologic procedures [3], spine surgery [2, 4], and even orthopedic soft-tissue repair [1].

Even though the advantages over open surgery were clear and led to frequent and widespread use when MIS first came about, eventually, its integration with automatic systems in Robot-assisted Minimally Invasive Surgery (RMIS) seemed promising, as its imagined capabilities greatly exceeded the limits of the actions MIS allowed. Shortly after, the NASA Ames center developed research into telepresence for surgery, later being adapted into a fully-working system at Stanford Research Institute [5, 6]. Then, the first telemanipulators started to appear on the market, with ZEUS and daVinci being some of the most notorious products [7]. These robotic platforms consist of two main hubs, as may be seen on figure 1.1: the console, where the surgeon is seated and where control of the second hub, the patient-side cart, with the robotic end-effector, is possible. There's also a visual component, as the patient-side hub is broadcasting image to the surgeon's console [8]. All this serves to allow for finer control, better flexibility and reduced fatigue, thanks to its ergonomic design, ultimately working towards improving patient outcome [7–9]. As expected, however, these systems don't come without disadvantages, and their high upfront cost is one of the most significant, along with the extensive training required from medical practitioners to safely utilize this equipment and make the most of their capabilities [8, 10].

However, with the integration of robotic systems as an interface between surgeon and patient, a new issue arises: namely, the lack of force feedback [12]. This leads to a higher degree of danger for the patients, as excessive interaction forces may occur, resulting in undesirable tissue damage [12, 13]. Thus, the surgeon must train to rely on visual cues to judge the applied forces [10], which is less than ideal, and constitutes a disadvantage for these systems. As such, much research

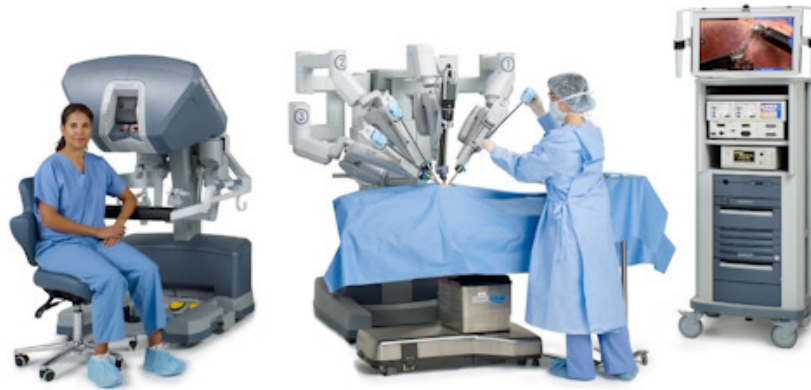


Figure 1.1: The daVinci surgical robot, adapted from [11]

has been conducted to convey some of that haptic information to the surgeon. Even so, the force measurement itself still constitutes a significant challenge, as the application of a force sensor on the robot end effector would need to be subject to strict sterilization and would add to the bulkiness of the system, while current trends point to a miniaturization of robotic systems [14]. With this, the most promising force measurement approach for these systems is indirect, through the use of some external tool. This approach will be further explored and discussed in chapter 2.

1.2 Objectives

This dissertation focused on two main points:

Firstly, the development of an external force estimator for Robot-assisted Minimally Invasive Surgery, with no requirement for force sensors. This later evolved to the development of a force estimation algorithm from monocular video frames, along with robot state vector.

Secondly, the development of a haptic feedback system, capable of harnessing the predicted force from the estimator and delivering it to the surgeon on the console-side.

By focusing on these two objectives, it would be possible to develop a haptic system working semi-independently from the surgical robot, so long as the monocular video frames and robot state vector can be provided to the system in real-time.

Finally, after the main objectives were concluded, a live demonstration was intended, with a simulation of live recording, force estimation and haptic feedback. Due to the early stage of development in this system, the feedback loop would remain open with the video feed being a simulation from pre-recorded files, and the operator not in control of the RMIS platform.

1.3 Structure

Besides the current introduction chapter, which focused on contextualizing and motivating the project, this dissertation will have the following structure:

Chapter 2 will discuss most of the current state-of-art solutions on force estimation and force feedback for RMIS, along with discussing future trends.

Chapter 3 will focus on describing the tools used and explain the choices made along the project.

Chapter 4 will present the obtained results and discuss their meaning and value.

Finally, Chapter 5 will focus on the conclusions taken from this project and future work that may be done to achieve better results.

Chapter 2

Literature Review

This chapter will focus on the literature review of Force Estimation Algorithms, Haptic Feedback, and their integration in the context of RMIS.

2.1 Force Sensing

One of the key challenges in generating force feedback in RMIS is force sensing itself. While, in other fields, that could be solved by the use of force sensing equipment such as tactile and strain sensors [15, 16], in RMIS, this solution isn't as practical. This is because direct force sensing, like with strain and tactile sensors, requires those sensors to be placed at or near the zone of contact between the tool and the tissue [17]. This imposes two main constraints on the sensors: their size must be miniaturized, and they must not only be biocompatible, but also water-resistant and insensitive to temperature changes [17–19]. As a consequence, the bulk of recent research for force sensing has shifted to indirect force sensing, a trend further reinforced by the increase in computational power. That said, direct sensing has also been explored [19, 20].

2.1.1 Direct Force Sensing

Most direct force sensing technology is done by way of strain gauges [21]. Even though these sensors are small and low-cost [20], to be usable in the context of RMIS, they must be subjected to costly special surface treatment and several sterilization cycles, with most losing reliability after the seventh cycle, even with specialized coating [22]. Strain gauges are even susceptible to electromagnetic noise, which is undesirable when working with other electric tools [23].

While strain gauges offer a low-cost solution, their drawbacks have led researchers to explore other direct sensing methods, such as optical sensing technology, not to be confused with vision-based sensor-less estimation, which will be discussed in section 2.1.2. Most of these are based on Fiber Bragg Grating (FBG), a technique that harnesses the photosensitivity of optical fiber to modulate the index of refraction in its core [24]. This makes it so when shining broadband light on the fiber, specific wavelengths are reflected. However, under strain, these wavelengths change proportionally to the strain and thus allow the measurement of applied force [23, 25]. FBG is widely

considered a good option for force sensing thanks to its high sensitivity and Signal-to-Noise Ratio (SNR) [20]. Even so, the costs associated with it can get excessive, due to the optical interrogator, whose commercially available options often exceed 10 000\$ [26]. The remaining option for optical approaches is based on Light Intensity Modulation (LIM), by way of phototransistors and photodiodes [20]. However, due to being vulnerable to light intensity variations from temperature shifts and fiber bending, these methods are not the focus of much research [20, 25].

The other main direct sensing technology, Capacitive sensing, works on the basis of capacitance differentials generated by movement between two different electrodes. This can be explained with equations (2.1), (2.2) and (2.3), where C is the capacitance, ϵ_r is the dielectric constant of the material between the two plates, A is their overlapping area and d is the distance between them, Q is the stored charge and V is the electrostatic potential.

$$C = \epsilon_r \frac{\epsilon_0 A}{d} \quad (2.1)$$

$$C = \frac{Q}{V} \quad (2.2)$$

$$F = \frac{\epsilon_r \epsilon_0 A V^2}{2d^2} = \frac{C V^2}{2d} \quad (2.3)$$

By varying either the distance between the plates or their overlapping area, a difference in capacitance is generated. This difference is proportional to the changes applied, and thus, force can reliably be measured by measuring the capacitance [27, 28]. Even though capacitive sensors provide a higher sensitivity than strain gauges, they suffer from a limited range and may lose reliability with temperature and humidity changes [27].

Even though Micro-Electrical Mechanical Systems (MEMS) work on the basis of previously described phenomena, they constitute a considerable evolution of the rest by allowing for further miniaturization, and as such, are considered a separate sensing technology [20]. MEMS make use of integrated circuit batch-processing to further enhance their cost-effectiveness and may even be coupled with specific coatings to better integrate with biomedical systems [29, 30].

Besides all the described systems, there was some research done on novel methods, albeit with reduced data. These include Piezoresistive sensors, such as the one proposed by Hou et al. [31], which harnessed MEMS technology to produce a 2 mm x 2 mm x 0.4 mm sensor, designed to be integrated in the tip of the robotic gripper. This type of sensors harnesses the piezoresistive effect, a phenomenon through which the electrical resistance of materials is altered by mechanical deformation. Such an effect is translated in the gauge factor (GF_l), which correlates relative changes in resistance (R) to relative changes in length (l), that can result from applied forces, as seen in equation (2.4) [32].

$$GF_l = \frac{dR/R}{dl/l} \quad (2.4)$$

Piezoelectric sensors were also considered, related as they are to piezoresistive: as opposed to mechanical deformation provoking a change in resistance, in piezoelectric sensors, mechanical deformation generates an electrical current that is proportional to the deformation applied [32].

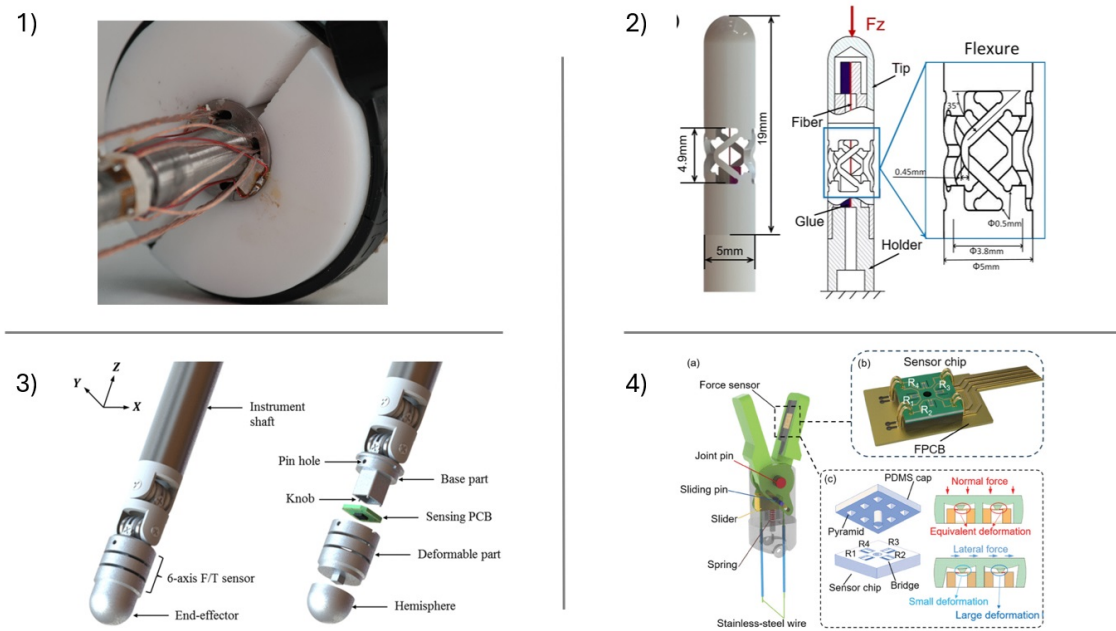


Figure 2.1: Examples of direct force sensing technologies. 1) Strain gauge sensor proposed by Trejos et al. [22]; 2) Fiber Bragg Grating-based sensor proposed by Lv et al. [25]; 3) Capacitive sensor proposed by Kim et al. [36], 4) MEMS sensor proposed by Liu et al. [30]

However, piezoelectric sensor readings suffer not only from charge leakage but also from changes in temperature, which reduces their reliability for RMIS [20]. With a similar working to the mentioned pressure (*piezo*) activated sensors, there are special composites, known as Quantum Tunneling Composites (QTC), which make for extremely sensitive force sensors, with their resistance varying greatly in response to mechanical deformation, being sometimes described as a "ultrahigh pressure-sensitivity" [33]. This is an interesting property for force sensing, albeit undermined by their low resistance to temperature changes [20].

Lastly, some research has been done into the use of accelerometers for measuring vibration frequency and phase shifting caused by applied forces in an attempt to achieve a low-cost force sensor. This, however, isn't a promising system for RMIS, as it is slow to generate a stable signal, on the scope of 0.1-0.15 seconds. [20, 34]

Mohareri et al. [35] went for a different approach when considering direct force feedback. They took advantage of the double patient-side manipulator (PSM) configuration of the daVinci Research kit and used one manipulator as an action manipulator, using the other one as the reaction manipulator. Essentially, while one manipulator performed the action, the other remained still, with the function of recording the forces and sending that information to the console on their respective side. The users could switch between the action hand as the left or the right, with the reaction changing accordingly, by pressing one of the pedal buttons on the tray.

Even though some of the solutions presented for direct force sensing may be promising, apart from the issues mentioned, they include issues in miniaturization, biocompatibility and cost, with

some of the solutions dependent on a specific RMIS platform, such as daVinci. This further encourages research into indirect sensing, as the trend suggests most of the research done in the recent past is on sensorless force estimation [17, 20].

2.1.2 Indirect Force Sensing

The research into sensorless force estimation is extensive and, as mentioned, a very recent booming topic, with countless paradigms being applied. They can be sub-divided into two main areas: control-based estimation, which focuses on modeling the surgical tool, and uses information on the robot's state to estimate the grasping force, and vision-based estimation, which uses image-capturing tools, either stereo or monocular, to estimate interaction forces [17]. Even so, some image-based methods make use of robot state vector, and, as will be discussed, achieve better results than with single-approach methods [37]. It should be noted, however, that models that take into account robot state information are not entirely sensorless, since they are measuring the data for the state, even if they don't present the same issues as previously mentioned sensor-dependent force sensing [20].

Control, or model, based force estimation further subdivides into two main approaches: analytical models, encompassing both physics-based and observer-aided models, such as Kalman filters, and data-driven, where the model is a black box that is achieved by machine learning methods [20].

Analytical models often use observers, algorithms focused on estimating the state of a system from its inputs, with Kalman filters being specialized observers that take into account probabilistic data. First described in 1960 by R.E. Kalman, these filters provide a recursive solution to the least-squares method, capable of estimating future states, even when the precise nature of the modeled system is unknown [38]. To do this, they make use of sensor readings for a parameter of the state, such as joint motor currents, and, by having a set of equations for state evolution and another for measurement evolution, are able to achieve a better estimate than by just modeling the system [38, 39]. Li et al. [39] proposed a method based on Unscented Kalman Filter (UKF) and dynamic modeling of the full system, with the motor torque and motor encoder as the observations. Despite its good precision, it doesn't actually estimate force, but the torque applied by the force on the gripper joint, while also conceding that the dynamic model could benefit from further tuning. Liu et al. [40] also propose an observer-aided method, with a model based on inverse dynamics, that estimates the force from joint torque information. After modeling the system, they apply Stribeck friction equation to further correct the estimations, achieving a lower error than with Coulomb-viscous equation, which they used to determine the robot's dynamics parameters, but still couldn't fully describe the nonlinear characteristics of friction at low speed.

Data-driven, black box, models refer to the use of supervised machine learning methods, such as neural networks [41], to fit a force estimation output from a series of input states in training data, and then applying that model to unseen states [20]. Li et al. [42], for example, propose a method based on Gaussian Process Regression, which is similar to Bayesian Regression, insofar as both adapt the Bayes theorem, to model the relationship between the input X and output Y , by fitting a

function f to known input and output, X^* and Y^* . They differ where Bayesian Regression assumes linearity for state space function f , while Gaussian utilizes a kernel-based learning method that adapts better to non-linear data. This, however, has a disadvantage when it comes to generalizing this model, as the Gaussian kernel method shows issues when predicting data outside of its seen (i. e. in training) samples. While Li et al. justify this decision by mentioning the ease of simulating most of the workplace for a specific application, this system's expandability issues are undeniable.

The issues described for Li's method [42] lead to an overall conclusion related to model-based force estimation: by nature, it's limited to the robotic system with which it was developed, and as the dynamics can change wildly between different systems, this poses a severe issue for the transferability of these methods [20].

In an attempt to circumvent these issues and develop a system-agnostic force estimator, we can possibly look at vision-based estimators, that rely on video recordings as their main information [17, 20].

Indirect vision-based force sensing works on the basis of Hooke's law (2.5), which states that the applied force (F) in an elastic material is directly proportional to the displacement (x) it provokes [43]. k is a constant that is related to the mechanical properties of the material, such as its Young Modulus [44]. By learning to estimate displacement from images, the models can easily transfer that knowledge into estimating the applied force.

$$F = kx \quad (2.5)$$

However, this also reveals a possible flaw in such systems, as different tissues of the human body present wildly different mechanical properties, with the Young's modulus going from 1.0 MPa for arterial walls to 2500 MPa for tendons and ligaments, as seen in table 2.1 [45].

Material	Modulus (MPa)
<i>Soft tissue</i>	
Arterial wall	1.0
Hyaline cartilage	0.4–19
Skin	6–40
Tendon/ligament	65–2,500
<i>Hard tissue (bone)</i>	
Cortical	16–20 GPa
Cancellous	4.6–15 GPa

Table 2.1: Mechanical properties of biological tissues. Adapted from [45]

Vision-based methods are a convenient solution for RMIS in part because, initially, RMIS systems already need good imaging tools, in the form of either stereo or mono video, which is one of the requirements for vision force estimation [9, 23]. In fact, when lacking haptic feedback systems, surgeons end up judging the applied forces by visual means, applying Hooke's law unknowingly [20, 46, 47]. And, as mentioned, vision force estimators will simulate a similar approach, firstly quantifying estimated displacement and from that, applied force. [47]

Most vision-based applications use some sort of machine learning algorithm in their pipeline, with focus on supervised learning [20]. This is the most dominant method in machine learning and refers to training methods where the model has access to annotated data [48], which in the case of force estimation, would be video feed imaging, with force collected by force sensors [47].

Nonetheless, some of the research done doesn't rely on completely data-driven models and actually enhance these models with deformation modeling [47]. Examples of this are Zuo's [49], who leveraged inverse finite element analysis (FEA) to model the gripper's deformation in software and just used the frames gathered from the video feed to track the gripper points, submitting that information to the FEA analysis in order to achieve a force estimation. This was done with the objective of circumventing the long process associated with training neural network models, with their need for large amounts of data. Another work, by Kennedy et al. [50] also leveraged the finite element method to calculate forces from estimated displacement, itself resulting from the processing of stereoscopic video data. This processing is made possible by the marking of dots on the surface of the membrane, a stand-in for the patient in RMIS, and tracking the distance between them over time. Even if the results show promise, finite element analysis works with tool-specific, and, in the case of tissue deformation modeling, patient-specific data, leading to this approach still presenting the issues in scalability mentioned earlier [37, 51].

As for data-driven, AI-based models, these show some variability in the research corpus, with a plethora of architectures available [47]. Besides differing in the architectures used, they can differ in stereoscopic or monocular video frames and even the use of plain image data or the additional consideration of robot state information [17, 20].

In visual-based force sensing, some research focuses on stereoscopic video recordings. Aviles et al. [52] took advantage of this and proposed a 3D Force sensing method. While in 2D methods applying force in tissue results in a deformed contour, in 3D, as in this case, this translates to a deformed surface [17]. In their work, they separate the algorithm into two main blocks, named vision block and force estimation block. The former focuses on estimating deformation on the tissue by processing video frames, while the latter takes that information, along with kinematic data from the manipulator, and estimates a force vector. Even though their system only had access to RGB (as opposed to RGBD) cameras, by placing them in a stereoscopic parallel configuration, they were able to gather depth information and establish a 3D lattice, shown in figure 2.2. This lattice could then be used to model the deformation, and process this through a Recurrent Neural Network (RNN), achieving mean absolute errors of 0.0464 N.

RNN differ from normal Neural Networks mainly by containing feedback mechanisms across time-steps that allow them to achieve a memory-like process. However, they are plagued by issues such as exploding and disappearing gradients, where the model could be led to attribute, respectively, too much or not enough importance to the weights in the network, leading to worse results [53]. This will be mitigated by Long Short-Term Memory (LSTM) networks, incorporating gate mechanisms in order to allow better control of the gradients [54]. LSTM's building block is shown in figure 2.3. These networks specialize in modeling long-term dependencies, and, while achieving good results, there are other variations besides the vanilla LSTM block shown, such as

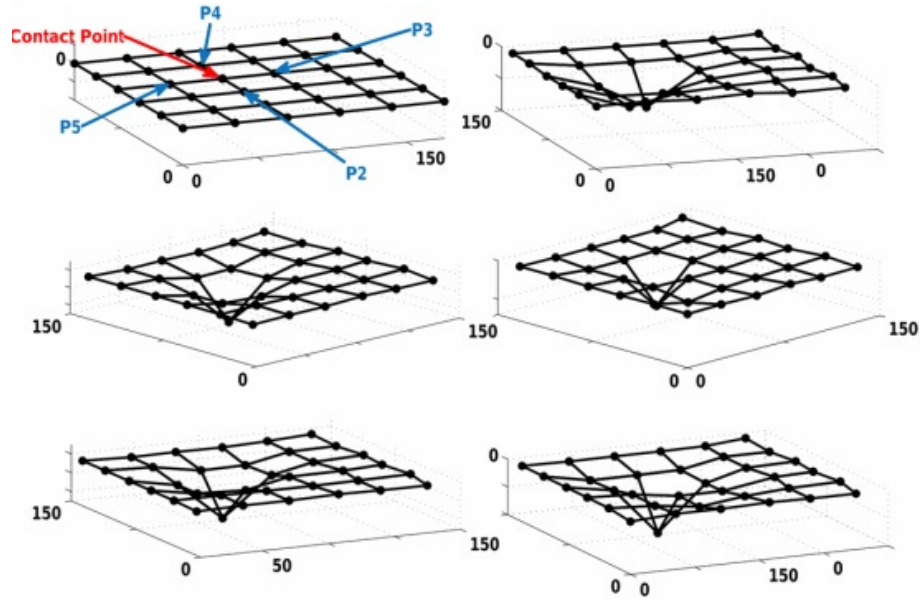


Figure 2.2: Reaction of the developed lattice when force is applied, Reproduced with permission from [52] © 2014 IEEE

Bilateral LSTM (BLSTM) which is reported to achieve even better results. However, in the context of haptic feedback for RMIS, BLSTM isn't applicable, as it takes into account future inputs, thus ruling it out for real-time integration on a haptic feedback system. Another notable LSTM variation is LSTM with coupled input-forget gates LSTM-CIFG. This means that the weight attributed to "forgetting" data from older time steps is tied to the weight of learning data from newer ones. This simplifies the network, by reducing the independent parameters it needs to learn [55].

In fact, taking into account the advantages of LSTM over RNNs, Aviles et al. later proposed a new network, working on a similar lattice for the displacement estimation, and then processing that data through an LSTM block [56].

For monocular force estimation, the combination of images to recover 3D shape isn't available. As such, algorithms must rely solely on 2D imaging. A promising architecture proposed in the lit-

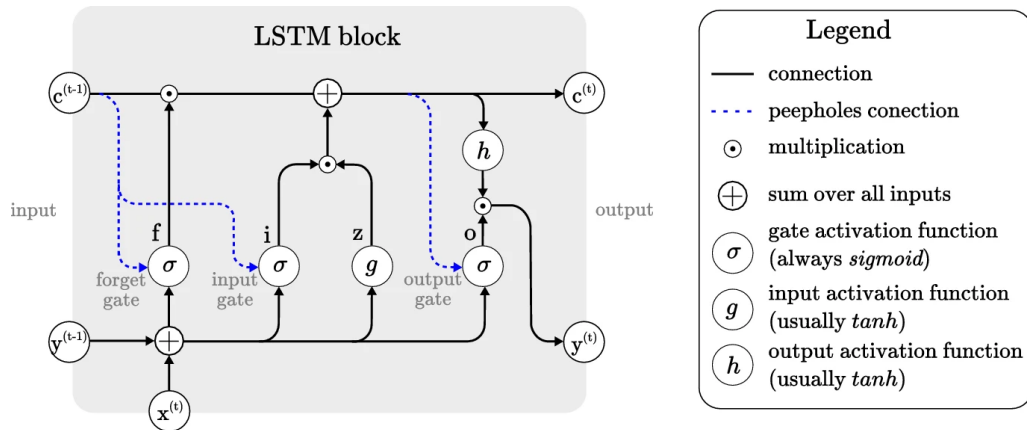


Figure 2.3: Elementary LSTM Block, reproduced with permission from Springer Nature [54]

erature is the combination of Convolutional Neural Networks (CNN) and LSTM networks [17, 47]. The idea behind this is the separation of the estimation in two blocks, as previously done by Aviles et al. [52]: the encoding block, generating spatial features from images, and the decoding block, taking those features, and their evolution over time, to reach an estimation of a 6-dimensional vector of the applied forces and torques [17]. This work will be further explored by Chua et al. [46] and Reyzabal et al. [37].

The encoding block is done by a CNN, making use of its transfer learning ability. This is a technique where neural networks simulate human learning by applying previously trained data in a source task to a new target task, instead of learning the isolated target from scratch. These tasks don't have to be the same, as transfer learning allows for reweighting to adapt the source network to the new task [57, 58]. This is expected to reduce training time, while also reducing the need for very large datasets to train for the target task, which is mentioned as one of the most worrying aspects for RMIS force estimation [20, 37, 46]. The source network used, VGG16, was developed by Simonyan et al. [59], as a very deep CNN pre-trained on ImageNet, a publicly available image classification large-scale dataset [60], achieving exceptionally good accuracy in the classification task and good generalization [59].

Leveraging all the mentioned characteristics above, Marbán et al. [17] first pre-process the video frames. Then, using their CNN adapted from VGG16, they process those frames, concatenating the result with kinematic data from the robot manipulator to form a feature vector for each time frame. A group of feature vectors for the current and previous time steps is then processed by an LSTM decoder, with the output being a 6-dimensional force and torque vector for the current time step. In this work, even the influence of using the kinematic data is studied, doing experiments with just the images, just the kinematic tool data and both combined. They also evaluated different decoder architectures between vanilla LSTM, LSTM-CIFG and Gated Recurrent Unit (GRU), with LSTM-CIFG achieving the best results. While other works such as [46] and [37] utilize mean-squared error as their loss during training, Marbán et al. also evaluate a novel loss, named Gradient Difference Loss, whose purpose it is to better model sharp transitions between time steps.

The results demonstrate the advantage of including both kinematic data and video processing for the force estimation. Besides the good results achieved by the study, some conditions need to be specified. Namely, the set-up used was static, while real conditions in RMIS are usually more dynamic [47]. Not only that, but the training and testing was done in a singular dataset created on artificial tissue, not taking into account generalizability. Besides fixing those issues, Marbán et al. also suggest, for future work, the use of an attention model to avoid lengthy pre-processing of the video, and automatically selecting important regions in each frame that most contribute to the output.

Even though it still represents an early-stage implementation, Marbán's [17] work shows promise for the force estimation task, and further research will build on its architecture. For example, Chua et al. [46] demonstrate the advantage of using not only position estimates, such as done in [17], but also velocity estimates and robot state information, such as force-torque estimates.

They also analyze architectures for Visual-only (V), State-only (S) and visual and state together (V+S). Their V architecture is composed of a Resnet50 architecture pretrained on the ImageNet dataset [60], modifying the last layer to just output a 3 dimensional force vector. Their S network was composed of 6 fully connected layers with the 3 dimensional force vector as output. The V+S network was a mixture of both, with the image encoder being the V network adapted with a 30-dimensional output, to which the state vector was concatenated. The result was then processed by 3 fully-connected layers with 3-dimensional output. From these architectures, V + S achieved the best results, with the lowest deviation from the ground truth, confirming the previous notion that Visual and State information work synergistically rather than conflicting.

Chua et al. also tried to placate the concerns related to processing time, raised by Marbán [17], by not including recurrency, instead relying on the velocity estimates to confer that time-related information. When comparing their networks to a benchmark, based on [17], they achieved similar results to the RNN, with faster processing time, showing a possible avenue for further research to follow. Nonetheless, Chua et al. still mention the issues related with the dataset, as their dataset was created from scratch on tissue phantoms with the daVinci Research kit manipulator. The lack of surgical scenarios and differing camera movements raises concerns about overfitting on the workspace position. Reyزابال et al. [37] specifically design their experiment to tackle these issues and follow on suggested work from both Chua's work and Marbán's.

Reyزابال's [37] main proposal is the creation of a larger dataset by mixing their own with Chua's [46] previously available dataset and using data augmentation to avoid overfitting to position of the workspace, lighting conditions and camera rotation. As for their architecture, they follow on the previous suggestion of using attention modules for this task and employ a Vision Transformer (ViT) [61], a state of the art architecture based on technology from Natural Language Processing (NLP), that has been achieving more promising results than CNN in classification tasks. They follow the previous research [17, 46, 52], separating the process between a ViT image encoder that converts each frame to a feature vector and a LSTM decoder, after concatenating the feature vector with kinematic and state data. However, as opposed to the work done in [17], they limit the recurrency window to 5 frames, stating that those are the minimum requirement to represent acceleration of the manipulator tool.

Vision Transformers are algorithms based on Transformers, from NLP, that take advantage of self-attention processes and were proposed by Dosovitskiy et al. in 2021 [61]. These models start by splitting the input images into patches and flattening them. Then, they linearly project each patch, resulting in tokens that represent areas of the original image. A positional embedding is added to these tokens and a special class token is appended. All of these tokens are then fed to a Transformer encoder, where they are repeatedly processed by attention modules. After the encoding, the prediction is made on the class token. Figure 2.4 shows an overview of this model. The main advantage of ViTs over CNNs in Vision-Based Force Estimation is their ability to model long-range dependencies [47], while requiring fewer training iterations [37].

Reyزابال et al. [37] compared their results to an architecture based on the work from [46], but also compared the recurrent and non-recurrent versions of both models. Another important aspect

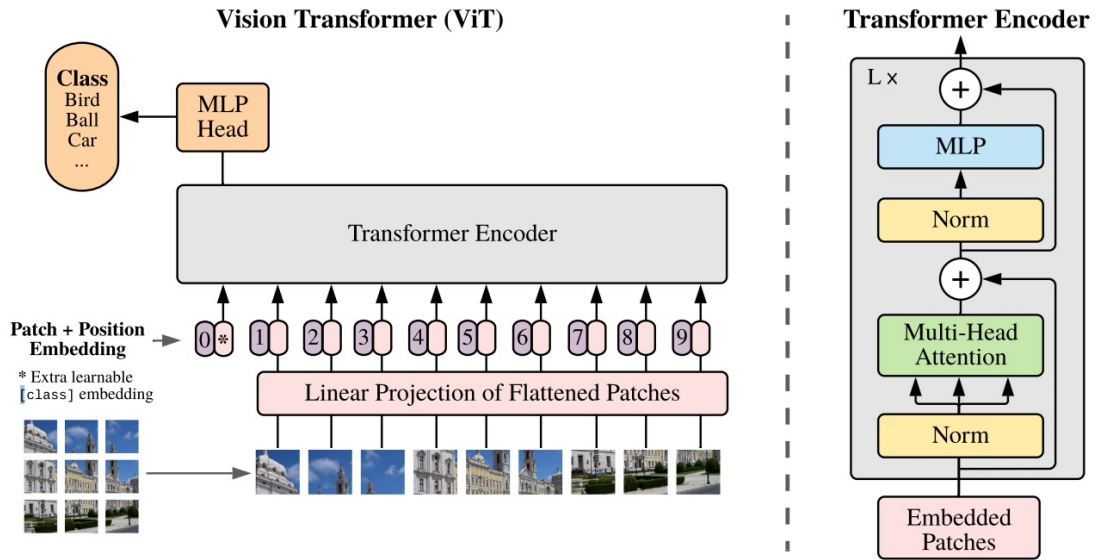


Figure 2.4: Vision Transformer architecture, as explained in the original source. Reproduced with permission from [61]

was generalizability, where they tested the results from training the architectures on one dataset and testing on the other, and the advantage of training on a mix of both, with data augmentation.

They showed that dataset mixing effectively improved the results for the networks, proving its viability. From their trained models, the Recurrent ViT achieved the best results, surpassing those of RCNN based on Chua's work. However, when comparing inference time, the difference between recurrent and non-recurrent models is apparent, with RViT and RCNN achieving inference frequencies of 12.45 Hz and their non-recurrent counterparts reaching 80 Hz. That said, another important aspect is that Reyزابal et al. [37] trained their vision transformer from scratch on a small dataset, which is undesirable for ViTs, as they benefit especially from pre-training on larger datasets [61].

Building on Reyزابal's work [37] and, consequently, on Chua's [46] and Marbán's [17], some of the issues can still be addressed: there is still a lack of real-case scenarios in the datasets, as most are done on tissue phantoms or otherwise simulated; ViTs need larger amounts of data to achieve better results and avoid the concerns with overfitting; the best model's inference time is too high for haptic feedback applications.

For the first issue, a complex set-up is needed, or even access to surgical images. However, labeling this real data would run into the same issues that direct sensing does, as it is the the most reliable method to identify interaction forces. Another alternative would be a semi-supervised learning model, as suggested by Marbán et al. [17], where a mix of labeled and unlabeled data is used [62].

For the model's inference time, simplifying the model is one of the solutions. While a 5 frame time window is enough to represent the acceleration of the tool, the solutions between 1 and 5 were not explored, and could yield promising results while speeding up inference time.

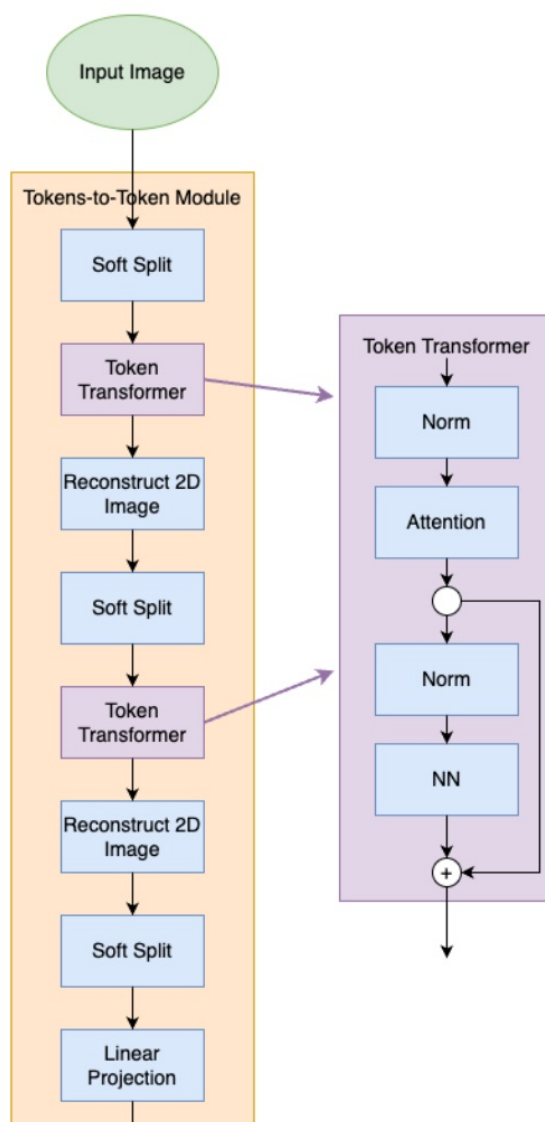


Figure 2.5: Token-to-token module and sequential image embedding instead of rigid patch tokenization, adapted from [65]

As for ViT's needs of larger data, two solutions appear: firstly, the previously mentioned transfer learning. While ViT is a recent technology, some pre-trained models are available [63], the application of which might yield more robust results. Another approach is the use of Tokens-to-Tokens Vision Transformers [64], an iteration on ViTs with the objective of achieving better results on mid-sized datasets by performing progressive tokenization of the input images. This token-to-token module is represented in figure 2.5 and is done by reconstructing an image from the tokens, applying a soft split on it and then processing it with a token transformer, which includes normalization layers and attention modules. This is then done iteratively, and the final result is processed through a ViT backbone, based on the vanilla ViT and CNN technology.

2.2 Haptic Feedback

Originating from the greek term "haptikos", which means "able to come in contact with", haptic currently refers to the science of touch, often mentioned in a techno-media context as pertaining to the touch feedback generated by artificial systems to simulate the sensation of touch [66]. In the context of open surgery, haptic feedback by directly touching the patient, or indirectly through surgical tools, allows the surgeon to feel not only shape, but also pressure, temperature, vibration and stretch to identify the tissue and the forces applied on it [67]. As mentioned in chapter 1, however, in RMIS, surgeons lose that ability and must work with other senses to estimate the applied forces, which motivates the work mentioned in 2.1. That said, once a force estimate is achieved, the question remaining is how to transmit that sensation to the operator and restore some of that function.

The literature separates the sense of touch into two different properties: Kinaesthetic feedback, referring to force and pressure feedback, and Tactile sensing, referring to the other properties measurable by touch, like temperature, vibration and texture [67–69]. Another important notion in the field of haptic feedback is Telepresence. This term refers to the ability of the surgeon to feel present in the remote operating room, even if not physically there [70, 71], which can only function if technology reaches a high level of transparency [71], in turn defined by the match between the remote environment impedance and the one perceived by the operator [70]. Besides transparency, haptic feedback systems must also be stable [67]. However, achieving high transparency can introduce instability, as the elements that allow for it can provoke oscillations and propagate errors in the closed loop [67]. For instance, the simulation of torques and forces on the leader manipulator generates friction and may unintentionally mask the signal, leading to erroneous feedback [70].

2.2.1 Force Feedback

Between kinaesthetic and cutaneous feedback, the most researched approach in the context of RMIS is kinaesthetic [67, 72], despite the recognized benefits for the inclusion of tactile information, explored in recent reviews [72, 73]. Several approaches for providing kinaesthetic force feedback are discussed in the literature, including motor-based actuation, pneumatic actuation, and electrostimulation. A common method, motor-based actuation, directly resists the movement of the surgeon's control handles to simulate the forces the surgical manipulator experiences when interacting with tissue [72]. While this approach is widespread and can achieve high transparency, it can suffer from instability and other issues because the system's feedback loop is closed at the console [67]. Another concern with motor-based systems, particularly in grasping applications, is the jerking phenomenon, which can result from the interaction between the human operator's reflexes and the motor's actuation.

To solve this issue, pneumatic actuation was proposed in [72]. This simple solution involves inserting a pneumatic tube between the handle's graspers. Increasing the air pressure within the tube resists the grasping motion, offering a straightforward way to provide feedback for this specific task. Furthermore, since this pneumatic solution is designed as a supplementary component,

it avoids the need for extensive modifications to the control handles, making it a more compatible option than motor-based systems.

More recently, Neuromuscular Electrical Stimulation (NMES) has been explored as a promising, though unproven, approach for use in RMIS. A study by Galofaro et al. [69] utilized a commercial soft robotic suit, the Teslasuit, to explore this method. The suit, which has inertial measurement units (IMUs) for body tracking and 80 electrode channels for Surface Nerve Electrostimulation [74], allowed researchers to simulate the sensation of holding a weighted cube. They achieved this by stimulating the antagonist muscles to mimic the joint torque that would be created by physical loading.

While the use of NMES for mechanical load simulation is a novel concept [69], soft robotic suits are increasingly being used in haptic feedback research due to their commercial availability and enhanced ergonomics [75]. Even so, this approach is still in its early stages. The study's results were limited by its lack of stimulation in agonistic muscles for certain movements and the absence of tactile feedback in the hand, indicating that more research is needed for complex tasks, such as its integration with RMIS.

To address the instability issue with direct force feedback, researchers have experimented with two alternative approaches. The first approach uses hybrid feedback modalities, such as a combination of kinaesthetic and cutaneous sensing [67], while the other approach is the avoidance of force feedback, conveying that information through sensory substitution, such as visual or vibrotactile feedback [67, 70].

2.2.2 Multimodal Force Feedback

Multimodal force feedback systems are considered the best compromise between transparency and stability, as kinaesthetic feedback is more transparent but less stable, while tactile feedback is the opposite [67]. To achieve this, both feedback types are decoupled and provided in separate closed loops to avoid force interference and further instability [67, 70].

Abiri et al. [72] propose both a bi- and a tri-modal feedback system, combining kinaesthetic feedback through pneumatic actuation, cutaneous normal force feedback also through pneumatic actuation and, later, vibrotactile feedback. The normal force feedback was implemented with a solenoid valve array at 5 pressure points in the operator's fingertips. Alternatively, a depressed-membrane pneumatic actuator was used, allowing for the integration of vibrotactile feedback as well, through haptic vibration motors. This type of feedback is especially useful as warning signs, as described by Abiri et. al, allowing the tri-modal system to achieve even better results than the bi-modal, while both reported a significant improvement over non-feedback.

Another implementation of hybrid feedback approaches is put forward by Lim et al. [73]. In their system, they implement both kinaesthetic and tactile feedback. The former is possible thanks to a commercial haptic system, PHANTOM Premium 1.0 [76] that emulates on the robot handle the forces felt by the manipulator. The latter is enabled by a moving platform pressed against the finger and surrounded by 4 balloon actuators on its plane and one under it. The differential inflation of opposing coplanar balloon actuators allows the platform to move and provoke shear

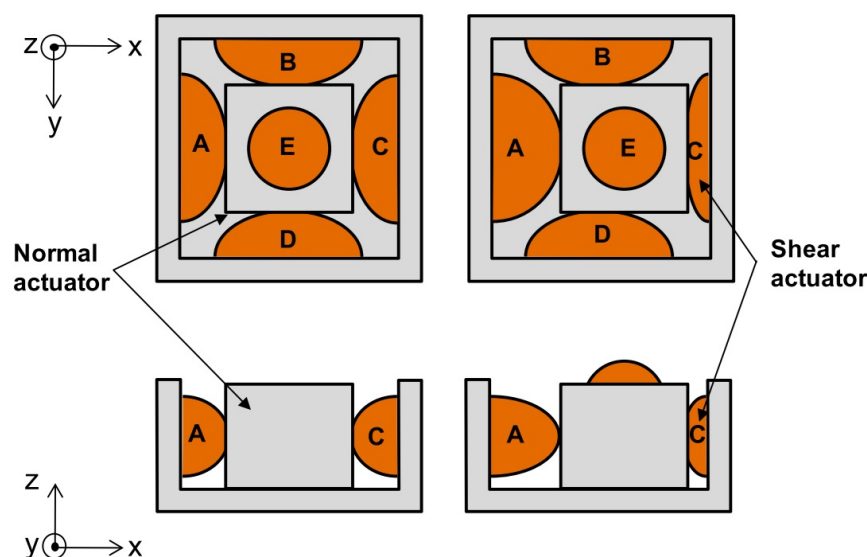


Figure 2.6: Moving platform operation schematic, with differential inflation of A, B, C and D for shear force feedback and E for normal force feedback, from [73]

forces on the operator's finger, while the inflation of the balloon under it simulates normal force. The working of this system is shown in figure 2.6.

Even if both these systems show promise, it should be noted that cutaneous sensing for different properties from contact force creates new issues. It requires not one but multiple sensors in an array configuration to adequately model local properties on the tissue [67, 70]. Additionally, while kinaesthetic information can propagate along the surgical tool, this is not the case for texture or shape information [67]. As such, direct cutaneous sensing must rely on miniaturized and biocompatible sensor arrays near the tool tip, which, as discussed in Section 2.1, is undesirable [68]. This won't be the case for most applications of sensory substitution.

2.2.3 Sensory Substitution

Sensory substitution methods are the other approach to deal with system instability, and allow for this by being decoupled from the handle's actuation by design. This approach refers to the substitution of force feedback by the use of another of the operator's senses [67].

One approach with good results in isolation was auditory feedback. In a work done by Kitagawa et al. [77], an audio tone signaled the moment the optimal force for the task of suture tying was achieved. This allowed the operators to achieve 50% more consistency than non-feedback. The authors note that the ideal objective for the auditory feedback would be a continuous feedback solution, where pitch and volume would vary according to the forces applied. However, such a system could lead to confusion or discomfort for the operator, and could even be a detriment to communication. That said, continuous feedback can be achieved with another sensory substitution approach: vision.

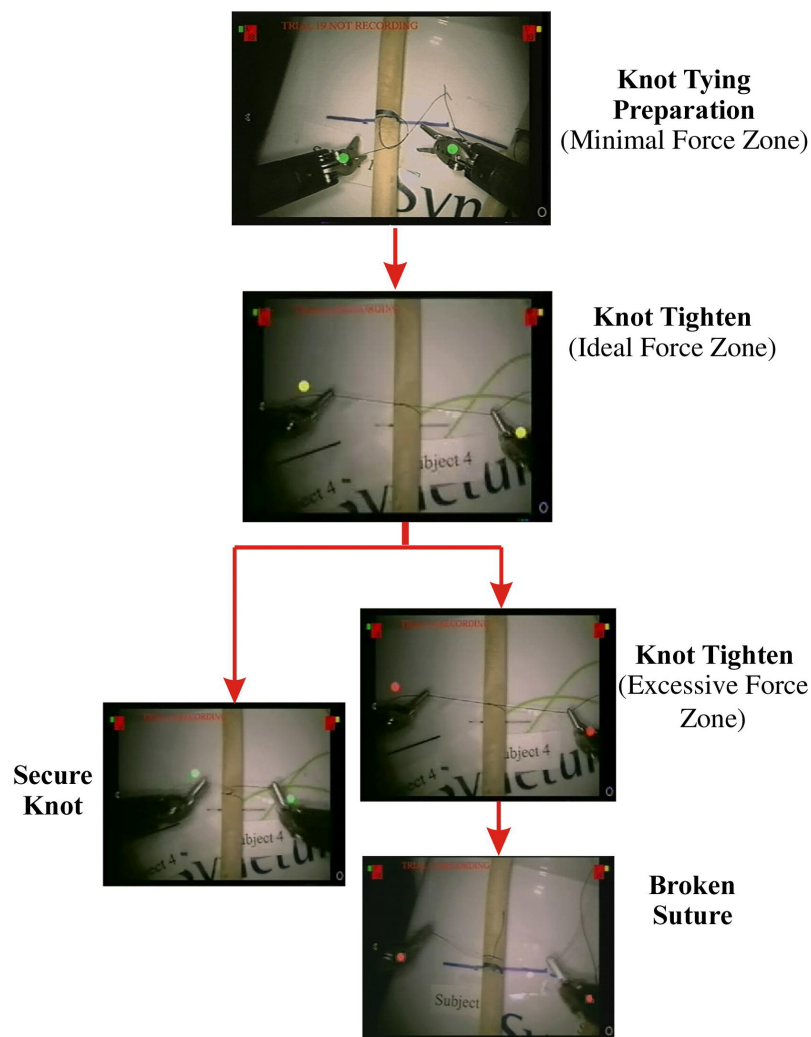


Figure 2.7: Video sequence of knot-tying trial, with translucent 3-color circles for visual feedback [78]

In the same work [77], visual feedback achieved even better consistency values than auditory feedback, with 84% increase in consistency from non-feedback tying, as opposed to 50%. This approach to sensory substitution is easily implemented: once force is measured, a plot of its values can be drawn and continuously updated with the new measured values. Properties such as size and color allow this feedback to convey intuitive and easily recognizable information. Reiley et al. [78] transmit feedback on 3 discrete force intervals, each with a different color, plotted through translucent circles on top of the manipulator tip on the surgeon's stereoscopic display, as shown in figure 2.7. Their system was evaluated in a suture task and achieved lower peak forces, lower standard deviation, less broken sutures and a slight decrease in task duration. One issue with visual feedback, however, is cited as sensory overload [77]: since the visual medium is also the medium by which the operator guides themselves, too much information can be distracting [78] or even lead to worse overall results [77].

A third approach to sensory substitution is by applying feedback on the skin without directly

simulating the forces being applied. This cutaneous feedback has been proposed in the form of skin stretching systems such as the one by Quek et al. [79], who designed three devices with this approach. The first, with 3 degrees of freedom, focused on applying in a coupled manner normal force by squeezing the operator's finger in the normal direction and tangential forces by using skin stretching on the operator's other fingers. The second device includes 6 degrees of freedom and besides communicating forces, focuses on communicating torques. And lastly, a wearable version of the first device, to allow freedom of movement. By emulating the applied forces through skin stretching, these devices can convey information on not only magnitude, but direction as well. The results show that the usage of this skin deformation approach led to an increased situational awareness and an improvement in task performance when compared to no feedback.

Besides skin deformation, cutaneous feedback can be conveyed by vibrotactile mediums. McMahan et al. [80] had the objective of conveying high frequency tool accelerations to the operator by vibrotactile feedback on the fingertip or audio feedback. While the system did not achieve statistically significant improvement in task performance, subjective reviews from the operators classify it as increasing intuitiveness and being highly preferable to no-feedback approaches. Another vibrotactile approach was discussed in section 2.2.2 with Abiri et al. [72] using vibrotactile feedback as an add-on to their bi-modal feedback system, and mainly consisting of an alert that signaled excessive force applied. This feature greatly enhanced results with inexperienced operators, unfamiliar with the RMIS system, showing its promise for learning platforms. Despite its promise, vibrotactile feedback is limited in the degrees of freedom it can convey, due to its simplicity [79] and due to the human hand not being able to discern vibration direction [80]. To circumvent this issue, spatial arrays of vibration motors were considered, where they would be triggered sequentially. Even so, the results were lackluster, since the temporal nature of this actuation limited the responsiveness of the system [79]. That said, cutaneous feedback is still promising, and can be successfully integrated in a larger haptic feedback system, when taking into account its shortcomings.

Chapter 3

Haptic Feedback Development

This chapter will focus on describing the work developed for this dissertation and some of the choices made, leading to the final pipeline, shown in figure 3.2.

3.1 Datasets

For the work developed in this dissertation, one of the first tasks was to gather a dataset of monocular video recordings from which to estimate the forces applied on the tissue. For this purpose, various authors of work developed on force estimation in the context of RMIS were contacted. From these, 2 different private datasets were obtained: DaFoEs and Single PSM Simple Tissue Phantom Dataset, from here onward referred to as Dvrk, which included both monocular video images and robot state information [37] [46].

These datasets were recorded on commercial tissue phantoms, and thus, still represent a very early stage for force estimation, with conditions being much more controllable than on live tissue. Even though this is a disadvantage, most other work on the subject is at the same stage, and it is not an issue characteristic of the obtained datasets.

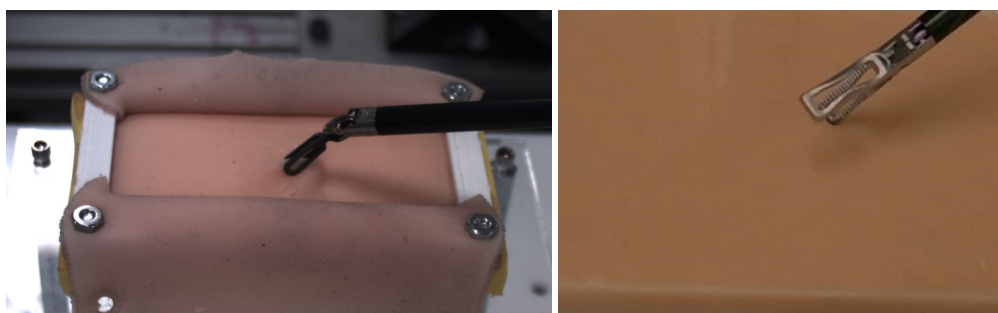


Figure 3.1: Examples from both Dvrk and DaFoEs dataset. On the left is Chua's configuration [46] to collect the images for the dataset, while on the right is an example from Reyزابال's [37] dataset.

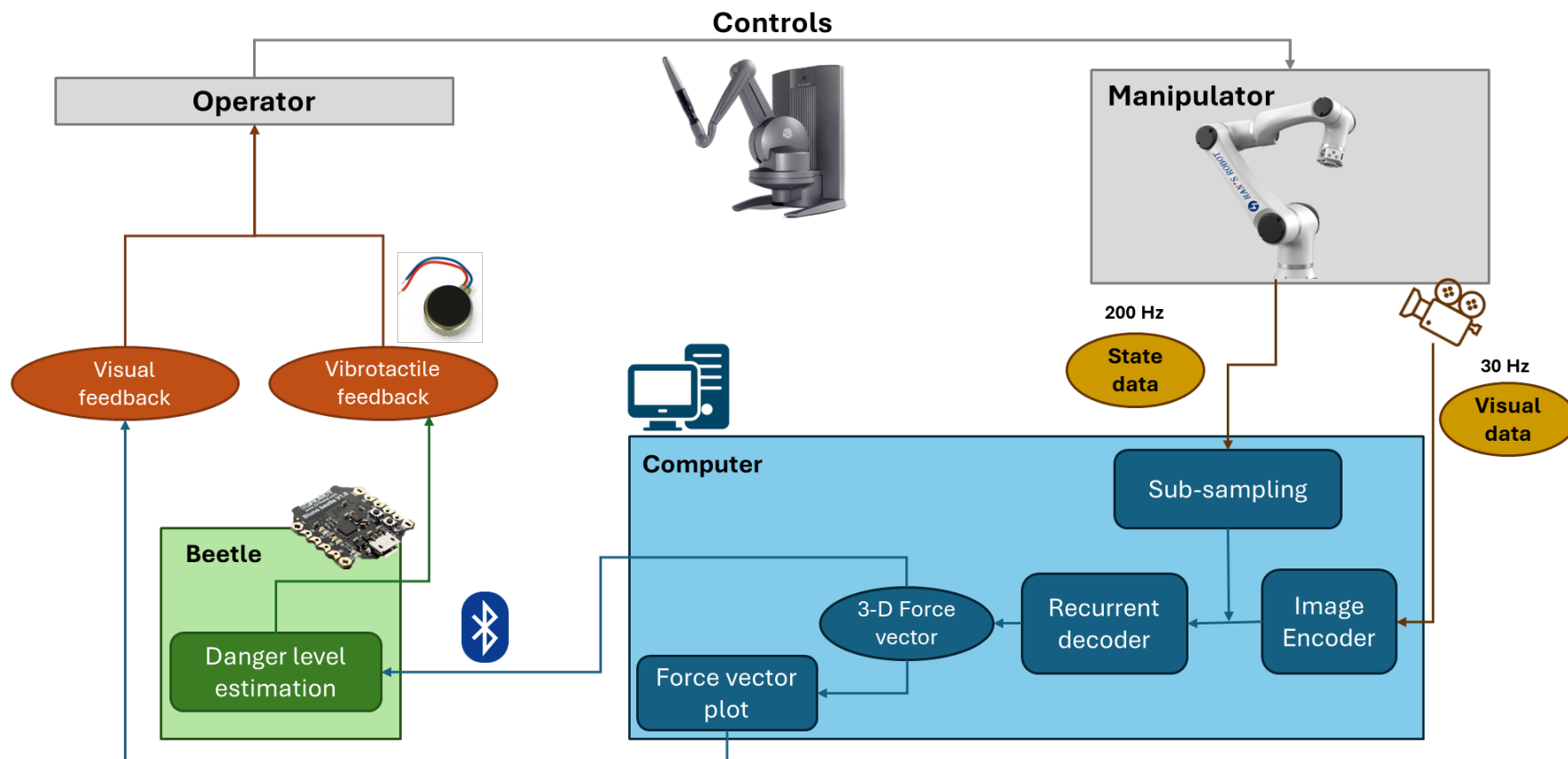


Figure 3.2: Schematic for the full pipeline of the developed system. Note that due to working with pre-recorded images, the operator to manipulator interaction is removed, even if it would be present in a final application of this system.

The Dvrk dataset was recorded over 21 different positions, 7 in the planar axis and in 3 different heights, while the DaFoEs dataset was only recorded at 1 height, over 5 planar positions. However, DaFoEs includes different combinations of single and double-layered tissue phantoms, while Dvrk only includes single-layered. As for tasks, DaFoEs is mainly focused on palpation and pushing, while Dvrk includes mainly pulling and grasping tasks. Ultimately, this results in 41 scenes included in the Dvrk dataset and 90 in the DaFoEs. Both these datasets were recorded at 30 Hz.

As for the state information, the Dvrk dataset's state was recorded at 200 Hz and included both kinematic state information, s_{kin} , and force state information, s_{force} , shown in equations (3.1) and (3.2). In (3.1), p is the world Cartesian position, o is a quaternion representing world orientation, v is the cartesian velocity, w the angular velocity, q the position of each joint, $\dot{q}$ the velocity of each joint and q_{des} the desired position for each joint:

$$s_{kin} = [p_i, o_i, o_w, v_i, w_i, q_k, \dot{q}_k, q_{k,des}] \quad i \in \{x, y, z\} \quad k \in \{1, \dots, 7\} \quad (3.1)$$

As for equation (3.2), f stands for the cartesian force in the end-effector, t the torque, τ the torque for each joint and τ_{des} the desired torque for each joint:

$$s_{force} = [f_i, t_i, \tau_k, \tau_{k,des}] \quad i \in \{x, y, z\} \quad k \in \{1, \dots, 7\} \quad (3.2)$$

These vectors combine to make a state vector of dimension 54. The DaFoEs dataset, on the other hand, was recorded with a manipulator with 6 degrees of freedom as opposed to 7 in Dvrk. This dataset's state vector had dimension 26, including p_E as the cartesian position of the end effector, o_E as the quaternion for its orientation, J as the 6 joint positions for the manipulator, p_H and o_H the 3-dimensional position and 4-dimensional orientation of the handle

$$\rho = [p_E, o_E, J_{robot}, p_H, o_H, J_H] \quad (3.3)$$

Although the Dvrk dataset is extensive, DaFoEs represents a greater number of possible scenes, while coupled with a commercial forceps as the end-effector. With this, and taking computational cost concerns into account, DaFoEs was chosen as the dataset for this work, from which the force estimation algorithms will be developed.

It should be noted that DaFoEs was developed with the intent of complementing Dvrk, and suggests a pipeline for dataset mixing, in order to achieve better and more accurate results. However, dataset mixing was beyond the scope of this project and as such, not followed through. It would be interesting to develop the algorithm with all the possible data, and even consider data augmentations, as these have shown to greatly increase the model's results, by avoiding overfitting, which may be an issue with the current set-up.

3.2 Force estimation algorithm

The force estimation algorithms were designed to build upon the work discussed in chapter 2. They were based on the separation of image encoder and decoder proposed in [17, 37, 46, 52] and were designed with the objective of placating some of the issues still present and discussed in that chapter.

3.2.1 Pre-processing

Before being subjected to the algorithm pipelines, the video frames had to be pre-processed, to create a standardized input and optimize the algorithm's results. The pre-processing steps were chosen as the ImageClassification preset used during training of the pre-trained ViT B 32 model in Pytorch [81] and consist of 4 steps: Firstly, resizing the image into size 256x256, followed by centre cropping of size 224x224, converting it to a float tensor and lastly normalizing by the mean and standard deviation of the ImageNet dataset [82].

The state information was collected at a much higher sampling frequency (200 Hz) than the image data (30 Hz) [37], so it had to be subsampled to the image sampling rate, and then normalized according to the mean and standard deviation of all state samples.

3.2.2 Algorithm Architectures

The force estimation in this work was done via vision-based algorithms, taking into account state information and estimates, as first suggested by [46]. Recurrency was an important consideration as well, with a recurrency size of 3 time steps being used, corresponding to a window of 100 ms at 30 Hz sampling frequency. As discussed previously, in chapter 2, 3 time steps were chosen as a trade-off with the objective of achieving faster inference time with low penalty to system performance, getting the force estimation algorithm closer to an adequate state for live haptic feedback systems [37].

For the image encoder, two approaches were developed, represented in figure 3.3: the first was based on the transfer learning of a ViT similar to the one Reyzaal et al. [37] trained from scratch. The network used was the torchvision model ViT B 32 [63], taken from the original source [61] and pre-trained on the ImageNet dataset [60]. This network has a kernel of 32x32 pixels for the patch embeddings and includes 12 encoding blocks. In transfer learning, the first layers of the network are expected to extract baseline information, with the last layers encoding high-level relationships [83]. As such, initially, during training, only the last 4 encoding blocks were unfrozen and fine-tuned. The network was further altered by removing the last linear layer, and just extracting the class token. This will be the output of the encoder and represent the feature vector for each image.

The other approach was the training of a T2T-ViT from scratch. The network used was based on the original [64] and was adapted from the code available at [65]. The main difference was the addition of dropout layers and the removal of the last linear layer for classification. Similarly

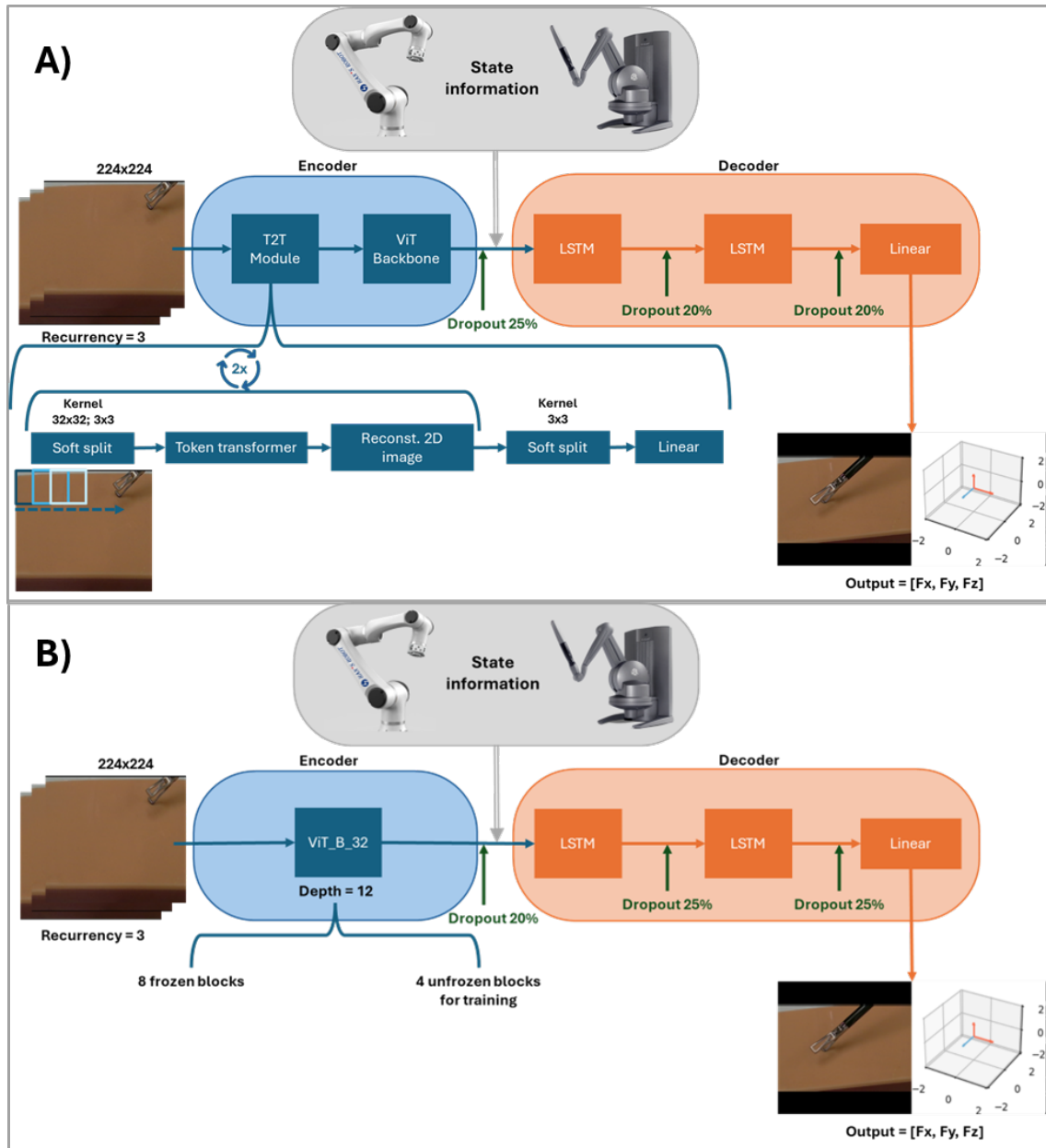


Figure 3.3: A) Architecture for the T2T approach. B) Architecture for the Transfer Learning approach

to the ViT transfer model, the classification token was extracted and used as the feature vector in the remaining pipeline. The kernels initially chosen for the soft splitting in the Tokens-to-Tokens module were 32x32, followed by 3x3 and again 3x3. A dropout layer of 20% was used to avoid overfitting.

The output from both encoders was a feature vector of length 768 for each image processed. Similarly to the work done by Reyzaal et al. [37], the processed images and the feature vectors associated with each were concatenated on the time domain, creating a vector with 3 tensors associated with the processed images and 3 associated with the state information. This full vector was then processed by the decoder.

The decoder used was an LSTM network composed of 2 LSTM blocks with Dropout layers of 25% after each, to avoid overfitting. A linear layer was added to the output of this block, with an output of length 3, corresponding to the 3 Dimensional force prediction vector.

The loss metric utilized was Mean Squared Error, described by equation (3.4), where N is the batch size. This metric mainly penalizes distance from the prediction to the ground truth and, while efficient, a combination with Gradient Difference Loss, as suggested by [17] may also be an interesting approach to tackle.

$$\ell(x, y) = L = \{l_1, \dots, l_N\}^\top, \quad l_n = (x_n - y_n)^2, \quad (3.4)$$

During training, the weight updates were optimized by Adam, also implemented with the Pytorch library [84]. The betas were chosen based on Reyzaal et al.'s implementation [37] as (0.9, 0.999) and the learning rate initially as 1×10^{-3} . The training was done for 100 epochs, with early stopping implemented for when validation loss increased constantly or when training loss stagnated with no improvement on validation loss.

Model inference time was evaluated for both networks, over a test run, discarding the first 10 and last 10 samples, and then averaged over that run. Care was taken to avoid disturbances from other processes running simultaneously. The inference time measured does not take into account pre-processing and only corresponds to the forward pass through the network, as done by Reyzaal et al. [37].

3.3 Haptic Feedback Integration

Once we had the 3D force prediction vector, the next step was the development of a wearable haptic feedback system and the integration of the full system. The approach chosen was a hybrid feedback mode with vibrotactile feedback in an alert configuration, similar to the one proposed by Abiri et al. [72], with an additional visual feedback for the representation of higher degrees of freedom.

3.3.1 Wearable Haptic Components

For the development of a wearable haptic feedback system, there were some requirements: the components had to be miniaturized, to fit in bracelets that could be worn by the surgeon while working on the console, but they also had to enable reliable communication with the image processing tool, in this case, a laptop. To this effect, the chosen microcontroller was the Beetle BLE. This is a 28.8mm X 33.1mm compact development board, that allows communication by Bluetooth Low Energy (BLE). This protocol is convenient for wearable technologies, as it is designed for very low power consumption, while maintaining performance. The requirement of a USB dongle to establish this connection also allows to easily generalize this haptic system to other computers, without the pre-requirement for bluetooth connection. The Beetle was powered by a 3.7V 1100 mAh Li-ion battery and controlled the haptic driver DRV2605L, designed with the expressed purpose of working with haptic vibration motors. In this case, a pancake ERM motor was chosen for its compact design and simple working. By driving this motor with the haptic driver, different vibration modes were available, along with different vibration intensities, fully allowing for a customizable haptic vibration system. The available modes are shown in appendix A.

The components were carefully chosen for the autonomy they needed. As RMIS mean duration can reach about 340 minutes, or 5 hours and 40 minutes [85], it was imperative to take system power consumption into account, as it should clearly be available with no issues throughout the whole procedure, and ideally available for another one without need for charging, or downtime. The estimations for this power consumption are shown in table 3.1. Taking the estimates into account, to last a full 340 minute surgery at the maximum power, the system needs a 416.5 mAh battery.

Table 3.1: Current Consumption Estimate

Component	Current Consumption Estimate
Beetle BLE	10 mA (low power mode – 2 μ A *) [86]
DRV2605L Driver	2.4 – 3.5 mA [87]
Haptic Motor (functioning)	3V – 60 mA; 2V – 40 mA [88]
Total (resting)	13.5 mA
Total (Max power)	73.5 mA
(*) – Even though hardware supports low power mode, the firmware for the Beetle BLE does not, as it hasn't been implemented yet	

The full wearable system was comprised of two bracelets, for comfort and ease of movement, with one being placed on the wrist and the other near the elbow. These will be referred to as the fore bracelet and the back bracelet, respectively. The fore bracelet housed the motor and its driver, with the motor being glued to the textile and the driver being sewn to it, ensuring stability, especially with the motor's vibration on the course of its functioning. The back bracelet housed the beetle, sewn to it, and the battery, powering it. The battery was not sewn to the bracelet but its

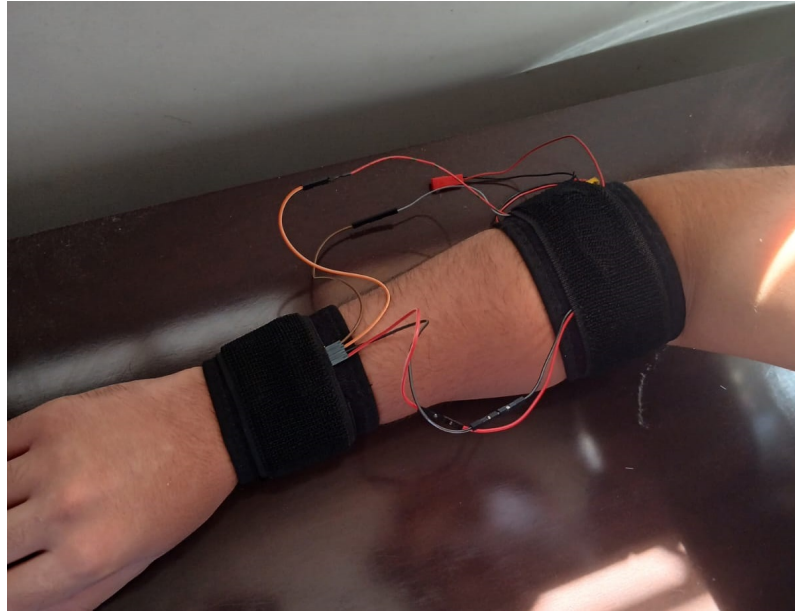


Figure 3.4: Full wearable vibrotactile system developed. The bracelets have a dual strapping mechanism, allowing to protect the electronic components by integrating them between both straps

stability was instead guaranteed by a custom sewn fabric strap, that allowed for its easy removal, for charging or substitution, while also ensuring stability. This system can be seen in figure 3.4.

This system was put through a throughput test to understand the speed of Bluetooth BLE connection between leader and follower. The test was done by the leader constantly sending a package of 200 bytes the fastest possible over 10 seconds, and the follower recording the number of bytes received over that time, then averaging those results. The package chosen was 0xAA due to its alternating pattern between High and Low in binary.

3.3.2 Full live haptic feedback system

To enable the full system, the first step is pairing the BLE USB dongle with the beetle. This is done through AT Commands in Serial communication, as suggested in [89]. In the chosen configuration, the USB dongle acted as the central device with the beetle as the peripheral. Once connection is established, both the dongle and the beetle light up LEDs to signal successful pairing. This is a one-time setup. All subsequent uses will be automatic.

With the force estimation algorithm prepared, as per described in section 3.2, and the haptic feedback components ready, the vibrotactile system could be implemented. The video frame and state data was fed to the estimation algorithm, that waited for 3 time steps before processing it. When these 3 steps are available, they are passed through the network and the output is printed in Serial to the USB dongle COM port.

On the Beetle side, a custom Arduino code parses the data received through Serial and compares it to customizable thresholds. 2 Thresholds for each direction are implemented, T1 and T2. T1 is assumed to be nearing high force levels and T2 is assumed to be exceeding safe levels. These

Table 3.2: Vibration modes applied per Danger level

Danger Level	Vibration Mode	Waveform ID	Waveform Effect
$danger_level < 1$	No vibration	0	None
$1 \leq danger_level < 2$	Low danger	53	Pulsing strong 2 - 60%
$2 \leq danger_level < 4$	Medium danger	66	Transition hum 4 - 40%
$4 \leq danger_level$	High danger	64	Transition hum 2 - 80%

thresholds then contribute to a points-based system. When threshold T1 is exceeded in a direction, a point is added to the variable for *danger_level*. If T2 is exceeded in that direction, an additional 2 points are added, so exceeding level T2 essentially adds 3 points to *danger_level*. The vibration mode is then chosen based on the value at *danger_level*, with 4 modes currently, described in table 3.2. This vibration mode only changes when the danger level leaves the current interval, achieving a continuous alert system. It should be noted that the code developed is highly customizable, with new vibration modes easily added. To make up for the lack of directional feedback in this system, visual representation of the force vectors in each direction is implemented in the real-time simulation.

To test the feasibility of the complete system, a real-time simulation was developed, with the video frames loading from a pre-recorded unseen test run. This simulation plotted the 3-dimensional force estimation vector next to the image, for visual force feedback, and sent this vector to the wearable haptic system through Serial communication allowed by the USB dongle. A framerate chosen as 20 Hz allowed the system to not suffer from the speed bottleneck associated with the force estimation algorithm.

Chapter 4

Results and Discussion

4.1 Force estimation

The training of the force estimation network was done with 30 training scenes from the DaFoEs dataset, balanced with the objective of including the highest representation for each different characteristic: different viewpoints, different materials and different material configuration, as in single or double-layered. The validation dataset, in turn, included 10 unbalanced scenes, chosen at random, as with the test dataset. Due to the complexity of the chosen networks, backpropagation took a long time, with mean epoch time going from 30 to 50 minutes on the mentioned train-test split, depending on the current model. When experimenting with the the same split on the whole dataset of 79 scenes, training times could reach over 1 hour and 10 minutes per epoch, which was not compatible with the time frame for this dissertation. With this, a trade-off was achieved, after testing, effectively halving the number of scenes for training, but minimizing penalties to performance by balancing the dataset, while keeping the training time confined to reasonable intervals.

The initial training of the force estimation networks was plagued by issues related to overfitting to the training data, with most of the early models reporting loss over time that follows a pattern shown in figure 4.1. These consequences were expected, as both transformers and recurrent networks are highly complex and require a lot of data to train properly. However, one of the main objectives of using both T2T and transfer learning approaches was reducing the effect of overfitting on the training of these models, without the need for such an increase in data availability. As such, measures had to be taken to counteract this overfitting tendency.

The first networks trained with a T2T approach used kernels of 32x32 for the first soft split and 3x3 for the other two, with a depth of 2 encoding blocks and a hidden size of 512 for the LSTM decoder. Besides this, initially, 4 LSTM blocks were used as opposed to 2. Coupled with a learning rate of 1×10^{-3} , taken from the literature, these networks showed an erratic validation loss and poor generalization, as can be seen in figure 4.1.

The next two steps in training were the addition of dropout layers to the networks and the reduction of LSTM blocks from 4 to 2. Dropout layers discard a percentage of randomly chosen parameters of the network, stopping it from relying too much on specific aspects of the training

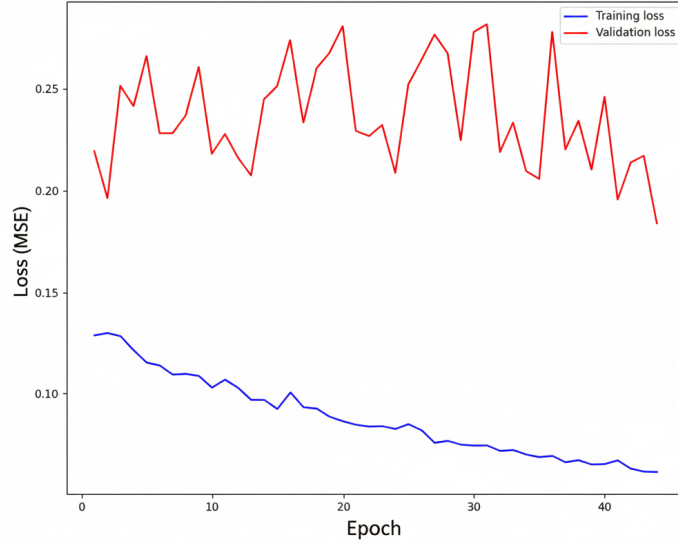


Figure 4.1: Training and validation loss evolution over epochs for the first badly-performing T2T networks. Due to showing no meaningful learning, this model’s training was interrupted

dataset. Even with the steps mentioned, the network was still too complex, as evidenced by the still too high validation loss, when compared with training loss, so the hidden size of the LSTM blocks was reduced gradually over different runs, arriving at 12. This is an exceptionally low number of parameters and its implications will be further discussed. While these steps improved the performance of the model, it was still overfitting early in the pipeline, and the validation loss was erratic. This was a sign that learning rate was too high, with the model overshooting its updated values. As such, the learning rate was reduced from 1×10^{-3} to 5×10^{-4} .

Once dropout layers were introduced and learning rate was reduced, as a way of opposing overfitting, the training of the models showed better generalization, with a much less erratic validation loss, as shown for the best T2T model in figure 4.2.

Between the training of the T2T networks, the transfer learning approach was also being developed. The network used had a kernel of 32x32 for the patch tokenization and its weights had been trained on the ImageNet dataset. It was adapted so the output would be just the class token, as mentioned before, to which the state vector would be concatenated and the result would be submitted to the same image decoder as the T2T, with, at the time, 2 blocks of LSTM and hidden size 512.

For the initial training of the transfer network, learning rate was set as 8×10^{-4} and the last 4 encoding blocks in its 12-depth encoder were unfrozen, since, as mentioned in chapter 3, earlier layers in the models encode baseline relations, while later layers can encode high-level information for specialized tasks. This, however, led to a network that wasn’t learning, shown in figure 4.3, with the training loss reducing, but the validation loss hovering over the same level throughout the whole training.

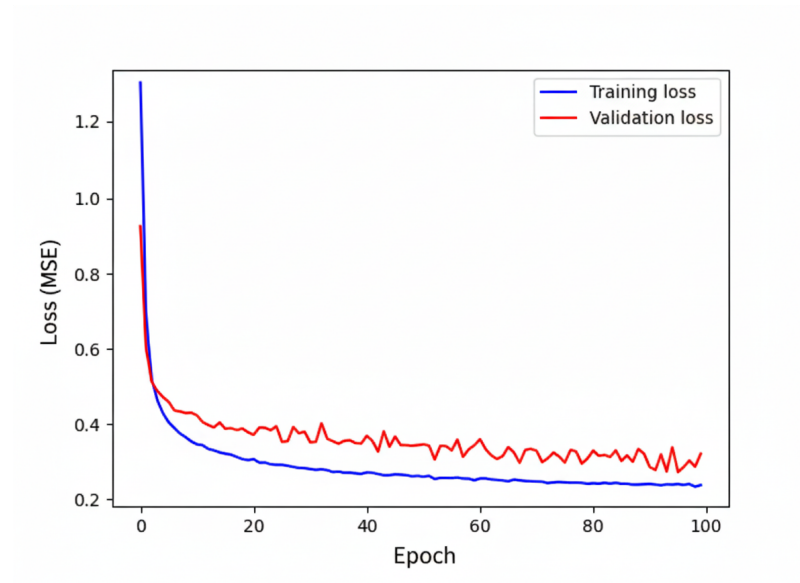


Figure 4.2: Training and validation loss evolution over epochs for the best performing network in the T2T approach

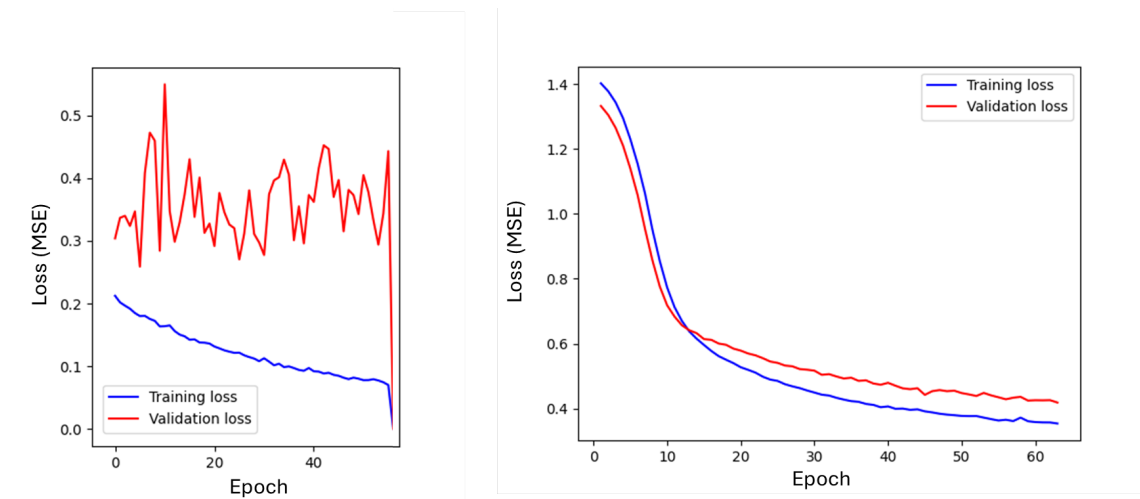


Figure 4.3: Training and validation loss evolution over epochs for the Transfer learning approach. On the left, the results for the first trained network and on the right, the results for the second best performing network of this approach.

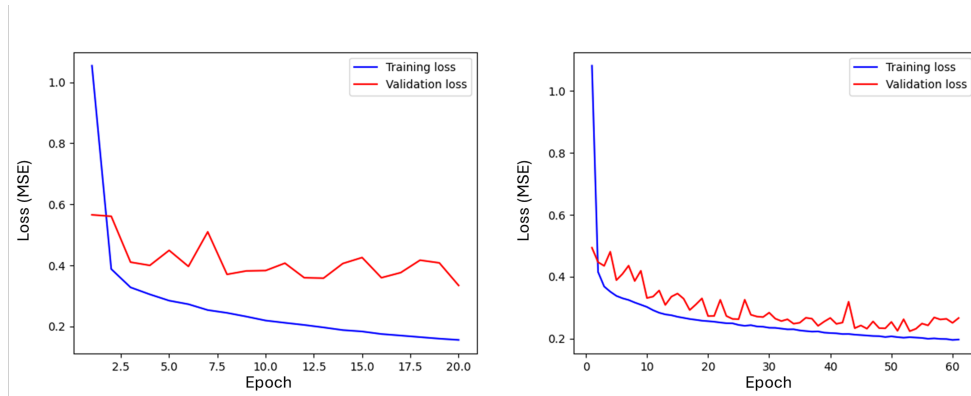


Figure 4.4: Training and validation loss evolution over epochs for both approaches with hidden size 128 for the decoder. On the left, the results for the T2T model and on the right, the results for the Transfer learning model. Due to lack of significant development for the training of the T2T approach, it was interrupted. The Transfer learning training also was interrupted, but only when overfitting began.

This was again a sign of overfitting, so similar steps to the T2T network were taken in this model, insofar as the learning rate was changed from 8×10^{-4} to 1×10^{-4} and the hidden size was also reduced from 512 gradually until 12 produced curves with good behaviour, as shown in figure 4.3. This network also had the 3 final layers unfrozen, as opposed to the 4 that were being used when the overfitting was observed.

Even if good results were achieved for models with hidden size 12, there's an effect at play that should be discussed. Going from tokens of size 768, by default, to a block with hidden size 12 is what can be considered a bottleneck layer. While these can be beneficial to networks, by reducing the dimensionality of the information, and can represent the extraction of the most important aspects for the task, a reduction from dimension 768 to 12 is considered extreme and may instead signal another issue with the task itself. One such possible issue was that the validation size, and in turn the dataset, was too small for the complexity of the model, resulting in instability of the loss results. In order to understand this, the hidden size in both networks was again increased from 12 to 128, while reducing the learning rate further. Even if going from 768 dimensions to 128 still constitutes a bottleneck layer, this is not as extreme as the previous bottleneck, so a good generalization in these networks could still mean that the model is learning important aspects for the task.

The results are shown on figure 4.4 and show that even increasing from 12 to 128 in the hidden size of the decoder is enough to provoke instability in the model's performance, despite reducing the learning rate. That said, this was the network that achieved the best results for the transfer model in the test set, with a meaningful loss evolution. However, this configuration for the T2T model wasn't able to learn meaningful data, with the validation loss hovering around the same values after the initial descent. This suggests that the transfer learning method generalizes better, or in other words, is able to identify the more meaningful patterns in the data, being able to apply it to unseen runs with moderate success.

The best performing networks are shown in figure 4.6. These two networks were compared by their RMSE over test dataset and their absolute deviations from the ground truth over an unseen, randomly chosen test run. The results are shown in figure 4.7. To ease the discussion of the models' performance, the RMSE values for Fx, Fy and Fz estimations in this run, and the respective mean and standard deviations for the absolute deviation of both models are presented on Table 4.1.

In this test run, the T2T network achieves, for strain forces Fx and Fy, lower RMSE, mean absolute deviation and a lower standard deviation, which means that the errors display less variation than the ones for the transfer network. However, looking at normal Fz forces, the transfer network achieves better results for all of these metrics, compared to the T2T. As for RMSE on the full test dataset, the models achieve similar performance, with the Transfer model reaching 0.4158 N and the T2T model 0.4411 N.

When examining the plots in 4.7, it becomes clear that the models routinely overestimate the force Fx, while undershooting the force Fy. The median value for the ground truth of Fy (0.1680) is higher than the value for Fx (-0.2884), while Fy reaches lower minimums than Fx, and Fx higher maximums than Fy. These values may explain the behaviour of the models: since the loss metric used was Mean Squared error, it punishes higher deviations more than low deviations, with a deviation of 2 leading to a loss of $2^2 = 4$, while a deviation of 1 leads to a loss of $1^2 = 1$. This means that, even if a model had a good fit for much of the run, if it didn't model the transition to the peaks correctly, it would be severely punished, leading the training in a more conservative direction, which prioritizes reducing those errors to the high peaks by allowing small deviations from the median values. Even if this explains why the models over and underestimate those values, it doesn't explain why they still overshoot on the mentioned peaks. This may be due to the learning rate and momentum during training. Since these peaks represent a sharp variation of the force values, the training algorithm could associate these with the necessity for a higher weight update than necessary, overshooting the optimal solution at the local minimum. With this, training with an even lower learning rate may yield a better approximation for these values.

As for the normal force Fz, the overall trend is similar to Fx, overestimating the values, even if not as pronounced. Similarly to Fx and Fy, it also shows overshooting in peaks, although, Fz also shows that both models have issues estimating one specific segment, shown in figure 4.5, which may recontextualize the issue with the peak estimation. As may be seen in the plot, the ground truth for Fz during that segment displays a high-frequency variation, rapidly changing between increasing and decreasing values. In similar segments across the other force directions, it is clear that neither network can completely model such variations, and a smoothing is applied instead. However, this can result in a catastrophic error, as the latter part of this segment leads to the peak error for the transfer approach, in this direction. While the explanation presented before for the peak overshooting may explain this behaviour as well, it may signal instead an intrinsic difficulty of the model in keeping track with such rapid variations. In fact, this was discussed in the literature by both Marbán et al. [17] and Reyزابal et al. [37].

Marbán's [17] solution for this issue was the design of a new loss metric, taking into account a mix of Mean Squared Error (MSE) and Gradient Difference Loss (GDL), the formula of which

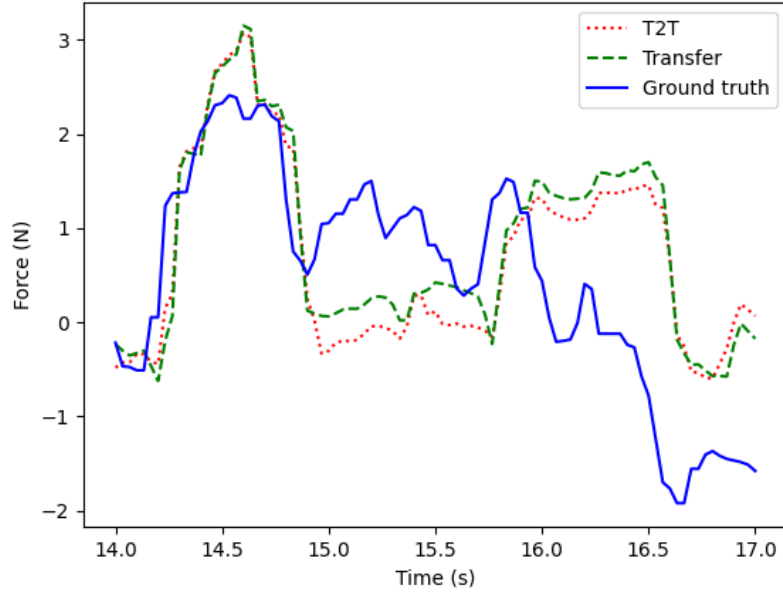


Figure 4.5: Segment taken from figure 4.7, for Fz estimation, for interval from 14 to 17 seconds

is presented in equation (4.1), where N is the number of force dimensions in the prediction, Y represents the ground truth and $\hat{Y}$ the estimation.

$$\mathcal{L}_{GDL} = \sum_{i=1}^{|\mathcal{D}|} \rho(x_i), \quad x_i = \sum_{j=1}^N |Y_i^{(j)} - Y_{i-1}^{(j)}| - |\hat{Y}_i^{(j)} - \hat{Y}_{i-1}^{(j)}| \quad (4.1)$$

In this equation, $\rho(x_i)$ refers to a function that was applied to x in training, and it varied for the training of the encoder and the decoder, being $\ln(x^\gamma + \varepsilon)$ for the fine-tuning of the encoder and x for the training of the decoder, where γ is 1.0 and ε is an infinitesimal to stop the $\ln$ function from being undetermined. It should be noted that $\ln(x^\gamma + \varepsilon)$ was also applied to the mean squared error. Essentially, using GDL as a component of the loss metric along with mean squared error allows the model to optimize for reduced deviations not only on actual value, but also on its first derivative, allowing it to better model sharp changes.

Reyzabal's [37] solution, on the other hand, is the utilization of a window of 5 time steps for the recurrency of their models, stating this window as allowing the network to model acceleration, its second derivative. Effectively, their recurrent models are able to follow complex high frequency variations in the test run. Since one of the main objectives of choosing the 3-time step window for the current work was to address the issues of inference time for Reyzabal's recurrent models without compromising performance, inference time was evaluated.

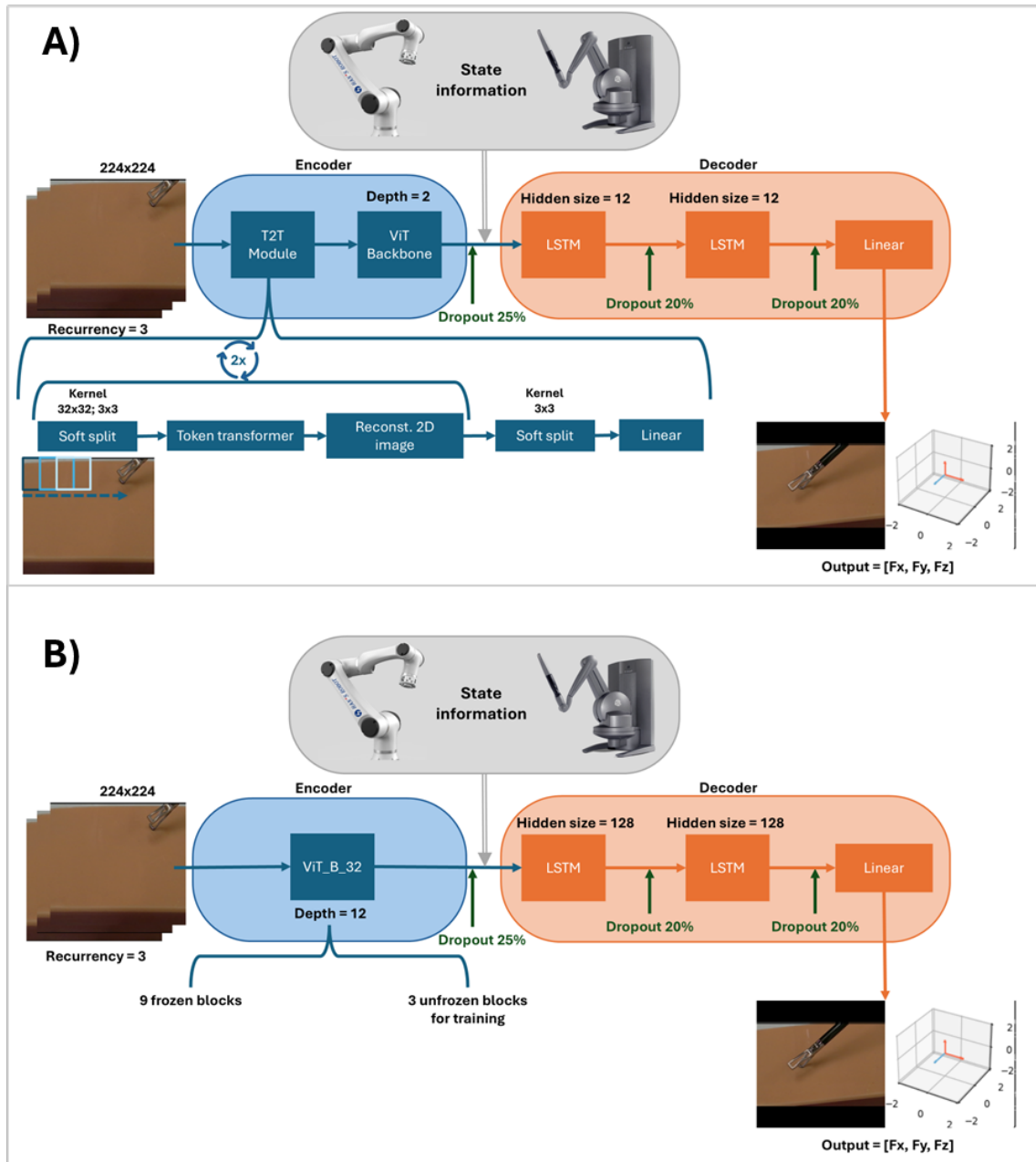


Figure 4.6: A) Best performing model for the T2T approach. B) Best performing model for the Transfer Learning approach

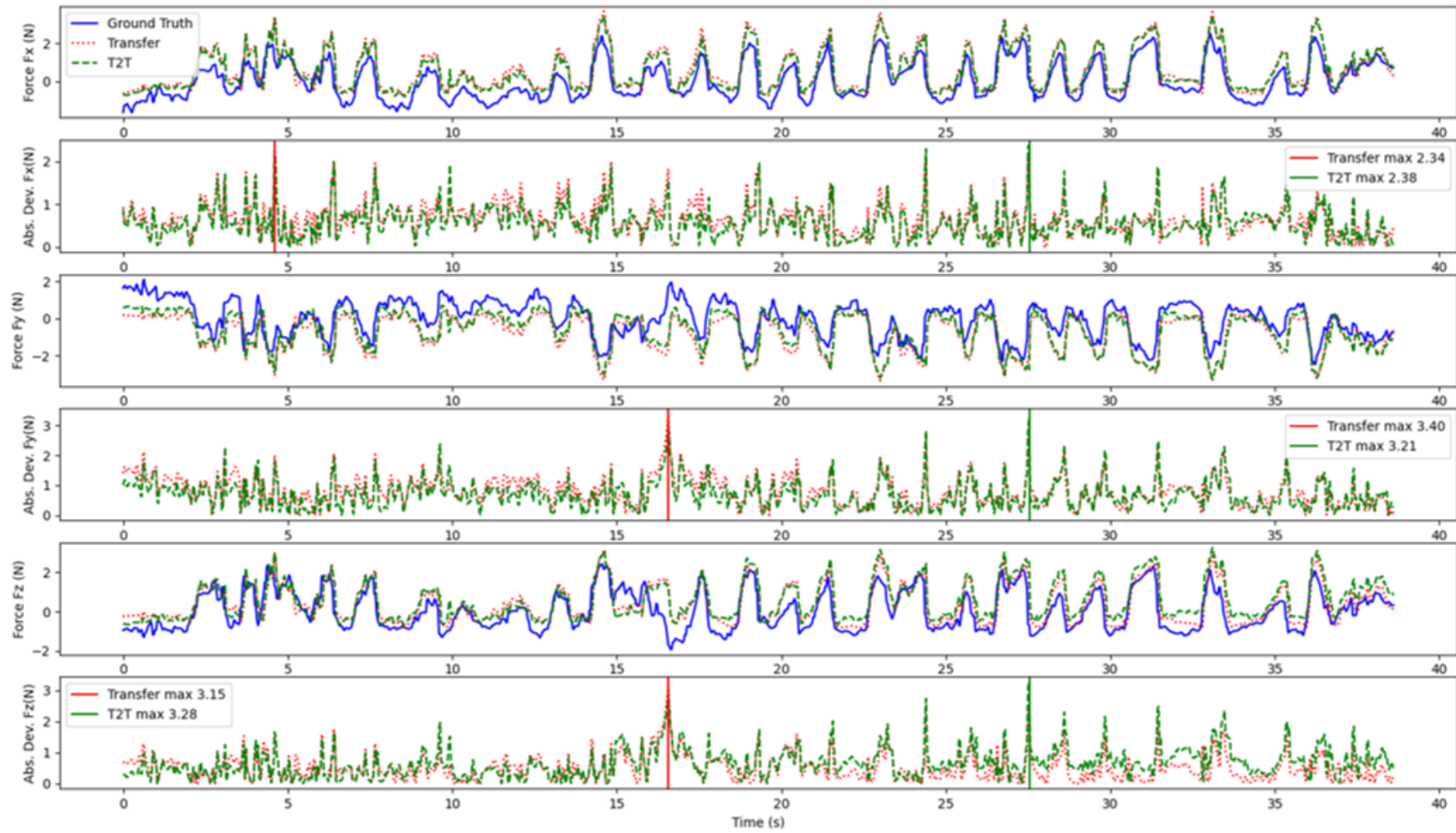


Figure 4.7: Results obtained for the comparison of the model's predictions to the ground truth over an unseen test run.

Table 4.1: Comparison of Transfer and T2T Models for RMSE and Mean and Standard Deviation of the Absolute deviation for directions Fx, Fy and Fz

Metric			Transfer	T2T
Fx	RMSE		0.7435	0.6741
	Abs dev.	Mean	0.6455	0.5714
		Std.	0.4494	0.4054
Fy	RMSE		0.9437	0.8330
	Abs dev.	Mean	0.8118	0.6898
		Std.	0.5659	0.5596
Fz	RMSE		0.6594	0.8028
	Abs dev.	Mean	0.5091	0.6564
		Std.	0.5475	0.5925

Inference rate was calculated as described in section 3.2.2. The T2T model achieved 105.58 Hz, while the Transfer model achieved 66.31 Hz. When comparing these values to the ones presented in [37] for the recurrent models, the inference frequencies for the current work are better suited to the needs of live haptic feedback systems. However, as discussed, this increase in inference time was achieved at the cost of performance, which is undesirable for such systems. Even so, since the models implement changes to the networks presented in Reyزابal’s work [37], in the future, their performance and inference time with 5 time steps should also be addressed, with the objective of finding the best trade-off between inference rate and model performance.

Due to the discussed errors, the models developed in this work are still far from being ready for implementation in live haptic feedback systems for RMIS. Even the lowest mean absolute deviation of 0.5091, for the estimation of Fz with the transfer model, can constitute about 12% of the full interval of roughly -2 N to 2 N to which the forces are confined, which is too high for a reliable force sensitivity in RMIS. For example, in the current haptic implementation, with the threshold T1 at 1 N and the threshold T2 at 2 N, a ground truth force Fz of 1.5 N could be estimated by this model as 0.99 N, under T1, and as such, provoking no response, or 2.01, over T2, being classified as highly dangerous, severely misguiding the operator and thus leading to a higher risk of tissue damage for the patient.

Taking all of the results into account, none of the present networks achieve a clear advantage over the other. Even if the T2T network has faster inference than the Transfer network, the recording is done at 30Hz, so any inference rate that manages to be above this value is enough to not bottleneck the system. In cases where the inference rate of the whole pipeline, including pre-processing, is lower than 30 Hz, a frame-skipping approach can be implemented, where every other frame is used for the prediction. Even so, the impact of this approach in the performance of the model should be evaluated.

As for prediction accuracy, both models achieve similar RMSE on the test set, with the Transfer model reaching 0.4158 N and the T2T model 0.4411 N. Despite being similar, the transfer model seems to more accurately generalize to unseen runs. This is a promising result for the future, and intuitive, as transfer learning was developed with the objective of leveraging previous

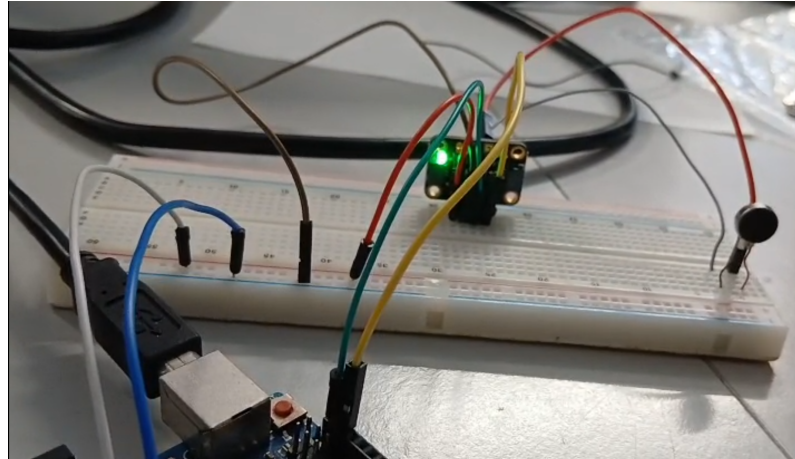


Figure 4.8: Prototyping of the DRV2605L driver and the haptic motor

learned large amounts of data and apply it to a new task. Going further in this direction, in future work, applying transfer learning with the T2T model might yield even better results, combining both approaches from this work. This can be done with the open-access models provided in the original source [64] and may be a good solution for the issue of lack of data for force estimation in RMIS. Another approach for dealing with this lack of data is to utilize the dataset mixing and data augmentation procedures proposed in [37], as the models trained with this modality achieved promising results for two distinct datasets, achieving overall lower errors when compared to models trained on a single dataset and evaluated on the other.

Besides implementing data augmentation and dataset mixing, and implementing transfer learning with T2T, future work on further improving the force estimation techniques developed in this dissertation includes the implementation of the mixed GDL and MSE loss proposed by [17], as a compromise of optimizing for errors in both force estimation and its first derivative to allow for high frequency modeling. Training the networks with a lower learning rate may also improve these results, especially considering the overshooting at the peaks presented in figure 4.7. Increasing the time steps to 5 should also be implemented to evaluate if it improves performance on complex statements without compromising inference time for these networks. With that, the current networks still represent a step in the right direction, showing promising results and interesting venues to explore in future developments for sensor-less force estimation in RMIS.

4.2 Haptic system

The developed haptic feedback system was achieved with off-the-shelf components and its implementation was simple. Firstly, the DRV2605L driver and motor's settings were prototyped with an Arduino and solderless breadboard, to better understand the vibrational modes and the library included with the driver. This prototyping is shown in figure 4.8. After understanding the control of the driver and the motor, the beetle's connection to the bluetooth USB dongle was set-up with AT commands, after which, the system would be ready for the wearable integration.

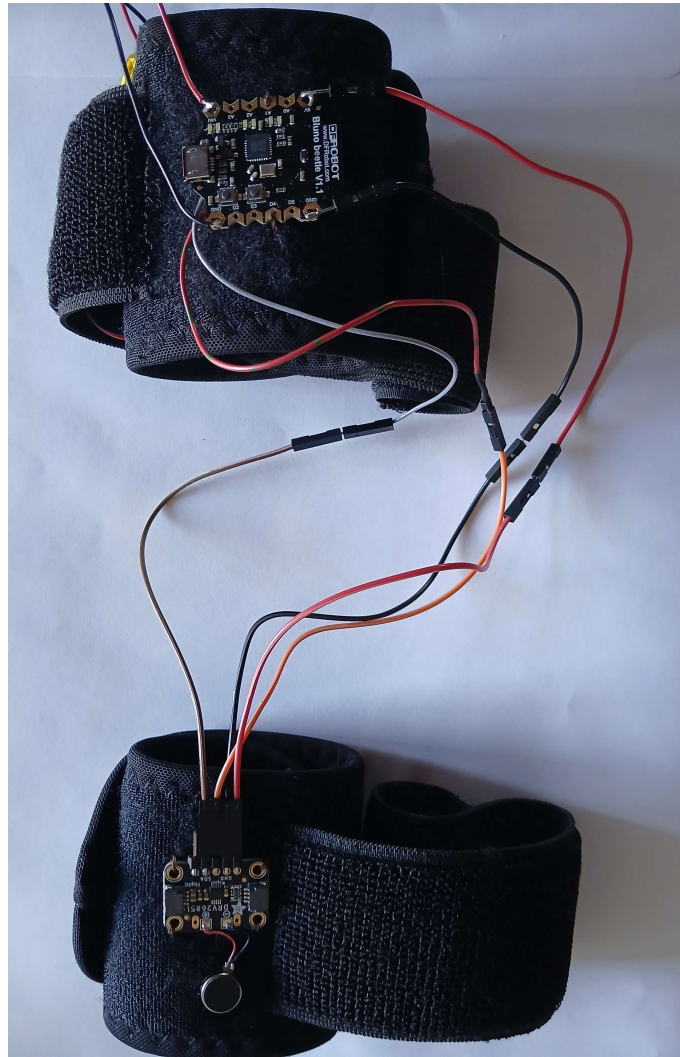


Figure 4.9: Components integrated in the wearable bracelets. On top, the back bracelet, with the Beetle and battery (not shown). On the bottom, the fore bracelet with the driver and haptic motor

The haptic feedback system, as described in chapter 3, was composed of two separate bracelets, each with their separate components, shown in figure 4.9. These components, with the exception of the haptic motor and the battery, were sewn to the bracelets, with the stitches represented in figure 4.10. While the haptic motor was glued to the fore bracelet, the battery was held in place by a cloth strap (figure 4.11). With this, the battery could easily be replaced for charging, substitution or for safekeeping when out of use, as mentioned in chapter 3.

In this wearable bracelet system, one concern was the difficulty in vibration transmitting over the textile bracelet. To account for this, the vibration modes selected had generally high amplitude. However, this may result in lack of discernibility between different modes and danger levels, which would result in even less degrees of freedom for the system. With this, the vibration mode chosen for danger level 1 differs from the other modes by pulsing instead of humming. This approach can be further escalated to differ between multiple danger levels.

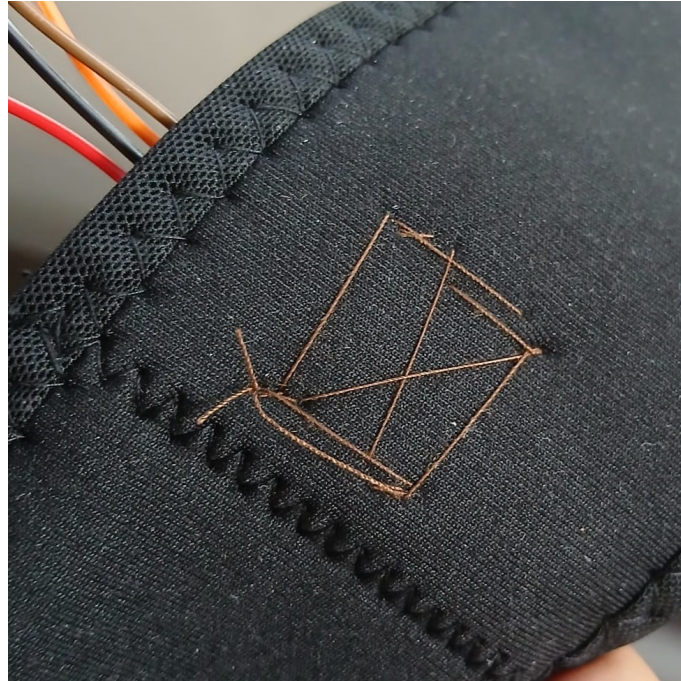


Figure 4.10: Photograph of the stitching on the back of the fore bracelet, for stabilization of the DRV2605L haptic driver



Figure 4.11: Demonstration of the sewn clothstrap, stabilizing the battery in the bracelet.

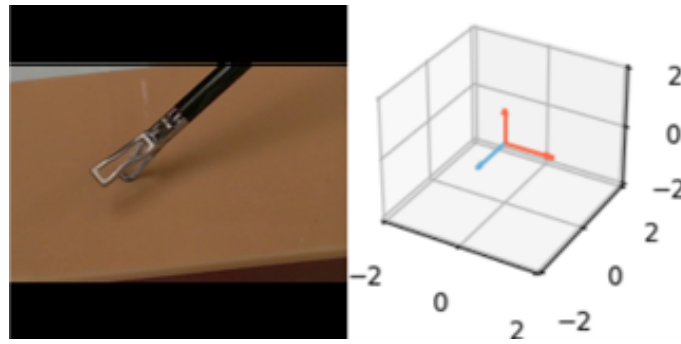


Figure 4.12: Visual feedback system developed for the live implementation. On the left side is represented the current image being processed and on the right side, the 3-dimensional force vector plot corresponding to the model's estimation.

The other important step for this system's transparency was the communication throughput between the platform and the processing system, the laptop. For this, the throughput test described in section 3.3 yielded the result of 1.84 KB/s. Taking into account that each force measurement outputs to serial a string with the form "< X.XX Y.YY Z.ZZ >", whose length is 18 Bytes (1 for each character, due to UTF-8 being used as the encoding), this means that the system has a maximum transfer rate of $1.84/0.018 = 102 \text{ Hz}$. When considering the wider system, and the inference rate measured in section 4.1, this information transfer rate is high enough to not bottleneck the system, and enables transparency by allowing for high responsiveness.

As for the visual feedback, this was done with the cv2 library on python. A custom function was built that received the predicted force vector and from it created 3 vectors to display in a 3D force component plot. This function also included the T1 and T2 thresholds that were used in the Beetle code for the danger level evaluation, and when each of the vectors surpassed these levels, its color would intensify. This had the objective of allowing the operator to relate the current vibration mode to the danger levels measured by the code for each direction, restoring some of the degrees of freedom lost with the choice of vibrotactile actuation. This plot is shown on the right side of figure 4.12, from the live feedback implementation, with the left side being the current image being processed, simulating the real-time monocular video acquisition. To further enhance the simulation, the image would only update with the frame rate, a pre-defined value.

The vibrotactile feedback system can easily be integrated with any RMIS system that outputs a force vector, as it constitutes a separate loop wearable and wireless feedback approach that conveys alert information to the operator. This means it can be integrated with RMIS systems with force vector estimation algorithms or with direct force sensing. Besides, the usefulness of these alert systems was discussed in [72], with an highlight being its integration with learning platforms for novice surgeons with RMIS technology. The current system can be utilized without concerns for mobility, thanks to its double bracelet wireless approach, shown in figure 4.13, and autonomy, due to its low power consumption, and the integration of the USB Bluetooth dongle allows any computer to communicate with it, being easily accessible to most available technology. It can even be combined with different haptic feedback modalities thanks to its independent actuation.

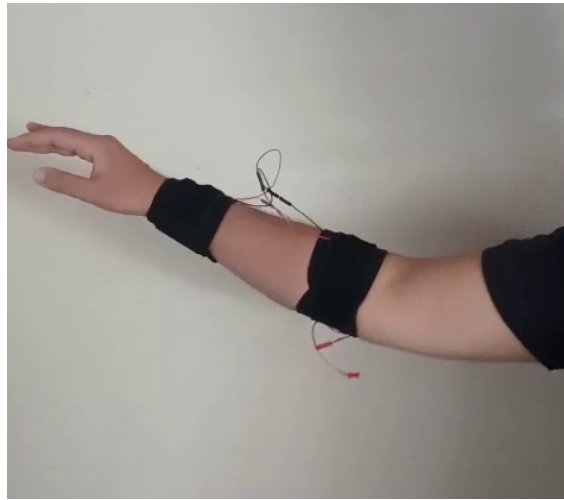


Figure 4.13: Snapshot from mobility demonstration, available at [90]

While the benefits of similar systems are well-supported in the literature, the one developed in the current work merely constitutes a proof-of-concept and would need a future critical evaluation to further fine-tune its integration. Mainly, three aspects should be understood: firstly, the distinctness of the vibrational modes' feel when operating a RMIS system should be evaluated, as the operator should be able to easily discern between danger levels and the current modes were chosen based on subjective testing on the author's sensitivity; secondly, the intuitiveness of the vibrational feedback should be quantified, with other design approaches available, such as intermittent clicking, as opposed to the current continuous vibration; thirdly, the system's comfort should be quantified in long-term use, since the mean time of RMIS procedures can reach 340 minutes, as mentioned in chapter 3. These measures are considered crucial to evaluate fully the integration of the current haptic system in the wider RMIS system.

That said, the system's current issues should be addressed. Firstly, the user needs assistance to equip or remove it, due to its position in one of the forearms. This could be solved with a change in design from strap bracelets to a wearable sleeve or another approach that may be slipped on the arm and allow the system to be set-up by the user alone. Then, as stated before, the system's lack of degrees of freedom impedes it from working as the sole haptic feedback approach without loss of performance. Even the current approach with visual feedback as the main sensory substitute is imperfect, since the visual medium is prone to sensory overload in RMIS systems. As such, taking the disadvantages into account, a future system should include another kind of feedback modality, such as the pneumatic approaches discussed in [72] or an electrical approach, taking advantage of the miniaturization of connected microcontrollers like the beetle BLE. These could help achieve better performance and reduce the degree of danger for the patient by providing haptic information in a more meaningful and ergonomic way to the operator. The system should also be evaluated with the end-users to better understand the conditions it's subject to and the features the users prioritize, as well as understand design flaws when implementing it with the RMIS platform.

Chapter 5

Conclusions

The research developed in this work successfully implemented a proof-of-concept haptic feedback system for RMIS procedures.

An external 3-dimensional force estimator algorithm based on sensor-less force estimation from monocular video recording and manipulator state data was achieved. Two approaches were evaluated for this, differing in their image encoders and decoders, with the objective of placating issues mentioned in the literature, regarding lack of training data for such estimation tasks: the first was based on the training from scratch of a Tokens-to-tokens Vision Transformer encoder with two blocks of 12 hidden size LSTM decoder for 3 time steps, while the second focused on the transfer learning and fine-tuning of a Vision Transformer network pre-trained on the ImageNet dataset for its encoder, and a 128 hidden size 2-block LSTM decoder also for 3 time steps. These achieved Root Mean Squared Errors in the test set of 0.4411 N and 0.4158 N respectively, while achieving inference rates of 105.58 Hz and 66.31 Hz, trading off performance for faster inference, when comparing to results from the literature [37]. This error is still not enough for these networks to be safely implemented in RMIS systems, with the results showing that the models can misguide the operators by over or underestimating the values. Besides this, the models show issues in modeling high-frequency variations, smoothing over those results and further misleading the operator. That said, the results presented still achieved a low error and represent promise for further improvements to achieve the objective of ultimately leading to a lesser degree of danger for the patient by reducing excessive tissue damage. Future improvements in these approaches include the integration of both modalities, by applying transfer learning to the T2T approach, as well as the implementation of a mixed loss metric presented previously by Márban et al. [17] that punishes both force estimation errors and gradient differences, by combining Mean Squared Error with Gradient Difference loss. The training of these networks can be further improved by reducing parameters such as learning rate and momentum, allowing the models to achieve better optimal solutions. And lastly, 5 time steps should be evaluated to understand if higher frequency modeling can be achieved while maintaining high inference rate.

The haptic feedback system developed in this work constitutes a successful proof-of-concept, incorporating a wearable wireless design with off-the-shelf components. It makes use of a hybrid

visual-vibrotactile feedback modality to enhance the limited degrees of freedom from vibrotactile approaches and further enhance the performance of the system. The chosen components allow for high autonomy, with the design allowing for easy substitution or removal of the battery for charging. The current system provides a customizable implementation, with a points-based danger level approach, where, in surpassing one of two customizable thresholds T1 or T2, the force in each direction provokes an increase in a variable that defines the vibrational feedback to be conveyed to the operator. Even so, the evaluation of this system was subjective and critical evaluation metrics of its performance should be designed and applied in the future to further improve on its qualities and placate its issues, with characteristics such as comfort of use over long periods of time being crucial. Mainly, the visual feedback modality is suboptimal due to sensory overload for the operator, with vision being the medium by which they guide themselves on RMIS systems. Besides that, the double bracelet design necessitates assistance to be set-up on the forearm of the operator, which is a disadvantage for the ergonomics of the design. This design should be discussed with end-users and tried out with the RMIS platforms to better understand other design failures, as well as potential features to further improve the design. In future work, a different kind of second modality from visual should be explored, such as pneumatic actuation proposed in the literature [72].

After the implementation of its core parts, a live haptic open-loop simulation was successfully developed for trying out the proposed haptic system with the developed force estimation systems. In the future, these systems should be implemented in a RMIS platform, allowing the closure of the loop and further testing of its transparency and stability.

This dissertation successfully developed a proof-of-concept for a haptic feedback system in RMIS, demonstrating the feasibility of an external, wireless, and wearable solution for conveying information from sensorless force estimation algorithms, which despite their current limitations, offer a promising foundation for a future where haptic feedback can effectively restore the sense of touch to surgeons. By addressing the identified challenges, such as improving model accuracy, mitigating sensory overload, and enhancing the system's ergonomics, the system's transparency and stability can be improved, working towards a reduction in patient tissue damage in robot-assisted minimally invasive surgery, and thus an improvement in treatment quality and safety of the surgical environment. This work represents a significant step forward, building on concepts proposed in recent literature and advancing them, with the objective of contributing to continued research and development in the expansive area of haptic feedback for RMIS platforms. The next steps for this system would be the implementation of the mixed loss metric along with transfer learning for T2T-ViT, using data augmentation to increase the size of the training dataset, in the case of the force estimation algorithm. The haptic feedback system would need the implementation of another modality to serve as the visual feedback substitution, and a critical metric evaluation among the end-users to understand its strengths and weaknesses. In the future, a system such as the one developed in this dissertation may be implemented along with other feedback modalities to increase transparency and stability in RMIS systems to the point of restoring the touch sense to the operator and being able to take full advantage of robotic systems without risking patient safety.

References

- [1] Giulio Dagnino and Dennis Kundrat. “Robot-Assistive Minimally Invasive Surgery: Trends and Future Directions”. In: *International Journal of Intelligent Robotics and Applications* 8.4 (Dec. 1, 2024), pp. 812–826. ISSN: 2366-598X. DOI: [10.1007/s41315-024-00341-2](https://doi.org/10.1007/s41315-024-00341-2). URL: <https://doi.org/10.1007/s41315-024-00341-2>.
- [2] Gabriel Pokorny et al. “Minimally Invasive versus Open Surgery for Degenerative Lumbar Pathologies: A Systematic Review and Meta-Analysis”. In: *European Spine Journal* 31.10 (Oct. 1, 2022), pp. 2502–2526. ISSN: 1432-0932. DOI: [10.1007/s00586-022-07327-3](https://doi.org/10.1007/s00586-022-07327-3). URL: <https://doi.org/10.1007/s00586-022-07327-3>.
- [3] Rocco Ricciardi et al. “The COMPARE Study: Comparing Perioperative Outcomes of Oncologic Minimally Invasive Laparoscopic, Da Vinci Robotic, and Open Procedures: A Systematic Review and Meta-analysis of the Evidence”. In: *Annals of Surgery* 281.5 (May 2025), p. 748. ISSN: 0003-4932. DOI: [10.1097/SLA.0000000000006572](https://doi.org/10.1097/SLA.0000000000006572). URL: https://journals.lww.com/annalsofsurgery/fulltext/2025/05000/the_compare_study_comparing_perioperative.8.aspx.
- [4] Zhenxuan Shao et al. “Comparison of Open Surgery Versus Minimally Invasive Surgery in Nonsevere Adult Degenerative Scoliosis: A Systematic Review and Meta-Analysis”. In: *Spine* 49.14 (July 15, 2024), E210. ISSN: 0362-2436. DOI: [10.1097/BRS.0000000000005011](https://doi.org/10.1097/BRS.0000000000005011). URL: https://journals.lww.com/spinejournal/abstract/2024/07150/comparison_of_open_surgery_versus_minimally.10.aspx.
- [5] F. Pugin, P. Bucher, and P. Morel. “History of Robotic Surgery : From AESOP® and ZEUS® to Da Vinci®”. In: *Journal of Visceral Surgery*. Robotic Surgery 148 (5, Supplement Oct. 1, 2011), e3–e8. ISSN: 1878-7886. DOI: [10.1016/j.jviscsurg.2011.04.007](https://doi.org/10.1016/j.jviscsurg.2011.04.007). URL: <https://www.sciencedirect.com/science/article/pii/S1878788611000324>.
- [6] J C Bowersox. “Telepresence Surgery”. In: *Journal of British Surgery* 83.4 (Apr. 1, 1996), pp. 433–434. ISSN: 0007-1323, 1365-2168. DOI: [10.1002/bjs.1800830402](https://doi.org/10.1002/bjs.1800830402). URL: <https://academic.oup.com/bjs/article/83/4/433/6166795>.
- [7] Jay Shah, Arpita Vyas, and Dinesh Vyas. “The History of Robotics in Surgical Specialties”. In: *American journal of robotic surgery* 1.1 (June 1, 2014), pp. 12–20. ISSN: 2374-0612. DOI: [10.1166/ajrs.2014.1006](https://doi.org/10.1166/ajrs.2014.1006). PMID: 26677459. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4677089/>.
- [8] Blenda Johansson et al. “Robotic Surgery: Review on Minimally Invasive Techniques”. In: *Fusion of Multidisciplinary Research, An International Journal* 2.2 (July 14, 2021), pp. 201–210. ISSN: 3051-0155. DOI: [10.63995/GQNC2594](https://doi.org/10.63995/GQNC2594). URL: <https://fusionproceedings.com/fmr/1/article/view/29>.

- [9] Tamás Haidegger et al. “Robot-Assisted Minimally Invasive Surgery—Surgical Robotics in the Data Age”. In: *Proceedings of the IEEE* 110.7 (July 2022), pp. 835–846. ISSN: 1558-2256. DOI: 10.1109/JPROC.2022.3180350. URL: <https://ieeexplore.ieee.org/abstract/document/9805581>.
- [10] Khurshid A. Guru et al. “The Learning Curve for Robot-Assisted Radical Cystectomy”. In: *JSLs : Journal of the Society of Laparoendoscopic Surgeons* 13.4 (2009), pp. 509–514. ISSN: 1086-8089. DOI: 10.4293/108680809X12589998404128. PMID: 20042129. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3030784/>.
- [11] *Company History | Robotic Assisted Surgery | Intuitive*. URL: <https://www.intuitive.com/en-us/about-us/company/history> (visited on 12/08/2021).
- [12] Michael M. Awad et al. “Evaluation of Forces Applied to Tissues during Robotic-Assisted Surgical Tasks Using a Novel Force Feedback Technology”. In: *Surgical Endoscopy* 38.10 (Oct. 1, 2024), pp. 6193–6202. ISSN: 1432-2218. DOI: 10.1007/s00464-024-11131-z. URL: <https://doi.org/10.1007/s00464-024-11131-z>.
- [13] O. A. J. van der Meijden and M. P. Schijven. “The Value of Haptic Feedback in Conventional and Robot-Assisted Minimal Invasive Surgery and Virtual Reality Training: A Current Review”. In: *Surgical Endoscopy* 23.6 (June 1, 2009), pp. 1180–1190. ISSN: 1432-2218. DOI: 10.1007/s00464-008-0298-x. URL: <https://doi.org/10.1007/s00464-008-0298-x>.
- [14] Arturo Marbán et al. “Haptic Feedback in Surgical Robotics: Still a Challenge”. In: *ROBOT2013: First Iberian Robotics Conference: Advances in Robotics, Vol. 1*. Ed. by Manuel A. Armada, Alberto Sanfeliu, and Manuel Ferre. Cham: Springer International Publishing, 2014, pp. 245–253. ISBN: 978-3-319-03413-3. DOI: 10.1007/978-3-319-03413-3_18. URL: https://doi.org/10.1007/978-3-319-03413-3_18.
- [15] Junchang Zhang et al. “Force Sensing and Force Control of Flexible Gripper with Integrated Flexible Strain and Tactile Sensors for Strawberry Non-Destructive Gripping and Freshness Grading”. In: *Food and Bioprocess Technology* 18.6 (June 1, 2025), pp. 5700–5721. ISSN: 1935-5149. DOI: 10.1007/s11947-025-03795-9. URL: <https://doi.org/10.1007/s11947-025-03795-9>.
- [16] Petros I. Kaltsas, Panagiotis N. Koustoumpardis, and Pantelis G. Nikolakopoulos. “A Review of Sensors Used on Fabric-Handling Robots”. In: *Machines* 10.2 (Feb. 2022), p. 101. ISSN: 2075-1702. DOI: 10.3390/machines10020101. URL: <https://www.mdpi.com/2075-1702/10/2/101>.
- [17] Arturo Marban et al. “A Recurrent Convolutional Neural Network Approach for Sensorless Force Estimation in Robotic Surgery”. In: *Biomedical Signal Processing and Control* 50 (Apr. 1, 2019), pp. 134–150. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2019.01.011. URL: <https://www.sciencedirect.com/science/article/pii/S1746809419300114>.
- [18] Stamatia Giannarou et al. “Vision-Based Deformation Recovery for Intraoperative Force Estimation of Tool–Tissue Interaction for Neurosurgery”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.6 (June 1, 2016), pp. 929–936. ISSN: 1861-6429. DOI: 10.1007/s11548-016-1361-z. URL: <https://doi.org/10.1007/s11548-016-1361-z>.

- [19] Yupeng Hao et al. “Development of Force Sensing Techniques for Robot-Assisted Laparoscopic Surgery: A Review”. In: *IEEE Transactions on Medical Robotics and Bionics* 6.3 (Aug. 2024), pp. 868–887. ISSN: 2576-3202. DOI: [10.1109/TMRB.2024.3407238](https://doi.org/10.1109/TMRB.2024.3407238). URL: <https://ieeexplore.ieee.org/document/10542442>.
- [20] Amir Hossein Hadi Hosseinabadi and Septimiu E. Salcudean. “Force Sensing in Robot-Assisted Keyhole Endoscopy: A Systematic Survey”. In: *The International Journal of Robotics Research* 41.2 (Feb. 1, 2022), pp. 136–162. ISSN: 0278-3649. DOI: [10.1177/02783649211052067](https://doi.org/10.1177/02783649211052067). URL: <https://doi.org/10.1177/02783649211052067>.
- [21] Lingtao Yu, Wenjie Wang, and Fengfeng Zhang. “External Force Sensing Based on Cable Tension Changes in Minimally Invasive Surgical Micromanipulators”. In: *IEEE Access* 6 (2018), pp. 5362–5373. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2017.2788498](https://doi.org/10.1109/ACCESS.2017.2788498). URL: <https://ieeexplore.ieee.org/abstract/document/8253606>.
- [22] Ana Luisa Trejos et al. “Design and Evaluation of a Sterilizable Force Sensing Instrument for Minimally Invasive Surgery”. In: *IEEE Sensors Journal* 17.13 (July 2017), pp. 3983–3993. ISSN: 1558-1748. DOI: [10.1109/JSEN.2017.2703883](https://doi.org/10.1109/JSEN.2017.2703883). URL: <https://ieeexplore.ieee.org/abstract/document/7927399>.
- [23] Hyundo Choi, Yoan Lim, and Junhyung Kim. “Three-Axis Force Sensor with Fiber Bragg Grating”. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). July 2017, pp. 3940–3943. DOI: [10.1109/EMBC.2017.8037718](https://doi.org/10.1109/EMBC.2017.8037718). URL: <https://ieeexplore.ieee.org/abstract/document/8037718>.
- [24] Andreas Othonos. “Fiber Bragg Gratings”. In: *Review of Scientific Instruments* 68.12 (Dec. 1, 1997), pp. 4309–4341. ISSN: 0034-6748. DOI: [10.1063/1.1148392](https://doi.org/10.1063/1.1148392). URL: <https://doi.org/10.1063/1.1148392>.
- [25] Changhu Lv, Shuxin Wang, and Chaoyang Shi. “A High-Precision and Miniature Fiber Bragg Grating-Based Force Sensor for Tissue Palpation During Minimally Invasive Surgery”. In: *Annals of Biomedical Engineering* 48.2 (Feb. 1, 2020), pp. 669–681. ISSN: 1573-9686. DOI: [10.1007/s10439-019-02388-w](https://doi.org/10.1007/s10439-019-02388-w). URL: <https://doi.org/10.1007/s10439-019-02388-w>.
- [26] Daniel S. Yurkewich et al. “Low-Cost Force-Sensing Arthroscopic Tool Using Threaded Fiber Bragg Grating Sensors”. In: *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*. 5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics. Aug. 2014, pp. 28–33. DOI: [10.1109/BIOROB.2014.6913747](https://doi.org/10.1109/BIOROB.2014.6913747). URL: <https://ieeexplore.ieee.org/abstract/document/6913747>.
- [27] Naghmeh Bandari, Javad Dargahi, and Muthukumaran Packirisamy. “Tactile Sensors for Minimally Invasive Surgery: A Review of the State-of-the-Art, Applications, and Perspectives”. In: *IEEE Access* 8 (2020), pp. 7682–7708. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2019.2962636](https://doi.org/10.1109/ACCESS.2019.2962636). URL: <https://ieeexplore.ieee.org/abstract/document/8943415>.
- [28] Edin Terzic et al. “Capacitive Sensing Technology”. In: *A Neural Network Approach to Fluid Quantity Measurement in Dynamic Environments*. Ed. by Edin Terzic et al. London: Springer, 2012, pp. 11–37. ISBN: 978-1-4471-4060-3. DOI: [10.1007/978-1-4471-4060-3_2](https://doi.org/10.1007/978-1-4471-4060-3_2). URL: https://doi.org/10.1007/978-1-4471-4060-3_2.

- [29] Manish Kumar Mishra et al. “MEMS Technology: A Review”. In: *Journal of Engineering Research and Reports* (Feb. 18, 2019), pp. 1–24. ISSN: 2582-2926. DOI: 10.9734/jerr/2019/v4i116891. URL: <https://journaljerr.com/index.php/JERR/article/view/76>.
- [30] Huicong Liu et al. “An Integrated 3-D MEMS Force Sensing and Feedback System for Robot-Assisted Minimally Invasive Surgery”. In: *IEEE Sensors Journal* 24.11 (June 2024), pp. 17470–17477. ISSN: 1558-1748. DOI: 10.1109/JSEN.2024.3380594. URL: <https://ieeexplore.ieee.org/abstract/document/10485247>.
- [31] Cheng Hou et al. “A Sensorised Forcep Based on Piezoresistive Force Sensor for Robotic-Assisted Minimally Invasive Surgery”. In: *2021 IEEE 16th International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)*. 2021 IEEE 16th International Conference on Nano/Micro Engineered and Molecular Systems (NEMS). Apr. 2021, pp. 60–63. DOI: 10.1109/NEMS51815.2021.9451326. URL: <https://ieeexplore.ieee.org/abstract/document/9451326>.
- [32] A. S. Fiorillo, C. D. Critello, and S. A. Pullano. “Theory, Technology and Applications of Piezoresistive Sensors: A Review”. In: *Sensors and Actuators A: Physical* 281 (Oct. 1, 2018), pp. 156–175. ISSN: 0924-4247. DOI: 10.1016/j.sna.2018.07.006. URL: <https://www.sciencedirect.com/science/article/pii/S0924424718308434>.
- [33] Siqi Ding et al. “Pressure-Sensitive Behaviors, Mechanisms and Model of Field Assisted Quantum Tunneling Composites”. In: *Polymer* 113 (Mar. 24, 2017), pp. 105–118. ISSN: 0032-3861. DOI: 10.1016/j.polymer.2017.02.058. URL: <https://www.sciencedirect.com/science/article/pii/S0032386117301829>.
- [34] Winnie Kuang, Michael Yip, and Jun Zhang. “Vibration-Based Multi-Axis Force Sensing: Design, Characterization, and Modeling”. In: *IEEE Robotics and Automation Letters* 5.2 (Apr. 2020), pp. 3082–3089. ISSN: 2377-3766. DOI: 10.1109/LRA.2020.2975726. URL: <https://ieeexplore.ieee.org/abstract/document/9006802>.
- [35] Omid Mohareri, Caitlin Schneider, and Septimiu Salcudean. “Bimanual Telerobotic Surgery with Asymmetric Force Feedback: A daVinci® Surgical System Implementation”. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. Sept. 2014, pp. 4272–4277. DOI: 10.1109/IROS.2014.6943165. URL: <https://ieeexplore.ieee.org/document/6943165>.
- [36] Uikyum Kim et al. “A Surgical Palpation Probe With 6-Axis Force/Torque Sensing Capability for Minimally Invasive Surgery”. In: *IEEE Transactions on Industrial Electronics* 65.3 (Mar. 2018), pp. 2755–2765. ISSN: 1557-9948. DOI: 10.1109/TIE.2017.2739681. URL: <https://ieeexplore.ieee.org/abstract/document/8010319>.
- [37] Mikel De Iturrate Reyزابال et al. “DaFoEs: Mixing Datasets Towards the Generalization of Vision-State Deep-Learning Force Estimation in Minimally Invasive Robotic Surgery”. In: *IEEE Robotics and Automation Letters* 9.3 (Mar. 2024), pp. 2527–2534. ISSN: 2377-3766. DOI: 10.1109/LRA.2024.3356984. URL: <https://ieeexplore.ieee.org/document/10410871>.
- [38] Greg Welch. “An Introduction to the Kalman Filter”. In: (1997).

- [39] Yangming Li et al. “Dynamic Modeling of Cable Driven Elongated Surgical Instruments for Sensorless Grip Force Estimation”. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 2016 IEEE International Conference on Robotics and Automation (ICRA). May 2016, pp. 4128–4134. DOI: [10.1109/ICRA.2016.7487605](https://doi.org/10.1109/ICRA.2016.7487605). URL: <https://ieeexplore.ieee.org/abstract/document/7487605>.
- [40] Keping Liu et al. “A Sensorless Force Estimation Algorithm for Robot Based on Model Compensation”. In: *2023 China Automation Congress (CAC)*. 2023 China Automation Congress (CAC). Nov. 2023, pp. 712–716. DOI: [10.1109/CAC59555.2023.10450234](https://doi.org/10.1109/CAC59555.2023.10450234). URL: <https://ieeexplore.ieee.org/document/10450234>.
- [41] Trevor K. Stephens et al. “Conditions for Reliable Grip Force and Jaw Angle Estimation of Da Vinci Surgical Tools”. In: *International Journal of Computer Assisted Radiology and Surgery* 14.1 (Jan. 1, 2019), pp. 117–127. ISSN: 1861-6429. DOI: [10.1007/s11548-018-1866-8](https://doi.org/10.1007/s11548-018-1866-8). URL: <https://doi.org/10.1007/s11548-018-1866-8>.
- [42] Yangming Li and Blake Hannaford. “Gaussian Process Regression for Sensorless Grip Force Estimation of Cable-Driven Elongated Surgical Instruments”. In: *IEEE Robotics and Automation Letters* 2.3 (July 2017), pp. 1312–1319. ISSN: 2377-3766. DOI: [10.1109/LRA.2017.2666420](https://doi.org/10.1109/LRA.2017.2666420). URL: <https://ieeexplore.ieee.org/abstract/document/7847310>.
- [43] A. C. Ugural and Saul K. Fenster. *Advanced Strength and Applied Elasticity*. 4th ed. Upper Saddle River, N.J: Prentice Hall PTR, 2003. 544 pp. ISBN: 978-0-13-047392-9.
- [44] Jonathan C. J. Wei et al. “Allometric Scaling of Skin Thickness, Elasticity, Viscoelasticity to Mass for Micro-Medical Device Translation: From Mice, Rats, Rabbits, Pigs to Humans”. In: *Scientific Reports* 7.1 (Nov. 21, 2017), p. 15885. ISSN: 2045-2322. DOI: [10.1038/s41598-017-15830-7](https://doi.org/10.1038/s41598-017-15830-7). URL: <https://www.nature.com/articles/s41598-017-15830-7>.
- [45] Frederick H. Silver and David L. Christiansen. “Mechanical Properties of Tissues”. In: *Biomaterials Science and Biocompatibility*. Ed. by Frederick H. Silver and David L. Christiansen. New York, NY: Springer, 1999, pp. 187–212. ISBN: 978-1-4612-0557-9. DOI: [10.1007/978-1-4612-0557-9_7](https://doi.org/10.1007/978-1-4612-0557-9_7). URL: https://doi.org/10.1007/978-1-4612-0557-9_7.
- [46] Zonghe Chua, Anthony M. Jarc, and Allison M. Okamura. “Toward Force Estimation in Robot-Assisted Surgery Using Deep Learning with Vision and Robot State”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 2021 IEEE International Conference on Robotics and Automation (ICRA). May 2021, pp. 12335–12341. DOI: [10.1109/ICRA48506.2021.9560945](https://doi.org/10.1109/ICRA48506.2021.9560945). URL: <https://ieeexplore.ieee.org/document/9560945>.
- [47] Di Ding et al. “Vision-Based Contact Force Sensing in Robotic Surgery: A Technical Review”. In: *IEEE Transactions on Medical Robotics and Bionics* 7.3 (Aug. 2025), pp. 828–844. ISSN: 2576-3202. DOI: [10.1109/TMRB.2025.3583117](https://doi.org/10.1109/TMRB.2025.3583117). URL: <https://ieeexplore.ieee.org/document/11051000>.
- [48] Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany. “Supervised Learning”. In: *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*. Ed. by Matthieu Cord and Pádraig Cunningham. Berlin, Heidelberg: Springer, 2008, pp. 21–49. ISBN: 978-3-540-75171-7. DOI: [10.1007/978-3-540-75171-7_2](https://doi.org/10.1007/978-3-540-75171-7_2). URL: https://doi.org/10.1007/978-3-540-75171-7_2.

- [49] Kaiwen Zuo. “Vision-Based Force Estimation Using Deformable Grippers”. Case Western Reserve University, 2024. URL: https://etd.ohiolink.edu/acprod/odb_etd/etd/r/1501/10?clear=10&p10_accession_num=case1718120033936051.
- [50] C. W. Kennedy and J. P. Desai. “A Vision-Based Approach for Estimating Contact Forces: Applications to Robot-Assisted Surgery”. In: *Applied Bionics and Biomechanics* 2.1 (2005), p. 436897. ISSN: 1754-2103. DOI: [10.1533/abbi.2004.0006](https://doi.org/10.1533/abbi.2004.0006). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1533/abbi.2004.0006>.
- [51] Klaus-Jürgen Bathe. “Finite Element Method”. In: *Wiley Encyclopedia of Computer Science and Engineering*. John Wiley & Sons, Ltd, 2008, pp. 1–12. ISBN: 978-0-470-05011-8. DOI: [10.1002/9780470050118.ecse159](https://doi.org/10.1002/9780470050118.ecse159). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470050118.ecse159>.
- [52] Angelica I. Aviles et al. “A Recurrent Neural Network Approach for 3D Vision-Based Force Estimation”. In: *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA). Oct. 2014, pp. 1–6. DOI: [10.1109/IPTA.2014.7001941](https://doi.org/10.1109/IPTA.2014.7001941). URL: <https://ieeexplore.ieee.org/document/7001941/citations>.
- [53] G. R. Kanagachidambaresan et al. “Recurrent Neural Network”. In: *Programming with TensorFlow: Solution for Edge Computing Applications*. Ed. by Kolla Bhanu Prakash and G. R. Kanagachidambaresan. Cham: Springer International Publishing, 2021, pp. 53–61. ISBN: 978-3-030-57077-4. DOI: [10.1007/978-3-030-57077-4_7](https://doi.org/10.1007/978-3-030-57077-4_7). URL: https://doi.org/10.1007/978-3-030-57077-4_7.
- [54] Greg Van Houdt, Carlos Mosquera, and Gonzalo Nápoles. “A Review on the Long Short-Term Memory Model”. In: *Artificial Intelligence Review* 53.8 (Dec. 1, 2020), pp. 5929–5955. ISSN: 1573-7462. DOI: [10.1007/s10462-020-09838-1](https://doi.org/10.1007/s10462-020-09838-1). URL: <https://doi.org/10.1007/s10462-020-09838-1>.
- [55] Klaus Greff et al. “LSTM: A Search Space Odyssey”. In: *IEEE Transactions on Neural Networks and Learning Systems* 28.10 (Oct. 2017), pp. 2222–2232. ISSN: 2162-2388. DOI: [10.1109/TNNLS.2016.2582924](https://doi.org/10.1109/TNNLS.2016.2582924). URL: <https://ieeexplore.ieee.org/abstract/document/7508408>.
- [56] Angelica I. Aviles et al. “Towards Retrieving Force Feedback in Robotic-Assisted Surgery: A Supervised Neuro-Recurrent-Vision Approach”. In: *IEEE Transactions on Haptics* 10.3 (July 2017), pp. 431–443. ISSN: 2329-4051. DOI: [10.1109/TOH.2016.2640289](https://doi.org/10.1109/TOH.2016.2640289). URL: <https://ieeexplore.ieee.org/abstract/document/7784717>.
- [57] Lisa Torrey and Jude Shavlik. “Transfer Learning”. In: *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI Global Scientific Publishing, 2010, pp. 242–264. ISBN: 978-1-60566-766-9. DOI: [10.4018/978-1-60566-766-9.ch011](https://doi.org/10.4018/978-1-60566-766-9.ch011). URL: <https://www.igi-global.com/chapter/transfer-learning/www.igi-global.com/chapter/transfer-learning/36988>.
- [58] Sinno Jialin Pan and Qiang Yang. “A Survey on Transfer Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (Oct. 2010), pp. 1345–1359. ISSN: 1558-2191. DOI: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191). URL: <https://ieeexplore.ieee.org/abstract/document/5288526>.
- [59] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Apr. 10, 2015. DOI: [10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556). arXiv: 1409.1556 [cs]. URL: <http://arxiv.org/abs/1409.1556>. Pre-published.

- [60] Jia Deng et al. “ImageNet: A Large-Scale Hierarchical Image Database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009 IEEE Conference on Computer Vision and Pattern Recognition. June 2009, pp. 248–255. DOI: [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848). URL: <https://ieeexplore.ieee.org/document/5206848>.
- [61] Alexey Dosovitskiy et al. *An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale*. June 3, 2021. DOI: [10.48550/arXiv.2010.11929](https://doi.org/10.48550/arXiv.2010.11929). arXiv: 2010.11929 [cs]. URL: <http://arxiv.org/abs/2010.11929>. Pre-published.
- [62] Y C A Padmanabha Reddy, P Viswanath, and B Eswara Reddy. “Semi-Supervised Learning: A Brief Review”. In: *International Journal of Engineering & Technology* 7.1.8 (Feb. 9, 2018), p. 81. ISSN: 2227-524X. DOI: [10.14419/ijet.v7i1.8.9977](https://doi.org/10.14419/ijet.v7i1.8.9977). URL: <https://www.sciencepubco.com/index.php/ijet/article/view/9977> (visited on 08/26/2025).
- [63] *VisionTransformer — Torchvision Main Documentation*. URL: https://docs.pytorch.org/vision/main/models/vision_transformer.html (visited on 08/26/2025).
- [64] Li Yuan et al. *Tokens-to-Token ViT: Training Vision Transformers from Scratch on ImageNet*. Nov. 30, 2021. DOI: [10.48550/arXiv.2101.11986](https://doi.org/10.48550/arXiv.2101.11986). arXiv: 2101.11986 [cs]. URL: <http://arxiv.org/abs/2101.11986>. Pre-published.
- [65] Skylar Jean Callis. *Tokens-to-Token Vision Transformers, Explained*. Towards Data Science. Feb. 27, 2024. URL: <https://towardsdatascience.com/tokens-to-token-vision-transformers-explained-2fa4e2002daa/> (visited on 08/28/2025).
- [66] Valentina Bartalesi. “Rethinking Contact: The Haptic in the Viral Era”. In: *Aisthesis. Pratiche, linguaggi e saperi dell'estetico* 14.1 (Nov. 19, 2021), pp. 23–36. ISSN: 2035-8466. DOI: [10.36253/Aisthesis-12481](https://doi.org/10.36253/Aisthesis-12481). URL: <https://oajournals.fupress.net/index.php/aisthesis/article/view/12481>.
- [67] Issam El Rassi and Jean-Michel El Rassi. “A Review of Haptic Feedback in Tele-Operated Robotic Surgery”. In: *Journal of Medical Engineering & Technology* 44.5 (July 3, 2020), pp. 247–254. ISSN: 0309-1902. DOI: [10.1080/03091902.2020.1772391](https://doi.org/10.1080/03091902.2020.1772391). PMID: 32573288. URL: <https://doi.org/10.1080/03091902.2020.1772391>.
- [68] Mohsin I. Tiwana, Stephen J. Redmond, and Nigel H. Lovell. “A Review of Tactile Sensing Technologies with Applications in Biomedical Engineering”. In: *Sensors and Actuators A: Physical* 179 (June 1, 2012), pp. 17–31. ISSN: 0924-4247. DOI: [10.1016/j.sna.2012.02.051](https://doi.org/10.1016/j.sna.2012.02.051). URL: <https://www.sciencedirect.com/science/article/pii/S0924424712001641>.
- [69] Elisa Galofaro et al. “Rendering Immersive Haptic Force Feedback via Neuromuscular Electrical Stimulation”. In: *Sensors (Basel, Switzerland)* 22.14 (July 6, 2022), p. 5069. ISSN: 1424-8220. DOI: [10.3390/s22145069](https://doi.org/10.3390/s22145069). PMID: 35890748. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9319132/>.
- [70] Claudio Pacchierotti, Asad Tirmizi, and Domenico Prattichizzo. “Improving Transparency in Teleoperation by Means of Cutaneous Tactile Force Feedback”. In: *ACM Trans. Appl. Percept.* 11.1 (Apr. 1, 2014), 4:1–4:16. ISSN: 1544-3558. DOI: [10.1145/2604969](https://doi.org/10.1145/2604969). URL: <https://dl.acm.org/doi/10.1145/2604969>.
- [71] Gordon M. Mair. “Towards Transparent Telepresence”. In: *Virtual Reality*. Ed. by Randall Shumaker. Berlin, Heidelberg: Springer, 2007, pp. 300–309. ISBN: 978-3-540-73335-5. DOI: [10.1007/978-3-540-73335-5_33](https://doi.org/10.1007/978-3-540-73335-5_33).

- [72] Ahmad Abiri et al. “Multi-Modal Haptic Feedback for Grip Force Reduction in Robotic Surgery”. In: *Scientific Reports* 9.1 (Mar. 21, 2019), p. 5016. ISSN: 2045-2322. DOI: 10.1038/s41598-019-40821-1. URL: <https://www.nature.com/articles/s41598-019-40821-1>.
- [73] Soo-Chul Lim, Hyung-Kew Lee, and Joonah Park. “Role of Combined Tactile and Kinesthetic Feedback in Minimally Invasive Surgery”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 11.3 (2015), pp. 360–374. ISSN: 1478-596X. DOI: 10.1002/rcs.1625. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.1625>.
- [74] *Full Body VR Haptic Suit with Motion Capture*. Teslasuit. URL: <https://teslasuit.io/products/teslasuit-4/> (visited on 08/30/2025).
- [75] Mengjia Zhu et al. “Soft, Wearable Robotics and Haptics: Technologies, Trends, and Emerging Applications”. In: *Proceedings of the IEEE* 110.2 (Feb. 2022), pp. 246–272. ISSN: 1558-2256. DOI: 10.1109/JPROC.2021.3140049. URL: <https://ieeexplore.ieee.org/abstract/document/9686045>.
- [76] *Phantom Premium Haptic Devices*. URL: <https://www.3dsystems.com/haptics-devices/3d-systems-phantom-premium>.
- [77] Masaya Kitagawa et al. “Effect of Sensory Substitution on Suture-Manipulation Forces for Robotic Surgical Systems”. In: *The Journal of Thoracic and Cardiovascular Surgery* 129.1 (Jan. 1, 2005), pp. 151–158. ISSN: 0022-5223. DOI: 10.1016/j.jtcvs.2004.05.029. URL: <https://www.sciencedirect.com/science/article/pii/S0022522304008876>.
- [78] Carol E. Reiley et al. “Effects of Visual Force Feedback on Robot-Assisted Surgical Task Performance”. In: *The Journal of Thoracic and Cardiovascular Surgery* 135.1 (Jan. 1, 2008), pp. 196–202. ISSN: 0022-5223. DOI: 10.1016/j.jtcvs.2007.08.043. URL: <https://www.sciencedirect.com/science/article/pii/S002252230701536X> (visited on 08/30/2025).
- [79] Zhan Fan Quek, William R. Provancher, and Allison M. Okamura. “Evaluation of Skin Deformation Tactile Feedback for Teleoperated Surgical Tasks”. In: *IEEE Transactions on Haptics* 12.2 (Apr. 2019), pp. 102–113. ISSN: 2329-4051. DOI: 10.1109/TOH.2018.2873398. URL: <https://ieeexplore.ieee.org/document/8478286>.
- [80] William McMahan et al. “Tool Contact Acceleration Feedback for Telerobotic Surgery”. In: *IEEE Transactions on Haptics* 4.3 (July 2011), pp. 210–220. ISSN: 2329-4051. DOI: 10.1109/TOH.2011.31. URL: <https://ieeexplore.ieee.org/document/5871614>.
- [81] *PyTorch*. PyTorch. URL: <https://pytorch.org/> (visited on 08/25/2025).
- [82] pytorch. *Vision/Torchvision/Models/Vision_transformer.Py at Main · Pytorch/Vision*. GitHub. URL: https://github.com/pytorch/vision/blob/main/torchvision/models/vision_transformer.py (visited on 08/28/2025).
- [83] Jason Yosinski et al. *How Transferable Are Features in Deep Neural Networks?* Nov. 6, 2014. DOI: 10.48550/arXiv.1411.1792. arXiv: 1411.1792 [cs]. URL: <http://arxiv.org/abs/1411.1792>. Pre-published.
- [84] Adam — *PyTorch 2.8 Documentation*. URL: <https://docs.pytorch.org/docs/stable/generated/torch.optim.Adam.html,%20/generated/torch.optim.Adam.html> (visited on 08/28/2025).

- [85] B. Q. Nguyen, C. Kim, and R. W. Farnam. “Operative Times of Robotic Hysterectomy Versus Other Modalities”. In: *Journal of Minimally Invasive Gynecology* 27 (7, Supplement Nov. 1, 2020), S95. ISSN: 1553-4650. DOI: [10.1016/j.jmig.2020.08.114](https://doi.org/10.1016/j.jmig.2020.08.114). URL: <https://www.sciencedirect.com/science/article/pii/S1553465020305045>.
- [86] *Bluno Beetle BLE Power Consumption*. RobotShop Community. Dec. 20, 2016. URL: <http://community.robotshop.com/forum/t/bluno-beetle-ble-power-consumption/27208> (visited on 08/25/2025).
- [87] Texas Instruments. *DRV2605L: 2- to 5.2-V Haptic Driver for LRA and ERM with Effect Library and Smart-Loop Architecture*. Datasheet, SLOS854D. Texas Instruments. Mar. 2018. URL: <https://www.ti.com/product/DRV2605L> (visited on 08/25/2025).
- [88] Adafruit Industries. *Vibrating Mini Motor Disc*. URL: <https://www.adafruit.com/product/1201> (visited on 08/25/2025).
- [89] *Bluno Beetle Simple Tutorial (With Explanatory Images) - DFRobot*. URL: <https://www.dfrobot.com/blog-283.html> (visited on 08/29/2025).
- [90] *Vibrotactile Feedback Images*. Google Drive. URL: <https://drive.google.com/drive/u/2/folders/1yPzZIIzpLYqMI6qGm8uOWSQCZFG5ps7-> (visited on 09/06/2025).

Appendix A

Table of Vibrational modes for DRV2605L driver

Table A.1: DRV2605L Haptic Driver Waveforms

Effect ID#	Waveform Name	Effect ID#	Waveform Name	Effect ID#	Waveform Name
1	Strong click - 100%	42	Long double sharp click medium 2 - 80%	83	Transition ramp up long smooth 2 - 0 to 100%
2	Strong click - 60%	43	Long double sharp click medium 3 - 60%	84	Transition ramp up long smooth 1 - 0 to 100%
3	Strong click - 30%	44	Long double sharp click 1 - 100%	85	Transition ramp up long smooth 2 - 0 to 100%
4	Sharp click - 100%	45	Long double sharp click 2 - 80%	86	Transition ramp up long smooth 2 - 0 to 100%
5	Sharp click - 60%	46	Long double sharp click 3 - 60%	87	Transition ramp up short smooth 2 - 0 to 100%
6	Sharp click - 30%	47	Buzz 1 - 100%	88	Transition ramp up long sharp 1 - 0 to 100%
7	Soft bump - 100%	48	Buzz 2 - 80%	89	Transition ramp up long sharp 2 - 0 to 100%

Adafruit DRV2605L Haptic Driver Waveforms - continued

Effect ID#	Waveform Name	Effect ID#	Waveform Name	Effect ID#	Waveform Name
8	Soft bump - 60%	49	Buzz 3 - 60%	90	Transition ramp up sharp medium 1 - 0 to 100%
9	Soft bump - 30%	50	Buzz 4 - 40%	91	Transition ramp up sharp medium 2 - 0 to 100%
10	Double click - 100%	51	Buzz 5 - 20%	92	Transition ramp up long sharp 1 - 0 to 100%
11	Double click - 60%	52	Pulsing strong 1 - 100%	93	Transition ramp up short sharp 2 - 0 to 100%
12	Triple click - 100%	53	Pulsing strong 2 - 60%	94	Transition ramp down long smooth 1 - 50 to 0%
13	Soft fuzz - 100%	54	Pulsing medium 1 - 100%	95	Transition ramp down medium smooth 1 - 50 to 0%
14	Strong fuzz - 100%	55	Pulsing medium 2 - 60%	96	Transition ramp down medium smooth 2 - 50 to 0%
15	750ms alert 100%	56	Pulsing sharp 1 - 100%	97	Transition ramp down sharp smooth 1 - 50 to 0%
16	1000ms alert 100%	57	Transition click 1 - 100%	98	Transition ramp down short smooth 1 - 50 to 0%
17	Strong click 1 - 100%	58	Transition click 2 - 80%	99	Transition ramp down short smooth 2 - 50 to 0%
18	Strong click 2 - 80%	59	Transition click 3 - 60%	100	Transition ramp down long sharp 1 - 50 to 0%
19	Strong click 3 - 60%	60	Transition click 4 - 40%	101	Transition ramp down long sharp 2 - 50 to 0%

Adafruit DRV2605L Haptic Driver Waveforms - continued

Effect ID#	Waveform Name	Effect ID#	Waveform Name	Effect ID#	Waveform Name
20	Strong click 4 - 30%	61	Transition click 5 - 20%	102	Transition ramp down medium sharp 1 - 50 to 0%
21	Medium click 1 - 100%	62	Transition click 6 - 10%	103	Transition ramp down medium sharp 2 - 50 to 0%
22	Medium click 2 - 80%	63	Transition hum 1 - 100%	104	Transition ramp down short sharp 1 - 50 to 0%
23	Medium click 3 - 60%	64	Transition hum 2 - 80%	105	Transition ramp down short sharp 2 - 50 to 0%
24	Sharp click 1 - 100%	65	Transition hum 3 - 60%	106	Transition ramp up long smooth 1 - 0 to 50%
25	Sharp tick 2 - 80%	66	Transition hum 4 - 40%	107	Transition ramp up long smooth 2 - 0 to 50%
26	Sharp tick 3 - 60%	67	Transition hum 5 - 20%	108	Transition ramp up medium smooth 1 - 0 to 50%
27	Short double click strong 1 - 100%	68	Transition hum 6 - 10%	109	Transition ramp up medium smooth 2 - 0 to 50%
28	Short double click strong 2 - 80%	69	Transition ramp down long smooth 1 - 100 to 0%	110	Transition ramp up short smooth 1 - 0 to 50%
29	Short double click strong 3 - 60%	70	Transition ramp down long smooth 2 - 100 to 0%	111	Transition ramp up short smooth 2 - 0 to 50%
30	Short double click strong 4 - 30%	71	Transition ramp down long medium smooth 1 - 100 to 0%	112	Transition ramp up long sharp 1 - 0 to 50%

Adafruit DRV2605L Haptic Driver Waveforms - continued

Effect ID#	Waveform Name	Effect ID#	Waveform Name	Effect ID#	Waveform Name
31	Short double click medium 1 - 100%	72	Transition ramp down medium smooth 2 - 100 to 0%	113	Transition ramp up long sharp 2 - 0 to 50%
32	Short double click medium 2 - 80%	73	Transition ramp down short smooth 1 - 100 to 0%	114	Transition ramp up medium sharp 1 - 0 to 50%
33	Short double click medium 3 - 60%	74	Transition ramp down short smooth 2 - 100 to 0%	115	Transition ramp up medium sharp 2 - 0 to 50%
34	Short double sharp tick 1 - 100%	75	Transition ramp down long sharp 1 - 100 to 0%	116	Transition ramp up short sharp 1 - 0 to 50%
35	Short double sharp tick 2 - 80%	76	Transition ramp down long sharp 2 - 100 to 0%	117	Transition ramp up short sharp 2 - 0 to 50%
36	Short double sharp tick 3 - 60%	77	Transition ramp down medium sharp 1 - 100 to 0%	118	Long buzz for programmatic stopping - 100%
37	Long double sharp click strong 1 - 100%	78	Transition ramp down medium sharp 2 - 100 to 0%	119	Smooth hum 1 (No kick or brake pulse) - 50%
38	Long double sharp click strong 2 - 80%	79	Transition ramp down short sharp 1 - 100 to 0%	120	Smooth hum 2 (No kick or brake pulse) - 40%
39	Long double sharp click strong 3 - 60%	80	Transition ramp down short sharp 2 - 100 to 0%	121	Smooth hum 3 (No kick or brake pulse) - 30%
40	Long double sharp click strong 4 - 30%	81	Transition ramp up long smooth 1 - 0 to 100%	122	Smooth hum 4 (No kick or brake pulse) - 20%
41	Long double sharp click medium 1 - 100%	82	Transition ramp up long smooth 1 - 0 to 100%	123	Smooth hum 5 (No kick or brake pulse) - 10%