Automation of Surgical Gestures on a da Vinci Research Kit

A Machine Learning Approach for Autonomous Surgical Robots

Aleks Attanasio

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

in

Electronic and Electrical Engineering

The University of Leeds
School of Electronic and Electrical Engineering
STORM Lab

March, 2021

Leeds, United Kingdom

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Abstract

This thesis addresses the problem of automating the execution of surgical gesture. In particular, the case of tissue retraction in which the surgeon or his/her assistant handles the tissue in order to increase visibility in the workspace is considered. In order to tackle this problem, initially a thorough analysis of the state of the art approaches is carried out. By means of this analysis the tissue retraction task is contextualized in the branch of autonomy in robotic surgery. Subsequently, a feasibility study regarding the possibility of performing tissue retraction based on image processing guidance is introduced. In this study a dataset and Convolutional Neural Network model (namely FlapNet) are proposed in order to extract the profile of candidate tissues for retraction in the endoscopic scene. Incremental work on the adoption of spatio-temporal layers such as Long Short-Term Memory cells and Attention Gates is reported showing the benefits of their embedding in the neural network model. Consequently, the robotic action to mobilise and retract the tissue is performed following the guidance of experienced surgeons to design the gesture. In the second part of the thesis a work focused on motion planning is proposed in order to address the translation from the surgeon's guidelines to the robotic platform. Concluding, results showed that, by means of the system proposed in the thesis it was possible to perform tissue retraction on a phantom.

Contents

Abbreviations

AI	Artificial Intelligence
AR	Augmented Reality
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CT	Computed Tomography
DM	Depth Map
DoF	Degree of Freedom
dVRK	da Vinci Research Kit
DVSS	Da Vinci Surgical System
ECM	Endoscopic Camera Manipulator
FDA	Food and Drug Administration
FoV	Field of View
GPU	Graphics Processing Unit
HMD	Head Mounted Display
LbO	Learning by Observation
LSTM	Long Short-Term Memory
MIS	Minimally Invasive Surgery
MRI	Magnetic Resonance Imaging
MTM	Master Tool Manipulator
NN	Neural Network
OR	Operating Room
PSM	Patient Side Manipualator
RAMIS	Robotic Assisted Minimally Invasive Surgery
RCM	Remote Centre of Motion
RGB	Red Green Blue
ROI	Region of Interest
SUJ	Set Up Joint
TR	Tissue Retraction
US	Ultra Sound
VR	Virtual Reality

List of Figures

List of Tables

Contribution

Minimally Invasive Surgery (MIS) presents several benefits to the patient such as reduced trauma and shorter recovery time. On the other hand, the manual dexterity required increases, resulting in long learning curves, lengthy procedures and surgeon's cognitive stress. The last two decades have witnessed the birth and growth of a new approach to MIS based on the adoption of robots in the operation room (OR). These machines ease the surgeons' learning curve while increasing the average precision of their movements, thus enhancing the procedure's outcome. Thus, robotic-assisted MIS is the golden standard. Currently, over four hundred thousands robotic-assisted procedures are conducted every year. Starting from 1999, Intuitive Surgical introduced the first Da Vinci Surgical System. Since then, the surging interest in innovative technologies attracted the attention of companies such as Cambridge Medical Robotics and Verb Surgical, fostering a multi-billion dollar industry. Although the industry has rapidly evolved in the last twenty years, requirements for current robotic systems change continuously, demanding continuing research to provide updated robots. In the last decade, improvements have been focused on autonomy in surgical robotics. However, incorporating autonomous technologies presents technical and regulatory challenges that hinder the deployment of commercial autonomous systems. To date, the surgical robots commonly used in the ORs are capable of solely replicating the surgeon movements, reducing trembling and filtering undesired motion, thus not performing direct actions on the patient. On the 5-level scale of autonomy proposed in [1], the commercially available systems are mostly at level 0, namely "No Autonomy", offering no autonomous features, and simply reproducing the surgeon's movements.

In order to advance towards more efficient and functional robots the paradigm of zero autonomy must be overcome. The research work reported in this thesis has the aim of proposing technical solutions to solve what are the main issues in the automation of surgical tasks. These limitations are:

- **Perception**: Commercial surgical robotic systems are not capable of distinguishing salient features from the endoscopic camera feed. Although many research projects focus on the detection of critical diseases and conditions (like colon polyps [2]), the accurate understanding and reconstruction of the 3D surgical workspace is still an open challenge. Such features would enable a robot to interpret the surrounding anatomy, and lay the foundation for a direct interaction with the anatomy, to pave the path towards autonomy.

- **Motion Planning**: Many commercial robots are equipped with actuated arms and their movement generally mimics the surgeon's motion. To provide a direct interaction with the target's anatomy, a surgical robot must rapidly plan and execute a motion that accounts for the patient's anatomical structure as well as the presence of other instruments to avoid collisions in the case of a cooperative system.

Most of the contributions in the literature on autonomous surgery focus on execution of surgical tasks (such as suturing and tissue retraction) in a highly controlled environment where salient features such as grasping points and obstacles are either defined a-priori or manually highlighted with visual markers. Although this might be satisfactory on the bench-top, real-world scenarios may become intractable when faced with the unique variations presented from case to case. For this reason, a robust system capable of extracting and understanding the workspace is required. To this end, the study presented adopts machine learning models for image feature extraction such as neural networks due to their adaptability, robustness and noise rejection capabilities. The detected features will help describe the surgical scenario in order to better plan the subsequent interaction with the patient's anatomy. As far as the motion planning is concerned, once the workspace is defined and the features of interest are detected, a planning algorithm is considered to rapidly estimate a trajectory and consequently execute a smooth path. The stochastic approach guarantees a short computational time allowing to take into consideration also the presence of obstacles such as anatomical structures or other instruments.

The work of this thesis is carried out on a Da Vinci platform from Intuitive Surgical adapted for research purposes, namely the Da Vinci Research Kit (dVRK). The dVRK [3] is an open-source library that grants access to the Da Vinci Surgical System internal variables. The Da Vinci platform is composed of two main parts: the patient-side robot, equipped with 3 Patient Side Manipulators (PSMs) and 1 Endoscopic Camera Manipulator (ECM), and the master console which instead presents 2 Master Tool Manipulator (MTMs) and a double

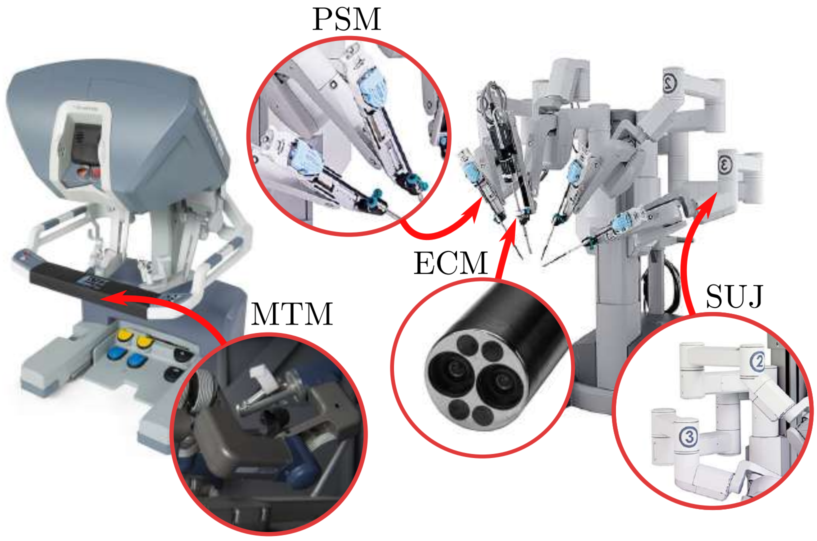


Figure 1: Main components of the daVinci Surgical System. The console on the left is equipped with two Master Tool Manipulators (MTMs), while the patient cart on the right is equipped with three Patient Side Manipulators (PSMs) and an Edoscopic Camera Manipulator (ECM). The PSM are mounted over the Set Up Joints (SUJs)

display visor for the surgeon. With respect to the original system, it presents an additional set of controllers to access the joints data of both the patient cart and the master console. The PSMs are mounted on the Set-Up Joints (SUJ) which kinematic data is not retrievable from the aforementioned controllers. This causes the relative position of the PSMs to be unknown and for this reason, a co-registration method is required.

The implemented trajectory planner of the dVRK allows planning a motion starting from the initial arm position to another reachable point in the robot arm space. This movement is constrained to have initial and final velocity set to zero and prevents the smooth execution of complex trajectories. Additionally, the movements planned are by default considered in the robot space, while usually the user carries out tasks in the camera space.

To date, the dVRK is not provided with any autonomous control to perform surgical gestures. In response to this, the contribution of this thesis is a framework based on Deep Learning models and a stochastic planner for the execution of autonomous tissue retraction, a common surgical task which may be challenging due to the elasticity and flexibility of human tissues. In the last two years, four peer-reviewed papers [4, 5, 6, 7] have been published in top tier journals and conferences, describing in detail the concepts behind the robot's awareness and mobility skills necessary to carry out tissue retraction. These publications are included in this document, as required by the University of Leeds' alternative thesis format, and constitute the core of the research work herein presented. The

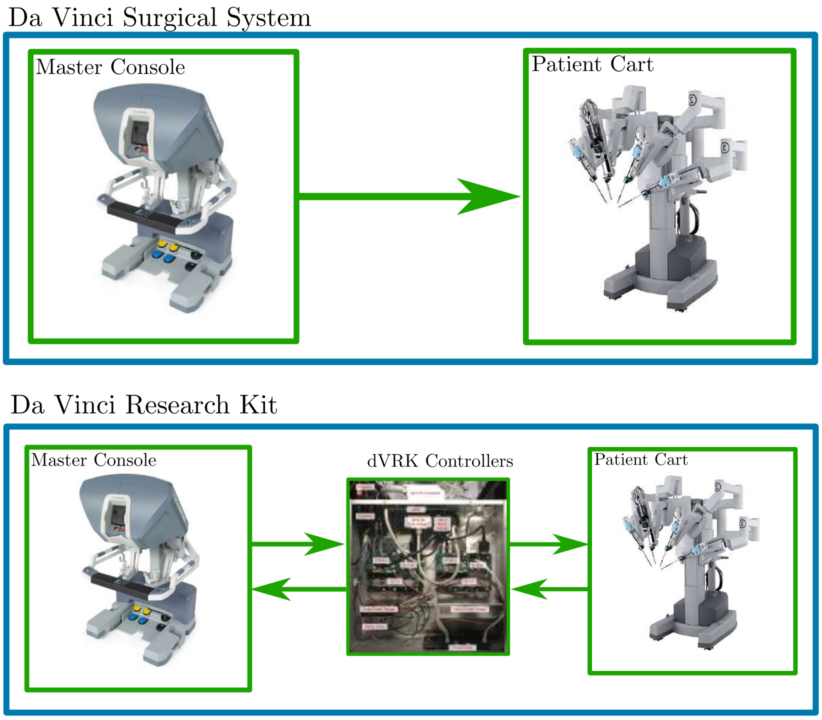


Figure 2: Difference between the DaVinci Surgical System and the Da Vinci Research Kit.

contributions of these publications are listed below.

- **Literature Review**: to understand how to approach the problem of autonomous surgical tasks, a thorough critical analysis of the state-of-the-art is conducted. An initial discussion of the most relevant problems and limitations of the current literature is proposed in the paper "Medical Robotics — Regulatory, ethical, and legal considerations for increasing levels of autonomy” [1]. Starting from this study, an in-depth analysis of the different levels of autonomy in surgery is conducted. On this topic, a comprehensive review of approaches to execution of autonomous tasks in robotic surgery is presented, sorting the most relevant ones into five ascending levels of autonomy. The analysis supports and validate the original idea at the base of this research to further develop the approaches for autonomous surgical gestures such as autonomous tissue retraction. Along with different technologies and related autonomy levels, this study considers ethical and legal aspects. In conclusion, it is possible to draw a comparison with autonomous vehicles, where the technology and the regulatory aspects, as well as the nomenclature, are much more established. As first author of this paper my contribution consisted in gathering, updating and checking all the references while writing down each section dedicated to the levels of autonomy as well as the conclusions. Paolo Fiorini and Pietro Valdastri contributed to the drafting of the introduction and Bruno Scaglioni gathered the information regarding ethical and regulatory aspects of the last section.

Relevant Publications:

  - Attanasio, A., Scaglioni, B., De Momi, E., Fiorini, P. and Valdastri, P. (2020). Autonomy in Surgical Robotics. Annual Review of Control, Robotics, and Autonomous Systems, 4.

- **Perception**: in the case of tissue retraction, as well as of other major surgical gestures such as suturing and ablation, a key aspect for autonomous robotic execution is the detection and localisation of the target tissues. The tissue flaps appearing in the surgical scene are subject to high variability in shape, dimension and pose. Since the tissues to detect appear continuously in camera images during a procedure, an image-based method is recommended to detect the targets candidate tissues for retraction. The detected flaps will be then shown to the surgeon who will acknowledge the tissue retraction execution. Due to the high variability of the light condition, texture, colour and shape of the tissue, traditional computer vision approaches for image segmentation are limited for this purpose. Additionally, the reflection of blood and the organs introduce disturbance which hinders the classic computer vision approaches. For this reason, I developed an adaptive approach for image segmentation based on machine learning to robustly extract the portion of images belonging to flaps of tissue. The FlapNet dataset as well as the code used to train the tissue detector model are available relatively at https://github.com/Stormlabuk/FlapNet and https://github.com/Stormlabuk/dvrk_tissueretraction. On this topic, an initial feasibility study is conducted to explore the possibility to train a neural network model with depth maps for tissue segmentation. Subsequently, I proposed an incremental study to include spatio-temporal layers to find an optimal model for segmentation. Similarly to the initial FlapNet model, the code for this incremental work is publicly available at https://github.com/Stormlabuk/dvrk_ULSTM. As first author my contribution consisted in: collecting, cleaning and labelling the dataset, develop and train the neural network model and testing it on an experimental setups. Chiara Alberti contributed to the technical development of the LSTM model while working to her master thesis. All the other authors supported the publications supervising and reviewing the results.

Relevant Publications:

- Attanasio, A., Scaglioni, B., Leonetti, M., Frangi, A. F., Cross, W., Biyani, C. S. and Valdastri, P. (2020). Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery–A Feasibility Study. IEEE Robotics and Automation Letters, 5(4), 6528-6535.
- Attanasio, A., Alberti, C., Scaglioni, B., Marahrens, N., Frangi, A. F., Leonetti, M., Biyani C.S., De Momi E. and Valdastri, P. (2021). A Comparative Study of Spatio-Temporal U-Nets for Tissue Segmentation in Surgical Robotics. IEEE Transactions on Medical Robotics and Bionics, 3(1), 53-63.

- **Trajectory Planning**: the execution of a surgical task may require the surgical robot to perform complex trajectories. When carried out by a human operator, these paths are planned and executed based on the surgeon's visual information which closes the control loop. In conventional teleoperated robots the continuous co-registration of the arms to the endoscopic camera, held by the ECM, simplifies the control on the surgeon's side. On the other hand, the dVRK, the PSMs and ECM controllers are able to retrieve only the kinematic information of the distal part of the arm, i.e. the actuated arm. The non-actuated arms connecting the PSMs to the base of the da Vinci, called Set-Up Joints (SUJ), require additional controllers to retrieve the joint values and close the kinematic chain from the arm base to the instrument tip. Unfortunately, such controllers are not available yet and a method to co-register these arms is still needed. Regarding this topic, I developed a method for registering the PSM to the ECM by means of visual markers. Secondly, an approach for trajectory planning of complex paths was elaborated in order to provide a trajectory constrained to predefined waypoints. This would allow the performance of complex trajectories that would eventually compose the execution of a surgical task. Both the registration and planning library and scripts are available on the STORM Lab's GitHub page at https://github.com/Stormlabuk/dvrk_stormolib. As first author my contribution consisted in the literature review and technical development of the library. More in detail, I developed the library herein proposed, modifying the dVRK robotic arm model to interfacing it with the MoveIt! framework. Nils Marahrens supported the testing phase by helping me managing the experimental setup.

Relevant Publications:

  - Attanasio A., Marahrens N., Scaglioni B. and Valdastri P. (2021). "An Open Source Motion Planning Framework for Autonomous Minimally Invasive Surgical Robots". IEEE International Conference on Autonomous Systems. [Accepted May 21st, 2021]. Third classified as Best Student Paper.

- **Gesture Design**: given a method to detect and localise tissue flaps in the surgical scene and an algorithm to plan the PSM motion, an adequate gesture must be designed. Several surgeons and doctors working at the St. James University Hospital of Leeds were interviewed and a widespread practice to carry out tissue retraction emerged. This has a fundamental role in this research since an effective and predictable behaviour of the robot is desired for direct and efficient human-machine cooperation.

  Relevant Publications:

  - Attanasio, A., Scaglioni, B., Leonetti, M., Frangi, A. F., Cross, W., Biyani, C. S. and Valdastri, P. (2020). Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery–A Feasibility Study. IEEE Robotics and Automation Letters, 5(4), 6528-6535.

Chapter 1

Literature Review

1.1 Minimally Invasive Surgery vs. Open Surgery

Compared to open surgery, Minimally Invasive Surgery (MIS) (in particular laparoscopy) presents several benefits for the patients such as limited trauma to anatomical structures, shorter post-operative recovery time and reduced blood loss. In laparoscopy, the tools used are characterised by a long shaft and a handle. These instruments allow the surgeon to operate through small incisions (5–12 mm), thus avoiding the access to the patient's anatomy through big cuts which will increase the post-operative recovery time. On the other hand, given the geometry of laparoscopic tools, these happen to be more difficult for the surgeon to handle, thus resulting in more complex and longer operations. The main problem related to laparoscopic tools is the so-called fulcrum effect. As the tools are inserted through the incisions, being them controlled from outside the anatomy, the movements are inverted in both direction and magnitude. An additional problem is posed by the necessity of a member of the surgical equipe to hold and move the endoscopic camera which gives the surgeon the possibility to see what he/she is doing. This leads the surgeon to focus on a monitor which, in most of the cases, is not aligned with the hands. This contributes in increasing the task difficulty by hindering the natural hand-eye coordination of the operator. In order to manage the added complexity, a long period of training is necessary for the surgeon to be independent during a procedure.

1.2 Robotic Assisted Minimally Invasive Surgery (RAMIS)

To overcome the problem of tedious training and difficult manoeuvrability, in 1999 Intuitive Surgical introduced its first model of da Vinci robot. The robot is composed of two main elements:

- **Surgeon's Console**: equipped with two Master Tool Manipulators (MTMs) to control the tools and camera of the patient cart and two stereo display for 3D vision.
- **Patient Cart**: provided with 3 Patient Side Manipulator (PSM) holding tools and an Endoscopic Camera Manipulator (ECM) holding a stereo endoscope for 3D vision.

The main purpose of this robot is to reduce the training time necessary for surgeons simplifying complex manoeuvrers by restoring the hand-eye coordination. This is achieved by solving the main problems regarding the limited dexterity of normal, non-robotic laparoscopic tools. With the da Vinci, the surgeons sight is automatically aligned with the tools which, being controlled robotically can better transpose the surgeons movement resulting in a more natural feeling for the user. With the advent of the da Vinci robot, shorter training times allowed more surgeons to operate on the platform needing less personnel in the operating room, thus granting a more efficient human resources distribution within the hospital.

Nowadays, Intuitive released several models of the da Vinci with new features such as single port access for throat surgery or vertical access for a better arm deployment in the initial phase of the intervention. Although the platform evolved with time along with key features, the paradigm of a slave-master system never changed. This means that no autonomous behaviour of the da Vinci is implemented yet and all the movements of the machine are simply the mirroring of the surgeon's motion. The research presented in this thesis is focused around the development of a framework for autonomous surgical gestures on a da Vinci platform. In particular, the execution of tissue retraction is considered. This task is frequently executed during the initial phase of the operation where the surgeon navigates through the patient's anatomy to reach the surgical site. In this stage, the surgeon resects and mobilises flaps of tissue in order to see and move towards the area of interest. The surgeon's assistant uses manual laparoscopic tools to facilitate the surgeon's activity during the intervention. This

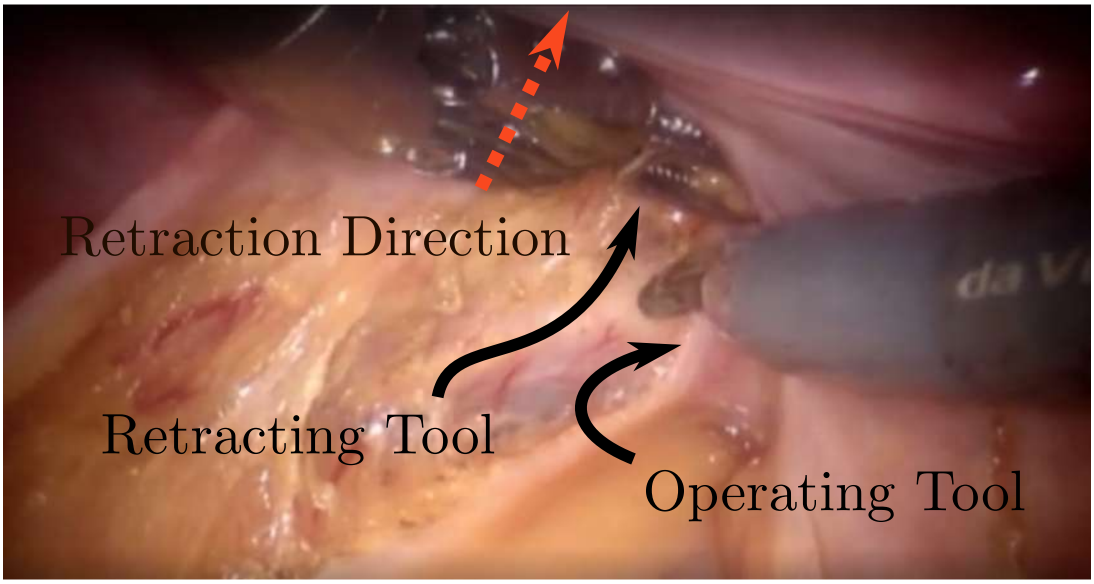


Figure 1.1: Example of tissue retraction. In order to reach the target area with the operating tool the surgeon (or the assistant) reach for the flap of tissue in order to mobilize it away from the region of interest. It is possible to identify the direction of retraction as represented in the figure.

cooperation is hindered by the delayed communication between the surgeon and the assistant and can lead to hazardous situations.

In addition to this, in case the surgeon can't rely on the assistant, the best solution would be to switch the configuration of the PSM taking control over the third da Vinci arm, retract the targeted tissue, and switch back to the original configuration. However, there is evidence in literature of the dangers and risks that may arise from changing the configuration of the arms during an operation. While changing the configuration of the PSMs the surgeon may accidentally move some arms out of the endoscope sight, tearing blood vessels or damaging critical anatomical structures. For this reason it is desired that the surgeon carries out the initial part of the operation working on the Da Vinci along with an assistant who would manually intervene if required by the clinician.

1.3 Autonomy in Surgical Robotics

© Annual Reviews. Adapted and reprinted, with permission, from Attanasio, A., Scaglioni, B., De Momi, E., Fiorini, P. and Valdastri, P. (2020). Autonomy in Surgical Robotics. Annual Review of Control, Robotics, and Autonomous Systems, 4. Supplementary material at https://doi.org/10.1146/annurev-control-062420-090543.

Authors: Aleks Attanasio, Bruno Scaglioni, Elena De Momi, Paolo Fiorini and Pietro Valdastri

Abstract: This review examines the dichotomy between automatic and autonomous behaviors in surgical robots, maps the possible levels of autonomy of these robots, and describes the primary enabling technologies that are driving research in this field. It is organized in five main sections that cover increasing levels of autonomy. At level 0, where the bulk of commercial platforms are, the robot has no decision autonomy. At level 1, the robot can provide cognitive and physical assistance to the surgeon, while at level 2, it can autonomously perform a surgical task. Level 3 comes with conditional autonomy, enabling the robot to plan a task and update planning during execution. Finally, robots at level 4 can plan and execute a sequence of surgical tasks autonomously.

In the last two decades, surgical robotics became an attractive field with a great potential economical impact and it is nowadays a very active research branch. This particular field is the common ground for three different communities: surgeons see the potential of such advanced technologies, engineers find rewarding the challenges posed by these technologies, and entrepreneurs understand the economical potential for future business activities. Although the elevated number of excellent prototypes developed in research laboratories, due to a strict and long certification process associated with the high costs for manufacturing, few of these project are eventually applied in clinical practice. For this reason, the three communities must converge on establishing the scientific and technological fundamentals for the current and future development.

This is the particular case of automation in surgical robotics, a challenging field where the regulatory and legal issues are still a relevant issue to be addressed. A first proposal to structure these research efforts was the Editorial "Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy", appeared on Science Robotics in 2017 [1] that classified the autonomy achievable by a surgical robot into six levels: no autonomy, robot assistance, task autonomy, conditional autonomy, high autonomy, and full autonomy. This classification is inspired by the "Automated Driving" level definition in the field of automotive [8], and adapts these concepts to robotic surgery. However, the transition from the road to the operating room adds additional complexity to the autonomy problem. This is observable also by the fact that while autonomous driving has already achieved Level-3 of autonomy, surgical robots still rest at Level-0 for what concerns commercially available platforms.

Before exploring the possibilities of autonomous surgery, it is fundamental to specify the difference between "autonomous" and "automatic". Automatic behaviors are completely predictable, as they follow well established theories, either deterministic or probabilistic. Although there are variations of behaviors for an automatic system, these are due to small adaptations of the controller parameters to external conditions. If variations are too large, an automatic system cannot adapt and consequently fails. An autonomous system instead, is able to make large adaptations to a change in the external conditions by planning its tasks. The planning function requires a wider domain knowledge and the use of cognitive tools, e.g. ontologies or logical rules that do not exist within an automatic system.

The current COVID-19 pandemic, transforming safe places such as schools and hospitals into hostile environments, made even higher the demand for autonomous technologies with the potential to remotely perform a given task. During the pandemic's peak, the majority of the non-emergency surgical interventions were suspended to prevent a possible infection between medical staff and patient. In this context, tele-operated robots and in particular surgical robots could prove extremely beneficial, potentially allowing a remote controlled or autonomous operation to be carried out without additional risks. These capabilities map the levels of autonomy defined in [1] and could inject new resources and ideas into autonomous robotics research. In robotic surgery however, the levels of autonomy have not been clearly mapped to specific surgical functions and, so far, there is an ample debate on which level of autonomy is appropriate for a given surgical task.

In this chapter the technologies that have already been developed towards the implementation of the desired autonomy level are discussed. In particular, since the terms used in research so far can be ambiguous, we provide here a clear map of the autonomy levels in surgical robotics.

1.4 Level 0 - No Autonomy

A vast amount of literature is available on systems at Level-0, often aimed at describing the commercial scenario and the platforms in a translational stage. The literature suggests a great commercial interest in the field of surgical robotics. The expiration of several patents owned by Intuitive Surgical has attracted the interests of venture capitals, consequently triggering the inception of many new robotic platforms. In year 2000, the DaVinci system from Intuitive Surgical introduced the paradigm of transparent teleoperation, where movements performed

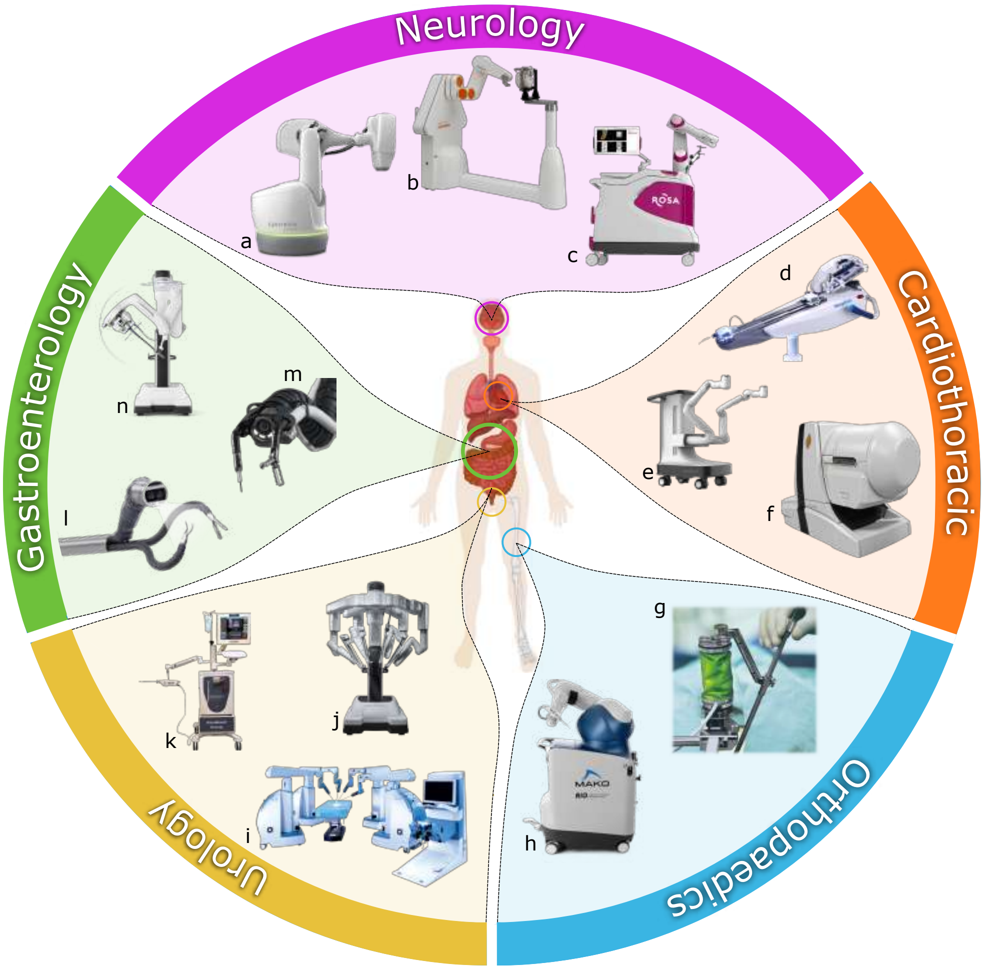


Figure 1.2: Commercially available systems organised by clinical application: (a) CyberKnife, M6 (b) NeuroMate, Renishaw (c) ROSA ONE, Zimmer Biomet Robotics (d) Magellan, Hansen Medical Inc. (e) Monarch, Auris Health (f) Niobe, Stereotaxis (g) Renaissance, Mazor Robotics (h) Mako, Stryker (i) Senhance, Transenterix (j) Da Vinci Xi, Intuitive Surgical (k) AquaBeam, Procept BioRobotics (l) SPORT, Titan Medical (m) Flex Robotic System, Medrobotics (n) Da Vinci SP, Intuitive Surgical.

by the surgeon on the control interface are exactly replicated by the surgical instruments on the patient side. This is the main characteristic of the Level-0 devices and the key feature that led to regulatory approval of surgical robotic platforms so far. The absence of a decision-making process by the machine in the transparent teleoperation paradigm leaves complete control to the surgeon. This feature has allowed Intuitive Surgical to claim that, in the absence of technical failures, the responsibility is totally held by the surgeon. This approach resulted in just two lawsuits reaching trial out of more than 3000 filed against the company up to 2016.

In light of the vast amount of literature available and the specific focus of this thesis on autonomy, we will only provide a brief overview of Level-0 platforms by citing existing review papers in the field. The interested reader can refer to [9] for an exhaustive list of surgical robotic platforms approved by the American Food and Drug Administration (FDA) as of 2018. In [10], a comprehensive description of the commercial systems intended for research purposes as of 2015 is provided. The platforms are classified in commercially available / developed for commercial use (but not on the market) and advanced research prototypes.

Several reviews are dedicated to specific sub-fields; in 2010 [11] provided a detailed description of micro-robots for surgical applications. The paper highlights how the target surgery (ranging from drug delivery to vessels repair) influences the design. The miniaturization trend is described in [12], where the research platforms are mapped on a decreasing dimensional scale. The paper predicts a considerable spread of intracorporeal devices, aimed at tackling pathologies at a cellular level. In a review published in 2013 [13], the authors map platforms in relation to their access route into the patient's body (intralumenal, extralumenal, translumenal and hybrid). For each category, strengths and weaknesses are discussed, along with the open challenges in each particular field.

Another popular classification method is toward clinical applications; in a 2018 review [14], robotic systems for otologic surgery are described, highlighting the need for a robot able to perform cholesteatoma surgery and indicating miniaturization as the main technical issue yet to be solved. In the field of neurosurgery, [15] provides a comprehensive review of the available systems as of 2016, while [16] reviews the available robotic systems for stereo-tactic approaches. From a different perspective, the work in [17] reports an interesting cross section of the evolution of surgical robotic systems, starting from the first commercially available devices: the voice-controlled endoscopic holders AESOP and ZEUS, both discontinued. In the field of colorectal surgery, [18] reviews flexible devices for ednolumenal and translumenal interventions, making a distinction between mechanical and robotic systems, and concluding that mechanical design of both would require massive upgrades to address the clinical needs of surgical endoscopy. Interestingly, no technical review is available in the field of urology, as the Intuitive's daVinci robot is dominating the field. As new competitors (e.g. Cambridge Medical Robotics, Transenterix) enter the market, reviews comparing robotic platforms for abdominal surgery in general, and urology in particular, may be expected.

The literature reviewed in this section suggests a clear trend: while the commercial scenario is flourishing in many directions, the research at Level-0 is focused on development of platforms for unmet clinical needs, such as microsurgery [19], endoscopic intervention [20] and MRI-compatible surgery [21]. On the other hand, for clinical applications where robotics is already well established (e.g. ab-

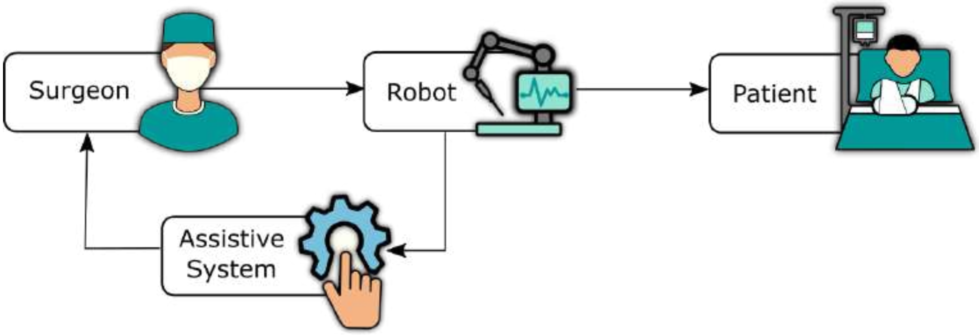


Figure 1.3: Representation of a Level-1 system. The surgeon interacts with the robot which, in turn, provides the clinician with manual guidance or virtual fixtures. In this case, the control loop is closed by the surgeon, who has the full control of the machine for the whole duration of the procedure.

dominal surgery), research is progressing towards higher levels of autonomy, as discussed in the following sections.

1.5 Level 1 - Robot Assistance

Defined as "Robot Assistance", this level includes platforms that provide some support to the operator/user, but never take control of the action being performed, as represented in Fig. 1.3. It is worth noting that, in the context of surgical robotics, the operator/user can always be identified with the surgeon in charge of the procedure, while the patient, often sedated, is the target. This is clearly different from other fields in medical robotics, such as rehabilitation, where the patient often plays the role of user and target at the same time.

The fundamental role of these technologies is detailed in the following section which is mainly focused on systems dedicated to supporting the surgeon in the execution of a specific surgical action. In our analysis, we first identify a number of enabling technologies that are crucial to achieve Level-1 autonomy. Then, we review approaches that provide passive assistance, producing information before and during robotic surgery and, thus, allowing a robot with Level-0 autonomy to reach Level-1. Examples include systems that suggest optimal robot deployment and ports placement, and systems providing augmented reality. We then analyse solutions that actively interact with the surgeon by providing mechanical support (i.e. guidance or compensation of periodic motion) and, lastly, we discuss the role of haptic feedback.

1.5.1 Enabling Technologies

Most research platforms operating at Level-1 acquire a limited amount of information characterized by relatively low complexity, typically related to either the robot, the surgeon, or the target tissues. Therefore, tool tracking, eye tracking and tissue interaction sensing can be considered as three enabling technologies to achieve Level-1 autonomy. In addition, surgical robot controls paradigms are the foundation on which all the autonomy levels poses [22]. However, as the control specifications varies for every medical application, an in-depth analysis of control methods for surgical robotics is not discussed.

1.5.1.1 Tool Tracking

Surgical tools tracking is a core component for developing assistive technologies such as augmented reality and haptic feedback. Tracking the appearance of a tool from an endoscopic view and project it to the robot space can in fact provide the required information to enrich the scene with augmented reality elements such as the pressure applied to an anatomical area or the force to which a tissue flap is subject while grasped. Different methods were proposed in the last twenty years starting from computer vision to sensor fusion of both cameras feed and kinematic information. Approaches in literature can be clustered in three main groups. During the first half of 2000's, just computer vision was used [23]. Then, with the launch of surgical open research platforms such as the daVinci Research Kit (dVRK) [24] and Raven II [25] and with the availability of kinematic data from the robot manipulators, new approaches were developed with improved robustness and lower dependency from light conditions [26, 27]. More recently, machine and deep learning became popular, with papers using either Gaussian Mixture Models (GMM), Convolutional Neural Network (CNN) [28] and Random Forests [29]. Video stream and kinematic data were combined to strengthen the generalisation capabilities of machine learning models such as Random Trees [30]. A 3D-CNN structure was implemented to account for the correlation between subsequent frames in [31]. Ultrasound imaging was implemented in [32] to enhance accuracy.

Overall, works from the literature show that it is possible to track surgical tools at a rate of 29-30 Hz [28], thus guaranteeing a smooth real-time video stream, while maintaining a sub-mm tracking error for most applications.

1.5.1.2 Eye tracking

Eye-tracking is nowadays an established technology, commonly adopted in research fields outside surgery (e.g. customer behaviour and user experience). In robotic surgery, it can be used to capture the surgeon's gaze in augmented and mixed reality applications or to study surgical task recognition.

In open or laparoscopic surgery the most common approach is the adoption of head-mounted devices (HMD). Such systems include glasses with tracking cameras [33] or optical trackers for estimating the pose of the head [34]. However, acceptance of HMDs by the surgical community is limited and a convincing clinical application, demonstrating effectiveness in a real surgical scenario, is still missing [35]. On the other hand, whenever a visualization device is already in use (i.e. surgical microscope, immersive user console), eye tracking is not disrupting the clinical workflow and can be easily adopted for guiding assistive tasks, such as instruments control by surgeon's gaze [36, 37].

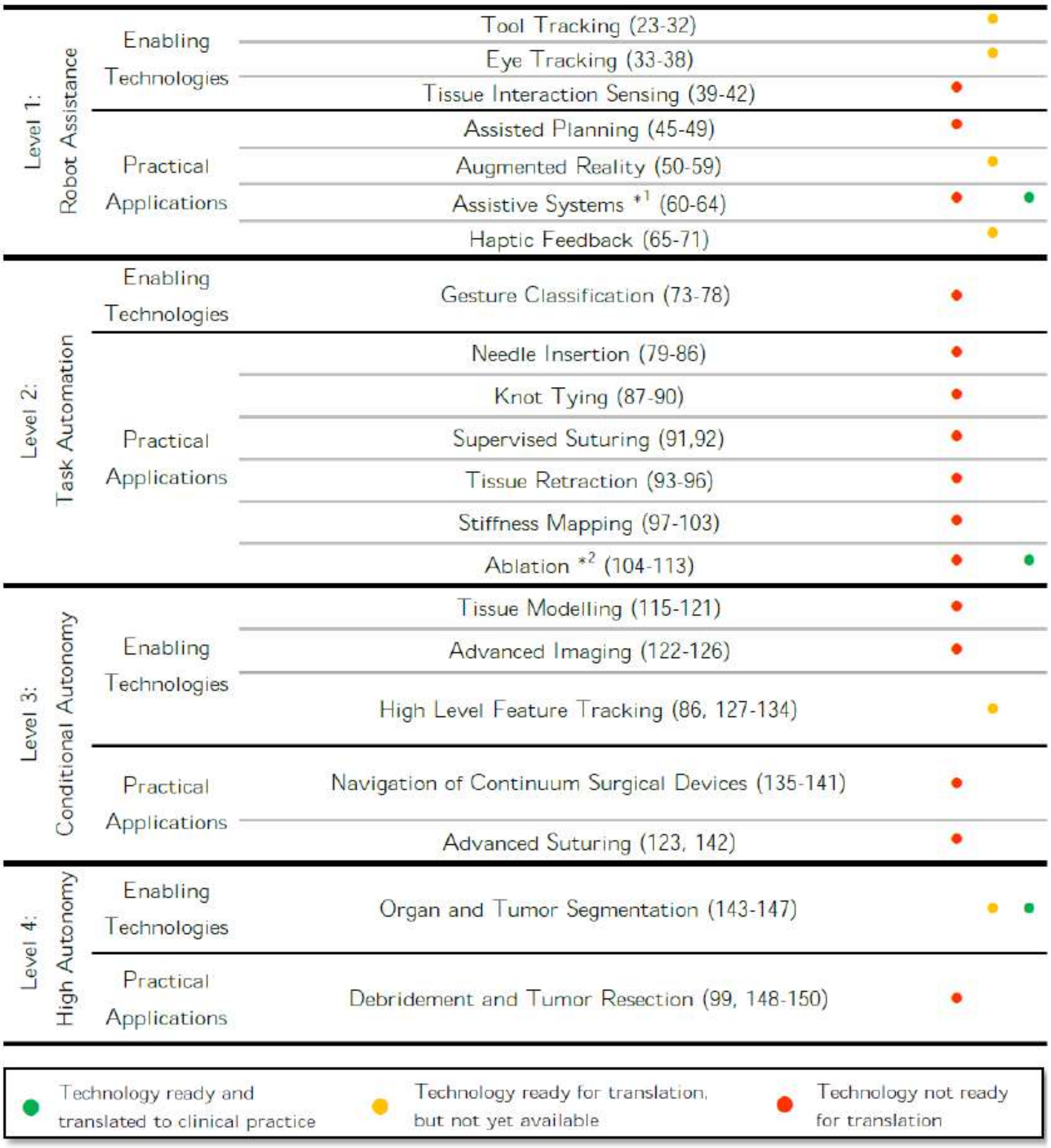
Overall, despite the availability of commercial systems for eye-tracking (e.g. Tobii [38] Pro or EyeLink by Sr Research), translating this technology to surgical applications is not immediate as a strong case for its adoption has yet to be demonstrated via convincing clinical studies.

1.5.1.3 Tissue Interaction Sensing

A feedback on the interaction between instrument and tissue is crucial for safety and efficacy in both open and minimally invasive surgery. Research in providing surgical instruments with force/torque/grasping/contact sensing capabilities has been extensive in the last two decades [39, 40]. However, this effort has been hampered by the additional complexity and cost that sensing adds to instruments that either need autoclave sterilization (if reusable) or extremely low fabrication costs (if disposable). For this reason, sensor-less options using force and torque estimation [41] or data-driven vision-based sensing [42] have recently gained traction in the research literature. More complex approaches adopt convex optimisation [43] or screw theory [44] to estimate tool dynamics. While extremely promising and straightforward to implement in a controlled environment, these approaches have yet to be demonstrated outside a research lab environment [42].

1.5.2 Passive Assistance

In the context of this section, passive assistance technologies are intended as systems that assist the surgical activity by providing additional information to



Level 1: Robot Assistance	Enabling Technologies	Tool Tracking (23-32)		🟡	
		Eye Tracking (33-38)		🟡	
		Tissue Interaction Sensing (39-42)	🔴		
	Practical Applications	Assisted Planning (45-49)	🔴		
		Augmented Reality (50-59)		🟡	
		Assistive Systems $*^1$ (60-64)	🔴		🟢
		Haptic Feedback (65-71)		🟡	
Level 2: Task Automation	Enabling Technologies	Gesture Classification (73-78)	🔴		
	Practical Applications	Needle Insertion (79-86)	🔴		
		Knot Tying (87-90)	🔴		
		Supervised Suturing (91,92)	🔴		
		Tissue Retraction (93-96)	🔴		
		Stiffness Mapping (97-103)	🔴		
		Ablation $*^2$ (104-113)	🔴		🟢
Level 3: Conditional Autonomy	Enabling Technologies	Tissue Modelling (115-121)	🔴		
		Advanced Imaging (122-126)	🔴		
		High Level Feature Tracking (86, 127-134)		🟡	
	Practical Applications	Navigation of Continuum Surgical Devices (135-141)	🔴		
		Advanced Suturing (123, 142)	🔴		
Level 4: High Autonomy	Enabling Technologies	Organ and Tumor Segmentation (143-147)		🟡	🟢
	Practical Applications	Debridement and Tumor Resection (99, 148-150)	🔴		

🟢 Technology ready and translated to clinical practice
🟡 Technology ready for translation, but not yet available
🔴 Technology not ready for translation

Figure 1.4: Readiness level of enabling technologies and research areas for the different levels of autonomy. $*^1$ in orthopaedics the problem of assistive systems is solved for specific applications (see the Mako, Stryker for joint replacement), the problem is not solved yet for soft tissue surgery. $*^2$ Ablation for specific application such as the treatment of benign prostatic hyperplasia is a commercially solved problem (see AquaBeam, Procept Biorobotics), however, ablation in endoscopic surgery is still matter of research. $*^3$ In neurosurgery the segmentation of tumors from MRI images is already implemented in the Brainlab technology. The challenge remains open for thoracic and abdominal surgery.

the surgeon. A considerable amount of work has been carried out in this field, particularly during the first decade of the century. Here, we focus on two specific research streams: assisted planning, relevant before the surgery starts, and augmented reality, which is available during the procedure to amplify surgeon's cognitive capabilities.

1.5.2.1 Assisted Planning

Optimal ports placement is a common issue in minimally invasive surgery (MIS) due to the limited reach, articulation and dexterity offered by endoscopic instruments. A poor placement at the beginning of the procedure may introduce undesirable delays and require re-placement while the patient is under anesthesia.

Laser pointers and light emitting diodes (LED) mounted on the tip of surgical tools were adopted in [45] to simplify the deployment of laparoscopic instruments. Other approaches capitalize on pre-operative analysis, such as computer tomography (CT) and magnetic resonance imaging (MRI), to develop virtual reality for planning in neurosurgery [46] or to minimize collisions in abdominal and thoracic surgery [47]. In orthopaedics, pre-operative 3D scans are used to manufacture patient-specific tool guides, thus increasing surgeons' accuracy during osteotomies [48]. All the mentioned contributions deal with hard tissues or instruments' geometry. Planning algorithms involving soft tissues are still an open challenge, due to the inherent complexities in modelling of the tissue and the lower reliability of registration with pre-operative imaging as reported in [49].

1.5.2.2 Augmented Reality

Introduced in surgery in 1986 [50], Augmented Reality (AR) gained momentum in the last three decades, enabled by the increased amount of computational power. With AR, additional information such as tumor location or hidden instruments can be shown to the clinician by superimposing virtual objects to the endoscopic image. Pre-operative images (CT, MRI, ultrasound) are used to extract the shape and location of the target. Subsequently, 3D models are registered to the anatomy. MRI- and CT-compatible fiducial markers can be adopted to address issues in registration [51]. Fluorescent fiducials are proposed to account for intra-operative deformations in [52]. The fusion of pre-operative annotated MRI and intra-operative trans-anal ultrasound is proposed in [53]. To provide high-level information in AR, context-awareness is required. An example is provided by [54] where different visualisations are proposed to the user depending on the tumor resection phase, autonomously detected by the system. The visualisation includes the targeted area, the resection margins or vital structures nearby the region of interest.

While most of the research target the surgeon as end user of the technology, [55] introduced ARssistant, an HMD that shows the location of robot instruments inside the patient's body to the assistant. This approach is particularly interesting from the clinical perspective, as literature reports many adverse events [56] caused by a lack of coordination between assistant and surgeon.

An exception worth mentioning is

Overall, research in the field of AR for surgery and robotic surgery is well established and is gaining momentum as a product. An example worth mentioning is the success of AR for surgical training [57], with several commercially available platforms already in use [58], and clear potential for expanding training programs to low-resource settings around the world [59].

1.5.3 Active Assistance

Robotic systems actively interacting with the surgeon at Level-1 are classified as active assistance systems. These devices perform actions that affect the surgical procedure, such as applying forces to the user interface or restricting motion of the surgical instruments, based on a limited knowledge of the environment (i.e. force sensors, pre-computed forbidden areas, periodic inputs, etc.). In this case, the robot does not have the ability to control the execution of tasks, but rather reacts to actions initiated by the surgeon.

1.5.3.1 Assistive Systems

Minimally invasive surgery on soft tissues may be affected by periodic movements such as respiration or heart beat. Compensation of oscillatory motions has the potential of reducing undesired interaction with the anatomy. Techniques such as Smith predictors [60] and Fourier series models [61] were adopted for heart beating motion forecasting. Validation tests proved the system to be able to reduce by a factor five the tracking error of the system compared to state of the art. Experimental results reported in [62] show that such technology enhances the clinicians' dexterity reducing by a third the rate of missed hits in a suturing task.

In surgeries where the environment is more "stable" or better constrained, such as neurosurgery, microsurgery or orthopaedic surgery, robot assistance can be provided to prevent undesired interactions with delicate structures. In this case, artificial repulsive force fields are generated and applied to the surgical tool tip. Passive control schema have been developed to guarantee the stability of the tool-tissue interaction [63].

In orthopaedic surgery, active constraints are already part of commercial platforms, such as the Stryker Mako and the Zimmer Biomet ROSA, shown in Fig. 1.5, which improve precision during interventions such as knee cap replacement,

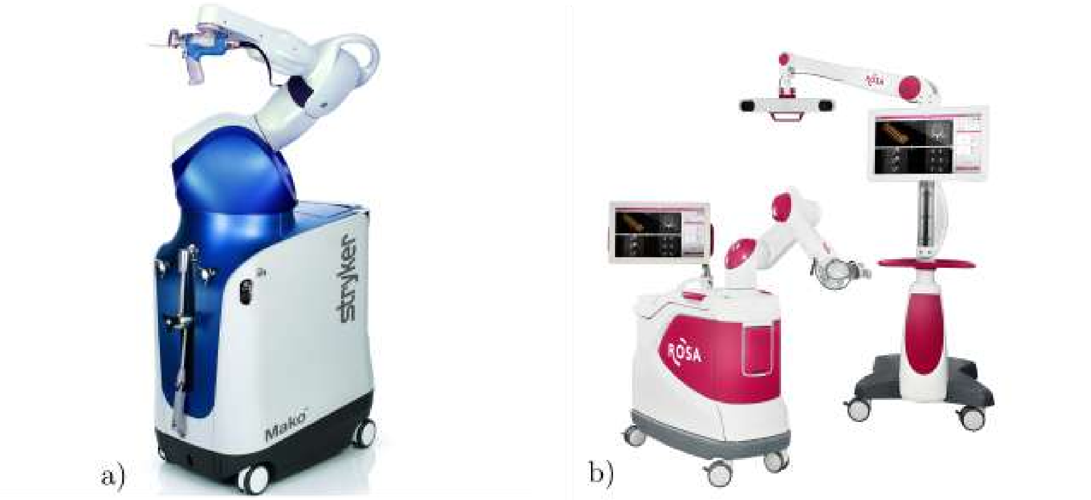


Figure 1.5: The Mako from Stryker (a) and the ROSA platform by Zimmer Biomet Robotics (b) are used in orthopaedics for joint replacement.

total knee arthroplasty, or total hip replacement. On the other hand, in cardiology a first example of a commercial platform employed for automated surgery is proposed by Corindus Vascular Robotics which provide a “Rotate-on-Retraction" gesture to simplify the navigation of a guidewire within blood vessels [64].

1.5.3.2 Haptic Feedback

Surgeons heavily rely on tactile and force feedback during open surgery. Such feedback is severely hampered in MIS and completely lost in current robotic surgery. Robotics offers the opportunity to restore haptic sensation by means of sensors placed at the instrument and actuators integrated within the user interface. A large body of research exist in this field, mainly driven by technological advancements in sensing and actuation.

The most common approach is based on mechatronics [65], but pneumatic [66] and hydraulic [67] systems have also been proposed. Haptic feedback can be used for tissue palpation, thus identifying buried tissue structures or stiffer regions, to develop assisted guidance of surgical tools in a confined space [68], or to prevent instrument collision [69]. So far, the consensus from the surgical community has been that high definition 3D vision, combined with the high dexterity and precision of robotic tools, were sufficient to cope with the lack of haptic feedback [70]. Nevertheless, platforms such as the Transenterix Senhance [71], that recently entered the market, are equipped with haptic feedback. It will be extremely interesting to see if clinical data from surgeries performed with new robotic platforms will be convincing enough to modify the opinion of the surgical community.

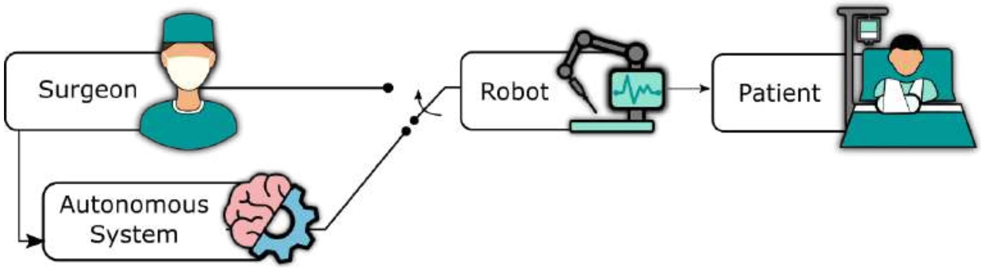


Figure 1.6: In systems belonging to Level-2, the surgeon provides the necessary information for the robot to accomplish a given task. Since during the autonomous execution the control passes from the surgeon to the machine for the time needed to perform the action, we refer to this as discrete control, represented by the switch.

1.6 Level 2 - Task Autonomy

The second level of autonomy is defined as “Task Autonomy”. At this level, the robot can take control of the procedure, but does not possess the ability to define any parameter for planning the task. The surgeon provides the information required to perform the action and the robot executes. The aim of task autonomy is to free the surgeon from the cognitive burden and/or fatigue associated to complex and/or repetitive tasks.

An example is tip retroflexion in magnetic colonoscopy. In colonoscopy, retrograde vision allows the operator to investigate a larger portion of the bowel. However, predicting how to change the controlling magnetic field and field gradient to achieve the desired motion at the tip of the endoscope is extremely complex for a human operator. In [72], an autonomous algorithm is proposed that tracks in real-time the pose of the endoscope tip and adjusts the pose of the external driving magnet accordingly in order to achieve retroflexion. The robotic colonoscopy platform normally works in transparent teleoperation with active constraints (Level-1) and, when the operator needs retroflexion, the algorithm kicks in.

Similarly to retroflexion in endoscopy, tasks such as tissue retraction, suturing and ablation can be automated in robotic surgery. To enable these and other autonomous tasks execution at Level-2, technologies like task recognition and tissue palpation are essential. Papers discussing these lines of research are reviewed in this section.

From an ethical standpoint, task autonomy is the first level where the machine takes full control of the surgical instruments, although for specific gestures and under the supervision of the surgeon. This "discrete shared control" introduces a paradigm shift in the ethical and regulatory framework that needs to be addressed by notified bodies and surgical robotics companies alike to allow autonomy to get into the operating rooms. This topic is discussed in more detail in Section 1.9.

1.6.1 Enabling Technologies

At Level-2, the robot does not own the ability to elaborate decisions; nevertheless, it is required to retrieve information with higher complexity with respect to Level-1. For this reason, we selected a technology that has a great impact on systems at this level: Gesture Classification. In robotic surgery, by gesture it is intended a sequence of actions (e.g. move towards a needle, grasp a thread, pierce a tissue with a needle). A surgical procedure is composed of a sequence of gestures aimed at, for example, navigate through the anatomy, reach a specific anatomical area, operate the target usually by removing the compromised tissue and recover the tools after having secured the anatomy. Gesture Classification can enhance the ability to activate the robot at the right time and minimize the disruption to the surgical workflow. By means of gesture classification, the robot is capable to follow the clinician's work plan, thus providing dedicated support depending on the phase of the operation.

1.6.1.1 Gesture Classification

A correct surgeon-assistant interaction is crucial to reduce the chance of mistakes during surgery [56]. In Level-2 the robot can be considered as an assistant, executing basic sub-tasks. To achieve a satisfactory coordination between surgeon and robot, the identification of the surgical task is crucial. To achieve this, several solutions have been proposed using source of information such as endoscopic videos [73], real [74] and simulated kinematic data [75] and depth images [76]. Video streams and kinematic data can introduce significant computational burden, preventing the system to work in real-time. To tackle this issue, machine learning models such as Hidden Markov Models [77], weakly supervised Gaussian Mixture Model [74], multiple kernel learning [73] and Recurrent Neural Networks [78] have been adopted. Promising results have been achieved on bench-top test scenarios. However, as the complexity and variability of real tasks increases, the model accuracy, generally ranging between 70% and 85%, tends to decrease, thus limiting applicability in their current implementations.

1.6.2 Suturing

Although widely performed in many surgeries, suturing remains a critical task as failures might lead to disastrous consequences. Surgeon ability and experience play a crucial role in the quality of a suture, thus automating this repetitive task would guarantee more uniform outcomes and relax the cognitive burden on the surgeon.

The execution of autonomous suturing is generally divided in two stages: the insertion of the needle, during which the needle pierces the tissue and is re-grasped at the exit point, and the tying of a knot to secure the suture with a surgical thread. A significant amount of literature regarding each sub-task is available.

1.6.2.1 Needle Insertion

The needle insertion stage entails high precision in estimating the optimal position, angle and applied force required to pierce the tissue. Moreover, the discontinuous grasp of the needle generates uncertainties in the pose estimation. Finally, as the needle pushes through the surface, the tissue is subject to deformation, thus increasing the needle pose uncertainty. For these reasons, autonomous needle insertion raises interesting technical challenges, mainly related to the needle geometrical model and tissue deformation. A combination of kinematic and geometric modelling is proposed in [79], where the trajectory is generated to minimize the tissue deformation. Estimation approaches such as Unscented Kalman Filter [80] and an online evaluated deformation matrix [81] were used to estimate the tissue and needle deformation.

A crucial aspect of needle insertion is the definition of the entry points. In order to simplify this problem, optical markers [82] and laser pointers [83] were integrated with optimisation techniques, with the aim of minimising the tissue strain. However, the intra-operative placement of optical markers could be undesirable in surgery, thus reducing the advantages of autonomous execution. [84] proposed a solution for a single-master dual-slave platform for semi-autonomous needle insertion. The surgeon controls one arm to insert the needle while the second arm, triggered by the insertion force, collects the needle and returns it to the surgeon.

More advanced approaches improve the success rate by adopting transfer learning (a method to transfer the learnt knowledge from an artificial intelligence model to address a different problem) [85], which reports a success rate of 87% in needle driving, or Sequential Convex Programming [86].

Even though satisfying results are reported on bench-top trials, validation on a realistic scenario considering tissue-specific mechanical properties and the presence of anatomical structure at risk, such as nerves and blood vessels, is still missing.

1.6.2.2 Knot Tying

The last step in suturing consists of tying a knot .The main technical challenge is related to the deformability and resistance of the thread, which could lead to undesired entanglement, thus damaging the tissue. The complexity of the task is further increased by limited dexterity, confined workspace and lack of tactile feedback. To mitigate uncertainties on the thread deformation, [87] proposed to apply a constant tension. The paper shows the feasibility of two different knots with performance comparable to human execution (nearly 10 seconds). Interestingly, during retraction, the tissue is subjected to external forces and deformation, thus requiring a continuous re-planning. More advanced techniques enhance the robustness of the autonomous system by using machine learning approaches such as Learning by Observation (LbO, also known as Learning by Demonstration) to extract the fundamental features from human gestures. In [88], manually performed tasks are used to train Long Short Term Memory Recurrent Neural Networks. This type of neural network is particularly interesting as it is capable of considering temporal evolution of features, thus allowing the algorithm to learn complex sequences of gestures typical of knot tying. In [89], LbO is used to generate trajectories on a phantom starting from recorded manual sutures, achieving an accuracy of 2 mm in the path execution.

As the methods proposed in the literature vary significantly, an objective comparative assessment of the performance is not straightforward. For this reason, [90] proposed an evaluation metrics, comparing 4 different approaches. However, thread deformation still hinders satisfactory results. The adoption of high visibility threads may simplify the tracking problem, reducing the uncertainty on the pose detection of the string.

1.6.2.3 Supervised Suturing

Literature that simultaneously tackle needle insertion and knot-tying on commercial robotic systems is limited. The most convincing solution at the moment entails the development of a dedicated platform, the Smart Tissue Autonomous Robot (STAR), for full autonomous anastomosis [91]. The system is composed of a 7-DOF KUKA LBR arm equipped with a custom suturing tool [92] (Figure 1.7).

Two working modes are available: in automatic mode, the system autonomously evaluates the position of each entry point, starting from the suturing outline defined by the surgeon, in manual mode each entry point is defined manually. Tests on phantoms demonstrated that the system is capable of completing a suturing task 5 times faster than a robot-assisted procedure and 9 times faster than an operator. It is worth to point out that, in both working modalities, the surgeon is required to define the suturing profile of the anastomosis. Systems capable of autonomously retrieving the suturing task specifications will be introduced in Section 1.7.3, at Level-3 of autonomy.

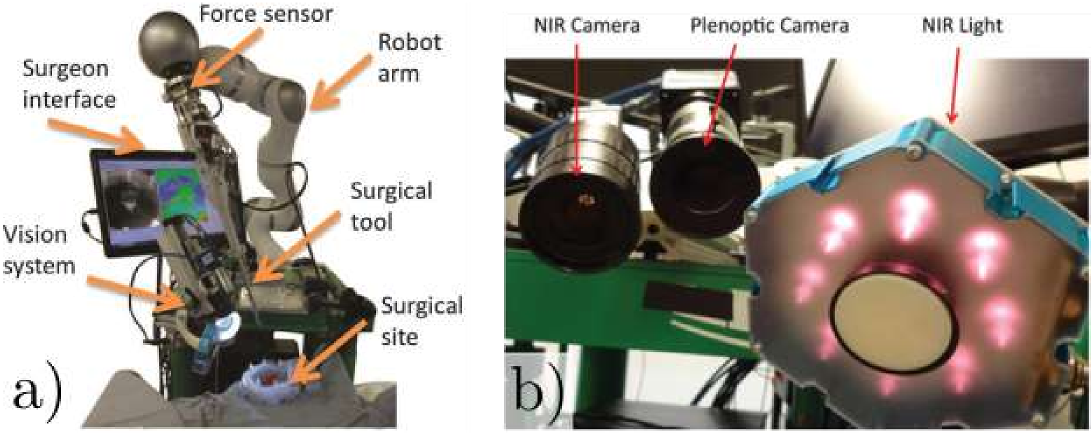


Figure 1.7: The Smart Tissue Anastomosis Robot (STAR) system (a). The system is equipped with Plenoptic camera to retrieve depth information while the Near Infra-Red (NIR) camera detects hidden structures in the tissue (b).

1.6.3 Tissue Retraction

During MIS procedures, a significant amount of time is spent mobilising and dissecting tissue to reach the area of interest. In this context, dissected tissue is often retracted to expose the surrounding area. Although this gesture is performed frequently and, thus, would make a good candidate for task automation, few contributions are available in literature. This may be due to the complexity associated with detecting and tracking deformable soft tissue during surgery. Simulation frameworks have been developed in [93, 94] to plan a grasp-and-retract gesture. The strategy aims at minimising the tissue strain, simultaneously avoiding tearing and guaranteeing an obstacle-free trajectory. Recent studies [95, 96] present tissue retraction on a dVRK involving visual markers to identify the flap grasping point and fuzzy logic to execute the gesture. Despite the promising results, only bench-top experiments are available. In a real scenario, the complexity of tissue detection may be significantly higher, considering the tissue elasticity and the presence of tools. For this reason, automating tissue retraction remains an open challenge.

1.6.4 Stiffness Mapping

Manual palpation is commonly used in conventional surgery to identify and dissect malignant masses below the surface of organs (i.e. kidneys, lungs). In robotic surgery, the lack of tactile feedback hinders the surgeons' ability to evaluate the tissue properties. Haptics, discussed in 1.5.3.2, aims at restoring this ability. A further step towards the execution of autonomous tasks such ad dissection and ablation (see Section 1.6.5) is the ability to autonomously estimate the tissue properties by mechanical contact. To provide palpation, array sensors based on different principles such as the measure of distributed pressure on a surface [97] and the Bernoulli pipe structure [98] have been adopted, detecting hard inclusions with a precision of 97%. Hall sensors were implemented on a daVinci instrument tip [99] to localise blood vessels to a maximum depth of 5 mm. A different approach for prostate palpation [100] is based on the adoption of an attachable sensor matrix. Approaches based on sensors are limited by increased complexity, cost and sterilization requirements. In a seminal paper, [101] demonstrate sensorless palpation with a multi-backbone continuum robots for the first time. Based on this work, [102] proposes a smart navigation approach supported by pre-operative images.

In [103], elastography had been used to collect a dataset and develop a machine learning model for autonomous detection of hard inclusions in a phantom. Most of the proposed strategies for palpation adopt custom, hand-held instruments, thus increasing the number of surgical accesses required. A solution integrated with commercial systems would be preferred. Moreover, the adoption of dedicated mechanical devices introduces complexities associated to reprocessing, possible contamination and production costs, thus significantly limiting the potential for clinical translation.

1.6.5 Ablation

Ablation consists in eradicating a portion of compromised tissue by transferring a high amount of energy to the target by means of electric cauterizers, cryoprobes or High Intensity Focused Ultrasound (HIFU). The major risk is undesired removal of healthy tissue from surrounding structures such as blood vessels or nerves bundles. The correct localization of the target tissue to remove and the precise identification of its margins pose technical challenges, especially in surgical excision of cancer, where tumors may be concealed underneath healthy tissue. In these procedures, it is also crucial to spare as much healthy tissue as possible to prevent organ failure and subsequent need of a transplant, should this be an available option. A possible approach is to perform mechanical palpation to create a local stiffness map for guiding ablation [104, 105], as discussed in Section 1.6.4. Alternative imaging methods such as ultrasound [106] and optical coherence tomography [107] may be adopted to guide cardiac ablation.

Following a common practice in surgery, several works considered laser ablation to reduce direct interaction with the anatomy [108, 109, 110]. Nonetheless, the lack of physical contact complicates the identification of the target area, which is manually selected by the surgeon before starting the procedure [109]. To relax the input required from the surgeon and increase the autonomy in detecting the target, preoperative scanning and voxel-growing on the 3D anatomic model were successfully implemented [108].

In order to avoid heat generation [111] adopted cryoprobes, while [112] used pressurized water jets. The latter is a commercially available system (branded as AquaBeam [113]) designed for the treatment of benign prostatic hyperplasia. Although the prostate profile is identified by the surgeon, on ultrasound scan, the resection is autonomously performed by a high-pressure water jet. This is a remarkable example of a Level-2 system reaching the operating room, enabled by the simultaneous use of intra-operative ultrasound imaging and robotics. Extending Level-2 systems for autonomous ablation to other surgical procedures will be challenging whenever the localisation of the target area is hindered either by tissue deformation or lack of visualization.

1.7 Level 3 - Conditional Autonomy

The main characteristic of Level-3, defined as “Conditional Autonomy’, is the ability to conceive strategies to perform a specific task, always relying on the human operator to approve the most suitable to be implemented. In the context of robotic surgery, this reflects the ability to autonomously extract the parameters required to plan a specific task from the information available to the system. During task execution, the environment is constantly monitored and the plan is updated in real-time. In case of performing a suturing task at Level-3, for example, the system would be able to extract the suturing points and the length of each suture from real-time imaging, then plan and execute the suture

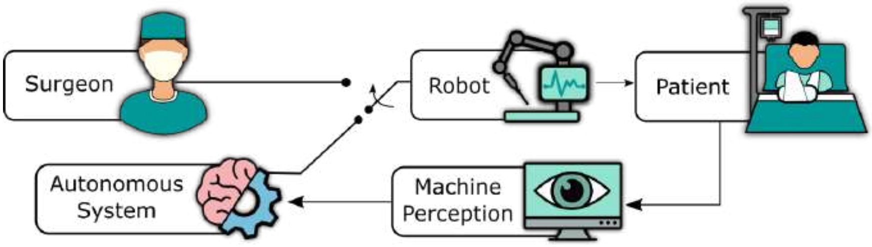


Figure 1.8: Systems belonging to Level-3 are capable of autonomously define the specifications to plan and execute a surgical task, differently from Level-2 systems where the surgeon was supposed to provide them to the system. Similarly to Level-2, discrete control takes place, as represented by the switch.

autonomously. At Level-2 this task would have required the intervention of a surgeon to explicitly designate the insertion point of the needle, while at Level-3 the system is capable of estimating the scene and retrieve the required feature without the human intervention. Real-time imaging would also provide continuous updates to the plan as the task is performed. This example is discussed in detail in Section 1.7.3. Other examples are autonomous navigation of flexible endoscopic robots in unstructured environments [114], autonomous navigation in the abdominal anatomy, and autonomous anastomosis.

1.7.1 Enabling Technologies

At Level-3, the system requires the ability to perceive, extract and analyze contextual elements to plan how to execute a task and to update the plan during execution. Similarly to the surgeon's cognitive process, systems at Level-3 are expected to extract high-level features from the surgical scenario and to act upon them in real-time. Some of the key elements to achieve this are computationally-efficient tissue models, advanced imaging capabilities, and algorithms to track high-level features in the environment.

A significant help in this context comes from the giant leap in computational power of graphic processing units (GPUs) that we have experienced in the last decade. Current GPUs, mainly developed for the gaming industry, can be used to run complex algorithms at an unprecedented speed.

1.7.1.1 Tissue Modelling

Predicting tissue deformation plays a crucial role in manipulating soft tissues. Understanding the mechanical properties is essential to avoid unintentional damages. Additionally the machine performances defined by the hardware setup heavily impacts the outcome of such technology. In fact, tissue modelling, demanding a significant amount of calculation for 3D rendering and physical analysis, puts a considerable stress on the computational hardware. In practice this translates to a trade off between fidelity, thus how realistic the analysis is, and computational speed. Considering how these two aspects affect the outcome of a surgical procedure an correct balance must be achieved to guarantee optimal procedures' outcome. Research has focused on deformation assessment by means of data-driven approaches in needle insertion [115], and real-time detection of tissue perforation during spine interventions [116]. Intra-operative real-time images [117] and pre-operatory CT scans [118] were used to obtain anatomy-specific deformation models. However, one of the main challenges is the intra-operative real-time identification of tissue parameters describing the elasticity and stiffness of a given surface. Recently, 3D displacements and kinematic data were combined to evaluate the deformation through optimization techniques in [119, 120]. Results demonstrated the system ability to evaluate the parameters in real-time, but quantitative assessments of the performances are not available. Only one work presented an approach to model the cardiac atrium for guided manual ablation with a Stereotaxis platform [121] with successful intraoperative results.

1.7.1.2 Advanced Imaging

Real-time feature extraction from sensing sources is crucial for the automation of surgical tasks. Even with state-of-the-art white light stereoscopic imaging, it is still a major challenge to have reliable online understanding of the surgical scene. For this reason, a number of advanced imaging approaches have been proposed, including "plenoptic" vision to retrieve depth from the scene. Plenoptic cameras are equipped with a micro-lens array capable of acquiring different points of view of the same scene in a single acquisition, thus allowing an accurate 3D reconstruction [122]. Tridimensional reconstruction accuracy is strengthen by the possibility of perceiving the light direction by this type of cameras. If combined with fluoroscopy imaging as in [122], plenoptic cameras allow to identify internal hidden structures such as blood vessels and nerves. Although a small amount of contributions is available in literature, interesting results have been obtained so far, such as the completion of a needle insertion [123], Level-3 suturing (as detailed in Section 1.7.3), and vitreoretinal surgery [124]. Alternative imaging technologies currently under evaluation to enhance feature extraction are hyperspectral imaging [125] and TeraHertz vision [126].

Considering the novelty of the field and the promising results achieved, future developments of these technologies will play a crucial role in surgical robotic research.

1.7.1.3 High Level Feature Tracking

Differently from the tracking discussed in Section 1.5.1.1, where the surgical tool was intended as a physical extension of the robotic platform, here we focus on tracking of tools or features that are physically disconnected from the surgical robot.

In the context of suturing, moving from Level-2 to Level-3 without taking advantage of a dedicated tool, as in the work discussed in Section 1.6.2.3, requires the ability to track the suturing needle and thread throughout the execution of the task. In the field of suturing thread detection a combination of color and geometry segmentation [127] are adopted to detect and model the thread as a spline. However, given the thin structure of a suturing thread, basic computer vision algorithm may suffer critical loss of performances in a real scenario where the light condition is insufficient and the environment is cluttered. For this reason, data-driven analysis of images to retrieve the 3D pose of the thread [128], image-based optimization techniques [129] and Markov Random Fields-based solutions [130] are presented to reject such disturbances.

In the context of suturing needle tracking, an effective approach consists of equipping the needle with highly visible markers [131], detectable by conventional white-light cameras. In [86], the detection by means of coloured markers is supported by a custom gripper that reduces the needle mobility.

Another interesting tracking problem is related to reconstructing in real-time the shape of biopsy or injection needles under ultrasound guidance. To address this challenge, motion features are explored in [132], while a Kalman filter is proposed in [133]. Optimization techniques based on gradient descent algorithms are adopted in [134] along with geometric needle models for tracking. Although the results of [132] reports a localization accuracy of $1.70mm$ while respecting the real-time constraint [129], there are no contributions addressing the problem of high level feature tracking in a realistic scenario, thus motivating further investigation in the field.

1.7.2 Navigation of Continuum Surgical Devices

Continuum surgical devices include, among others, steerable needles for biopsy sampling or local drug delivery and cardiovascular catheters. In this section, we discuss robotic platforms pursuing Level-3 navigation of these types of devices.

Robotically controlled needles may introduce a relevant benefit in brain, prostate and lung surgery, where the difficult access to the anatomy increases the complexity of the task. Due to their thin structures and tortuous paths, the manual navigation of these needles is demanding, if not impossible. To enable an effective use of these devices, autonomous navigation is crucial and continuous updates of the external forces acting on the needle are required to safely navigate towards the target. Moreover, as the needle pushes through, the system must compensate the tissue deformation to avoid undesired interaction with peripheral anatomical structures. [135] proposes a kinematic and a mechanics-based approach to evaluate needle-tissue interaction, thus predicting tissue deformation. A crucial aspect of steerable needles is localisation and registration to the anatomy. Ultrasound imaging is widely adopted [136] to develop image-based control strategies: as the needle advances through the tissue, an ultrasound transducer tracks and follows the tip. Alternative imaging approaches used for autonomous needle navigation include intra-operative MRI to localise and avoid obstacles [137] and Fiber Bragg Grating to track the needle tip [138]. Robotic needle guidance is a relatively new approach in robotic surgery, therefore no currently available clinical platform embeds this technology. However, promising results have been recently obtained in human cadaver trials [139] and with the support of preoperative analysis[140], demonstrating a possible translation to Level-4 autonomy in the near future.

In the context of autonomous navigation of cardiovascular catheters, a very advanced work is presented in [141], where force sensing and palpation are adopted to drive an autonomous catheter through blood vessels, up to the heart. The approach is validated by an in-vivo trial, demonstrating performances comparable to the manual execution.

1.7.3 Advanced Suturing

In order to achieve Level-3 suturing, plenoptic cameras have been adopted to extract the 3D profile of the scene and autonomously define the suture entry points in [123]. The algorithm is based on human demonstrations and validated on ex-vivo tissues, showing a superior performance in terms of time and accuracy when compared to a human operator. Point clouds were used in [142] to autonomously plan the needle path, including the entry points. In particular, the region of interest is identified manually by the surgeon to reduce the computational burden, but then the system takes over by extracting all the task specifications autonomously. The system is evaluated on a suturing phantom by assessing the thread tension and the displacement of the entry points. Results show a consistency almost three times higher than a human operator. While current results on advanced suturing are extremely encouraging, they are limited to anatomical phantoms or ex-vivo tissue models. As the approach is translated to a more realistic scenario, the performance of the suturing robot may be heavily affected. From the small amount of literature available, it is clear that full autonomous suturing is still far from being commercially available. Moreover, due to the high complexity of the task, no studies addressing the problem of failure modes, such as the accidental drop of the needle or the entanglement of the thread, have been carried out. Embedding the technologies included in Section 1.7.1.3 could potentially revolutionise the approach to autonomous suturing by providing a robust and continuous tracking of needle and thread, thus allowing the system to consider their presence in the scene.

1.8 Level 4 - High autonomy

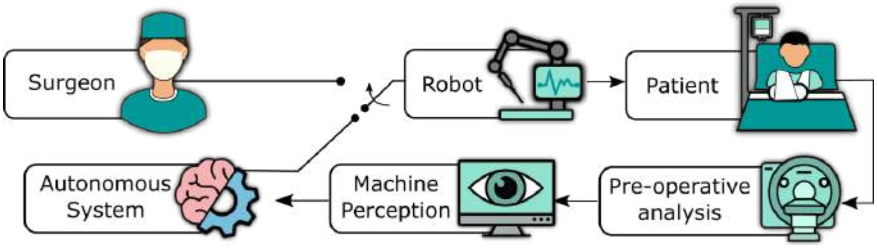


Figure 1.9: In "High Autonomy" systems, pre-operative and intra-operative information are used to devise an interventional plan composed by a sequence of tasks, and execute it autonomously, re-planning if necessary. A surgeon always supervises the procedure and can get back control at any time.

The fundamental characteristic of Level-4 systems is their ability to autonomously make clinical decisions and execute them, under constant supervision by the surgeon.

Beyond clear technical challenges, this level poses very relevant issues in terms of ethical and regulatory aspects.

While concrete examples are not yet available, we can easily see where these systems would clearly contribute to healthcare delivery, i.e. intelligent removal of cancerous tissue, from registration with pre-operative imaging, adaptation of the plan with real-time data, and ablation of cancer while maximising sparing of healthy tissue.

In this Section, we discuss how progresses in organ and tumor segmentation represent a stepping stone towards debridement and tumor resection. This section is then followed by a discussion of ethical and regulatory issues around autonomy in sugical robotics.

1.8.1 Enabling Technologies

1.8.1.1 Organ and Tumor Segmentation

Interpretation of pre-operative imaging (MRI, CT and US) is a requisite for a Level-4 system. Autonomous segmentation of organs such as brain [143], liver [144] and prostate [145] were investigated using different imaging techniques including CT [146], MRI [145]. Tumor profiles can be extracted from CT scans by adopting optimization techniques [145] and deep learning models [146, 144]. Subsequently, the extracted regions are merged together to obtain a 3D model of the target. Usually, the accuracy achieved with the help of these models is very high (mostly over 90%) since the feature to be recognized are highlighted in the capture thanks to the high contrast typical of both MRI and CT scans. These performance are not achievable in endoscopic images where, due to variation of lightning, presence of blood and reflection of the organs, usually the same type of model does not perform at a comparable level of accuracy.

Autonomous segmentation techniques are embedded in commercially available systems such as the Brainlab iPlan [147], which is capable of segmenting MRI scans and integrating them with other imaging techniques like ultrasound and elastography. Using this platform, the surgeon can validate the software segmentation, plan and deliver radiotherapy by means of robotic platforms, i.e. GammaKnife and CyberKnife. Although already implemented and used in the operating room for neurosurgery, autonomous segmentation remains an open challenge for thoracic and abdominal surgery, where tissue deformability prevents satisfactory results to be achieved.

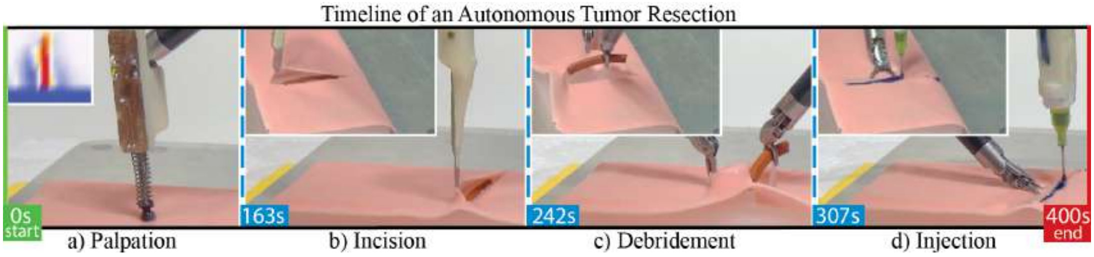


Figure 1.10: Stages of autonomous tumor debridement on a phantom reported in [99]. Initially, a palpation probe scans the tissue to define the tumor boundaries (a). Subsequently, an incision is performed (b) and the tumor removed (c). Eventually, the incision is sealed by means of surgical glue.

1.8.2 Debridement and Tumor Resection

To perform tumor resection, surgeons are required to fuse pre-operative (e.g. MRI, CT, US) with intra-operative (e.g. white-light endoscopy, fluoroscopy, Near Infrared (NIRF)) imaging modalities. Systems with "High Autonomy" must possess a similar ability. In [148], NIRF markers were adopted to retrieve the tumor boundaries and guide resection via electrosurgery, while maintaining a minimum margin of 4-mm. Despite the markers were applied manually, this shows the feasibility of a NIRF-based visual servoing for debridement on phantom, achieving a margin of 3.67 ± 0.89 mm. A similar approach is presented in [149], where the target area is detected by applying coloured markers on a phantom. The debridement is modelled by means of Finite State Machines (FSM), allowing a gesture execution time of 20.8 seconds. Alternatively, a palpation probe [99] mounted on a daVinci manipulator was used to identify and excise a tumor hidden below the tissue surface (Figure 1.10. Subsequently, the wound was sealed with surgical glue [150]. The probe initially scanned the area of interest, localising hard inclusions in the tissue and tracking the tumor profile. Then, a surgical scalpel performed the incision and grasped the tumor. Experimental results on phantoms showed a success rate of 50% in tumor excision. Despite the preliminary results, the system addressed some of the major challenges in surgical autonomy at Level-4 and did not require any intervention from the operator.

1.9 Legal and ethical aspects of autonomous surgical robots

The taxonomy and definition of the levels of autonomy are directly inspired by the SAE J3016[8] standard, which defines the same levels for on-road Autonomous Vehicles (AV). Despite the standard was published in January 2014, the first examples of semi-autonomous cars date back to the 1970's. The sector experienced great advancements in the last two decades with the support of initiatives like the DARPA Grand Challenge, attracting interest from private companies. The recent technical and regulatory advancements have been massive, to the point that the US department of transportation issued a document entitled "Ensuring American Leadership in Automated Vehicle Technologies" and pilot trials of Level-4 and above have started on public roads in US, Canada and Europe. In the context of surgical robotics, the first (baby) steps to move autonomous platforms (Level-3 or higher) out of university labs are in the military field, with the

US department of defense issuing a call for autonomous systems in combat zones [151]. However, in less extreme frameworks, ethical concerns arise regarding the consequences of decision errors and incorrect robot behaviours, potentially leading to serious injuries or even death. Being such a novel and fast-paced field of research, literature on this topic is scarce and mostly speculative. An interesting perspective is given by [152], in which three elements of responsibility are highlighted: Accountability, Liability and culpability. The first element is related to the ability of explaining decisions, which decreases as the system complexity increases and could be addressed by a combination of explainable AI and recording black boxes, similarly to aircraft. The element of liability, much discussed also for AVs, could be addressed by insurance coverage or alternative approaches, like the concept of electronic personhood, introduced by the European resolution of 16 February 2017 [153]. On the topic of liability, other documents issued by the EU discusses issues related to AI and robotics [154]. Finally, culpability (i.e. the possibility of punishing) constitutes the most complex topic, and could pose a significant legal and ethical barrier, having influence on the surgeons' role.

An interesting contribution [155] focuses on the ethical aspects of autopsies, concluding that explainable AI and machine learning could give powerful support to forensic analysis only in a context of human-robot collaboration.

On the topic of patients' perspective, initiatives like the iRobotSurgeon survey [156] tries to assess the public acceptance of autonomous surgical robots, while [157] discusses the issues related to privacy, suggesting to adopt the "contextual integrity" theory.

From the regulatory perspective, notified bodies such as the American FDA, the British Medicine and Healthcare Regulatory Agency (MHRA) and the German Federal Institute for Drugs and Medical Devices (BfArM, in German) do not have specific frameworks for autonomous robots. The FDA currently classifies surgical robots as Class-IIb devices, while implantable and self-activating devices like peacemakers (which have some degree of autonomy, not requiring any human intervention) belong to Class III. One reason for the current classification of robots is the absence of autonomy. On the other hand, Class-III devices are limited to low-complexity and simple functional mechanisms in which the failure modes and operating conditions can be evaluated extensively, thus performing a complete risk evaluation, as required by all the medical devices standards (e.g. ISO 13485, 14791, 62304). Moreover, different approval pathways for Class-II and Class-III devices require significantly increased investment, as thoroughly described in [158]. As discussed in [159], at low autonomy levels, the current legal frameworks could be suitable to evaluate new devices. A common approach to introduce autonomy is to leave the surgeon in charge of activating the autonomous features. Despite being simple, this method greatly limits the effectiveness of the devices.

The use of machine learning algorithms also presents several issues in current regulatory schemes. All the medical devices standards prescribe a development process based on risk evaluation and minimization, but modern deep-learning approaches treat information in such a way to prevent a detailed risk analysis. State-of-the art approaches based on the novel topic of explainable AI could solve this issue.

At high autonomy levels, the robotic systems are supposed to make clinically-relevant decisions. This could introduce another regulatory dilemma: notified bodies like FDA lack the legal authority to regulate medicine, as this practice is usually left to medical societies. The latter, on the other hand, lack the technical competence to dominate complex and continuously evolving technologies such as robotics. A possible solution is proposed in [160], suggesting to include ethical elements in the engineering development process from the very beginning. In this way, any further technical development paired to an ethical discussion in order to speed up the process of general acceptance.

1.10 Background Summary

The development of intelligent machines will be a long and difficult endeavor, marked by a number of incremental steps in which science and technology will drive changes in societal behavior and legislative framework. This is particularly evident in medicine, where novel solutions motivate regulatory changes and societal perception of how healthcare should be offered. Robotic surgery is no exception and the success of the daVinci surgical system demonstrates the gradual acceptance, and now the preference, of new robotic technology with respect to traditional surgical approaches. The addition of reasoning capabilities to surgical robots will require some time, primarily because of the many open regulatory and liability issues. However, if clear benefits are demonstrated, patients will accept and eventually demand devices that can provide additional cognitive and physical support to the surgeons.

This section aimed at providing a first comprehensive mapping of levels of autonomy that could eventually be added to surgical robots, their implementation through enabling technologies, and the translation of these abstract concepts into practical clinical examples (Figure 1.4). So far, only few laboratory experiments have shown a clear advantage of autonomy in surgical robotics when compared to conventional approaches, and no clinical evidence exists yet. However, if research keeps the current pace, positive evidence will soon emerge and build up to an extent that will motivate notified bodies and hospital's ethical committees to consider transition to clinical trials.

This progress will require unprecedented levels of collaboration among engineers, surgeons and healthcare operators to ensure that communication among all actors in the operating theatre is improved by the new technology. Human-machine interaction will be a key factor for the success of autonomy in surgical robotics. Only platforms that possess an effective way to communicate their intent and "explain" their decision to their human companions will find their way into the operating room of the future.

1.11 Tissue Retraction

Starting from the debate about autonomy in surgical robotics we can find a motivation for the following technical contribution of the thesis. As it is possible to notice from Figure 1.4 very few applications reached an adequate level of developing that allows its commercialisation. Augmented reality, Suturing and Tissue retraction are just some examples of such technologies. However, the technical issues related to each one of these tasks is particularly different and it is not possible to find a universal solution for all of them. Augmented reality is a settled and well established technology in other fields and even in surgery many works have been carried out to implement this technology in such a complex and dynamic environment. The same thing is true also referred to suturing: even if the problem is not solved yet, many studies have been carried out to find an optimal solution to this task. On the other hand, minor tasks such as tissue retraction or stiffness mapping are far away from being solved in an automated context. Surprisingly, even if these problems have not been solved yet, very few efforts have been put in this direction. As these problems are not less important than others, more focus is required on these tasks to allow their evolution and development towards commercialisation. For this reason, this thesis develops a framework for autonomous tissue retraction which aims to solve both the perception problem of understanding the surgical scene where the surgeon is operating, and the control problem to guarantee a safe interaction with tissues. The work of this thesis is divided as follows: in Section 2 a paper published in IEEE Robotics and Automation Letters about a feasibility study for tissue retraction is presented, in Section 3 an incremental work regarding the neural network model for tissue flaps segmentation is shown and Section 4 reports a paper where a motion planning framework is developed and adapted to work on a dVRK platform. Eventually, 5 terminates the thesis presenting the conclusions of this thesis.

Chapter 2

Autonomous Tissue Retracion: Surgical Gesture Design

The following chapters detail the technical contribution of this study. As mentioned in the introduction, the main contribution of this thesis consists of a method to assess both the perception and control problems related to the replication of a specific surgical task, namely tissue retraction. In this chapter, the surgical gesture model emerged from interviews carried out with surgeons is reported. From a thorough analysis of the clinicians' opinion, a good practice of tissue retraction is defined and algorithmically described. This chapter details the technical contribution to transfer the surgeon knowledge and common practice to the robotic framework. The test results proved the effectiveness of this method in clearing the clinician's field of view enhancing the visibility of the surgical scene, thus simplifying the anatomy access.

In order to plan an intelligent motion, an accurate estimation of the workspace is required. Once the key features for motion planning are identified, it is crucial to define a way to estimate the possible trajectory of the surgical tools. Finally, a motion planner must be implemented to estimate the joint velocities at every step in order to control the arm motion. The following chapters reports the contents of part of the publications where these problems have been analysed. Chapter 3 presents a comparative study of different neural network structures to accurately identify and segment flaps of tissue in the scene. The models are trained and estimated over a dataset collected at the Anatomy Facility of the Faculty of Medicine at the University of Leeds. In this case, the adoption of spatio-temporal layers in the model architecture proves to be effective in segmenting tissue flaps from depth maps, thus achieving a satisfactory level of accuracy for the next phases which guarantee a solid basis to perform the target task. Chapter 4 describes how the detected features coming from the previous phase are used in order to plan and execute a particular trajectory with the Da Vinci arm. By means of the MoveIt! environment, the PSM arm kinematic model is defined and used to plan and execute smooth trajectories in a 3D space. The chapter's conclusions present experimental validation to demonstrate that the robot is capable of following a smoothed trajectory constrained to the waypoints extracted from the previous analysis.

2.1 Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery – A Feasibility Study

© IEEE. Adapted and reprinted, with permission, from Attanasio, A., Scaglioni, B., Leonetti, M., Frangi, A. F., Cross, W., Biyani, C. S. and Valdastri, P. (2020). Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery–A Feasibility Study. IEEE Robotics and Automation Letters, 5(4), 6528-6535. Supplementary material at 10.1109/LRA.2020.3013914

Authors: Aleks Attanasio, Bruno Scaglioni, Matteo Leonetti, Alejandro F. Frangi, William Cross, Chandra Shekhar Biyani, Pietro Valdastri

Abstract: *In this work, we describe a novel framework for planning and executing semi-autonomous tissue retraction in minimally invasive robotic surgery. The approach is aimed at removing tissue flaps or connective tissue from the surgical area autonomously, thus exposing the underlying anatomical structures. First, a deep neural network is used to analyse the endoscopic image and detect candidate tissue flaps obstructing the surgical field. A procedural algorithm for planning and executing the retraction gesture is then developed from extended discussions with clinicians. Experimental validation, carried out on a DaVinci Research Kit, shows an average 25% increase of the visible background after retraction. Another significant contribution of this paper is a dataset containing 1,080 labelled surgical stereo images and the associated depth maps, representing tissue flaps in different scenarios. The work described in this paper is a fundamental step towards the autonomous execution of tissue retraction, and the first example of simultaneous use of deep learning and procedural algorithms. The same framework could be applied to a wide range of autonomous tasks, such as debridement and placement of laparoscopic clips.*

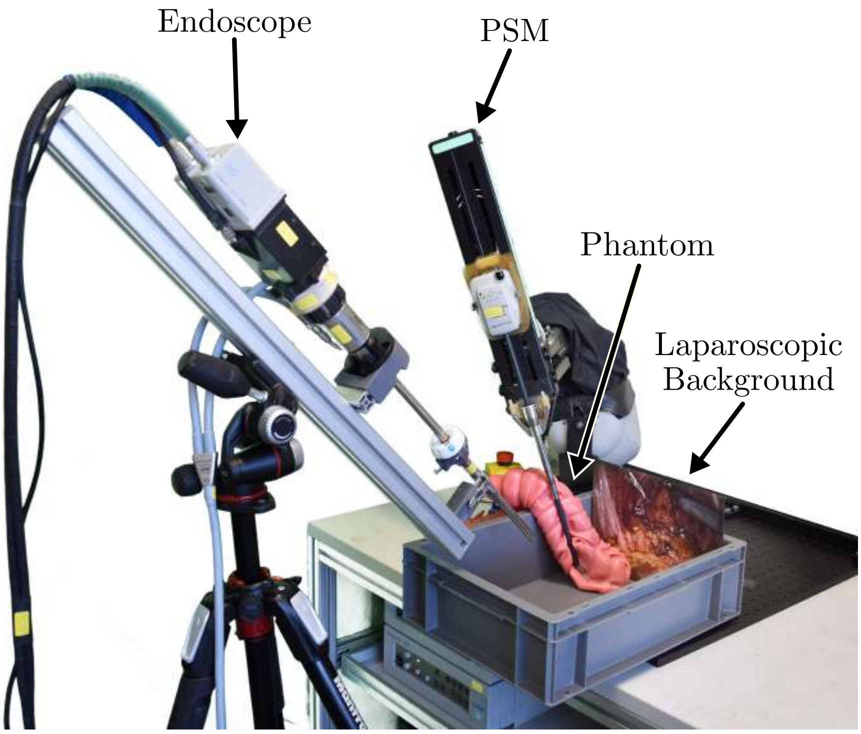


Figure 2.1: DVRK setup composed of a PSM and a stereo endoscope. A phantom and a printed laparoscopic background have been used to validate the semi-retraction approach.

2.1.1 Introduction

Minimally Invasive Surgery (MIS) presents several benefits for patients compared to open surgery, such as reduced trauma to the anatomical structures, shorter recovery time, and reduced blood loss [161]. A significant portion of each MIS procedure is devoted to Tissue Retraction (TR), which is conducted to access the area of interest (e.g. tumour) [162]. Exposing the surgical area is therefore a crucial task in MIS, as surgeons rely mainly on visual information, given that tactile feedback is absent or extremely limited. This is especially problematic in urology, where access to the bladder and prostate is obstructed by bowels and connective tissue [162]. In this clinical practice, robotic MIS is nowadays a common approach, with platforms such as the DaVinci Surgical System (DVSS) from Intuitive Surgical widely used worldwide. The DVSS is a master-slave teleoperated system, i.e. the movements of the surgeon on two Master Tool Manipulators (MTM) are replicated on the tip of laparoscopic instruments by means of three Patient Side Manipulators (PSM). During a typical robotic MIS procedure, the surgeon temporarily assigns one of the MTMs to the third PSM to perform tissue retraction, or requires the support of an assistant to carry out the task with an additional manual instrument. Retraction often involves manipulation of connective tissues or organs (e.g., liver or bowel). Switching robotic arms, or instructing an assistant on the desired retraction motion, significantly increases the surgeon's cognitive load [163] and raises severe risks with potentially catastrophic consequences [164]. TR can also be challenging in the context of manual laparoscopy, where the lack of coordination between surgeon and assistant can lead to hazardous situations, such as instruments collisions, tissue damage or unintentional tearing [165]. To tackle these issues, this paper presents a semi-autonomous system for TR that can be applied to surgical procedures using a robot-controlled instrument (i.e., full robotic MIS or hybrid manual-robotic procedures).

Our approach focuses on detecting tissue flaps obstructing the surgical field by using U-Net [166], a particular convolutional neural network structure, widely adopted in the segmentation of medical images. The network (henceforth: U-Net), fed with the endoscopic video stream, is trained via a dataset of surgical images recorded during procedures performed on Thiel-embalmed cadavers (i.e. an embalming technique that preserves the softness of human tissues [167]), and subsequently labelled manually. An algorithm is developed to identify the retraction grasping point and direction based on the size and shape of the detected flaps. This enables the TR to be planned and then planned and performed autonomously.

This methodology was validated on a DaVinci Research Kit (DVRK) [3] and experiments were performed on a benchtop platform. However, the proposed approach could be applied to any other surgical MIS platform fitted with stereo vision and at least one instrument manipulated by a robot [168].

Research in surgical robotics has recently focused on increasing the level of robots' autonomy, with examples of automating tasks such as suturing [92], and resection [150]. The research on task autonomy aims at relieving the surgeon of manual and repetitive tasks in a collaborative framework, rather than substituting the human action completely [169, 170]. Research in autonomous suturing and related sub-tasks, discussed in [86, 171] has been greatly facilitated by the availability of datasets dedicated to the analysis and automation of surgical gestures (JIGSAWS [172]). The use of automation for 3D tissue debridement of soft tissues presented in [173] is particularly interesting. In order to provide an accurate 3D mapping from the surgical scene, the method proposed in on stereoscopic imaging is proposed in [174] is capable of identifying the tool by means of marker and the tissue by the 3D reconstruction of the stereo pairs. The literature on TR is limited, despite this task being repeatedly performed during all typical procedures. In [94] a simulation framework to perform a grasp-and-retract task is presented along with a path planning method for retraction in the presence of an obstacle is reported. More recently, advanced approaches have been proposed in [95], where retraction is controlled by an image-based system, and in [96, 175], where three different approaches based on proportional control, hidden Markov models and fuzzy logic are developed. In these works, the start and end points of the retraction are manually indicated by the surgeon thus entailing no autonomous planning. Concerning the use of deep learning algorithms in the context of surgical data, the U-Net neural network has been developed for segmentation of biomedical images [166], and subsequently widely adopted in various surgical scenarios such as brain tumour detection [176], liver tumour tracking [146], and surgical tool detection [31]. In [177], segmentation is performed on MRI images, aiming at localising tumours by means of 3D reconstruction. However, U-net has not yet been applied to the detection of tissue flaps for the automation of retraction.

The main contribution of this work is a framework for semi-autonomous tissue retraction, including endoscopic image analysis and gesture planning. This contribution advances the field of robotic-assisted MIS, laying the foundations for future developments in the field of autonomous surgical assistance. Compared to other works in soft tissue retraction, such as [95] and [96], we increase the level of autonomy by providing autonomous tissue segmentation and gesture planning abilities directly on the endoscopic video sequence. Our system is capable of automatically extracting start and end points for tissue retraction, thus reducing the input required from the surgeon in defining task specifications. Other works, such as [178], adopt a similar workflow but focus on a different task (i.e. debridement) and therefore develop algorithms specifically dedicated to debris detection. Another major contribution of the present work is the introduction of FlapNet, a dataset of labelled surgical images dedicated to retraction, available at https://github.com/Stormlabuk/FlapNet. The dataset offers a valuable resource for research in the field of anatomy navigation. The approach described here leverages both deep learning techniques, well-suited to image analysis, and procedural algorithms, which offer the advantage of predictable behaviour and repeatability. The same approach can be adopted to perform other semi-autonomous tasks such as ablation, placement of laparoscopic clips, and debridement.

2.1.2 Materials and Methods

In Figure 2.2, a schematic diagram of the proposed method is represented. The approach is composed of three main elements: Tissue flaps detection (Fig. 2.2-a),

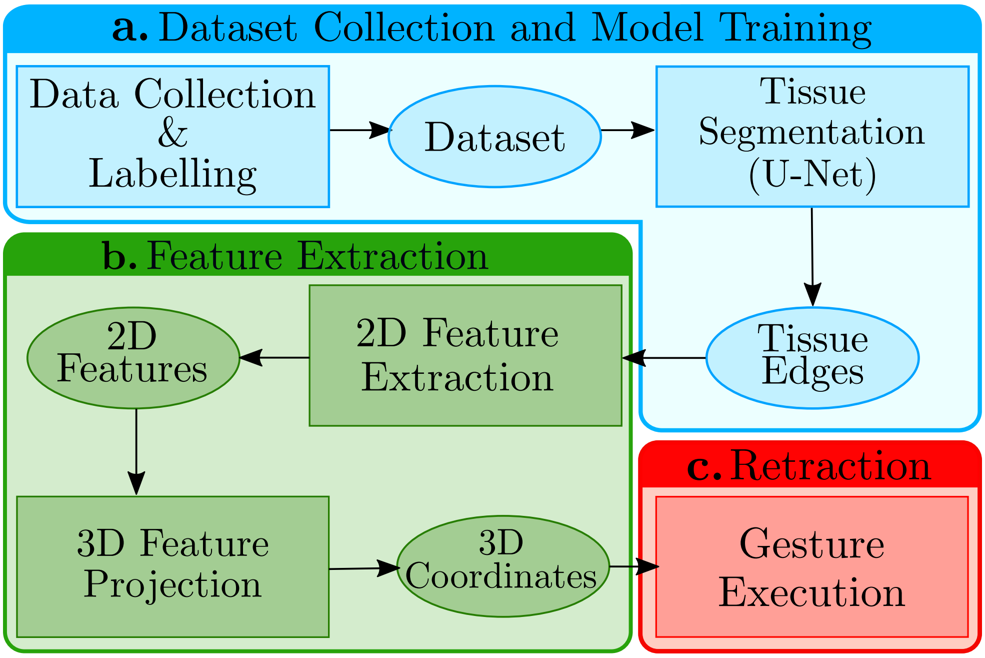


Figure 2.2: Tissue retraction pipeline: a U-Net is trained using manually labelled disparity maps evaluated from stereo images of a cadaveric lobectomy. Subsequently, 2D features such as grasping point, background and tissue centroids are identified on the tissue mask output by the network. Finally, the features are projected in the 3D space by means of epipolar geometry, allowing the DVRK controller to plan and perform the retraction.

extraction of relevant features (Fig. 2.2-b), and gesture planning and execution (Fig. 2.2-c). The output of each stage corresponds to the input of the following stage. In this work, a “detect-plan-execute" approach is adopted to allow the surgeon to maintain control over the execution of the gestures. The system is designed to plan the retraction and subsequently show the surgeon the grasping point, the retraction direction and the final position of the tool. The surgeon can acknowledge the execution by means of a pedal or voice control. The retraction gesture is performed for as long as the surgeon maintains pressure on the pedal. To avoid loss of visual control on the instrument, the camera field of view is mapped on the workspace, and motion of the tool is limited within the image's boundaries, whereby the boundaries correspond to the full-size image cropped by 5%. This restrained the motion to the visible workspace area without limiting the effectiveness of the retraction.

2.1.2.1 Tissue Flap Detection

The initial stage of the retraction process is the detection of the tissue flap to be retracted. This feature is provided by a U-Net developed in the Tensorflow [179] framework. The network is characterised by 5 encoder and decoder blocks. Each encoder, composed of 2 convolutional layers with batch normalisation and

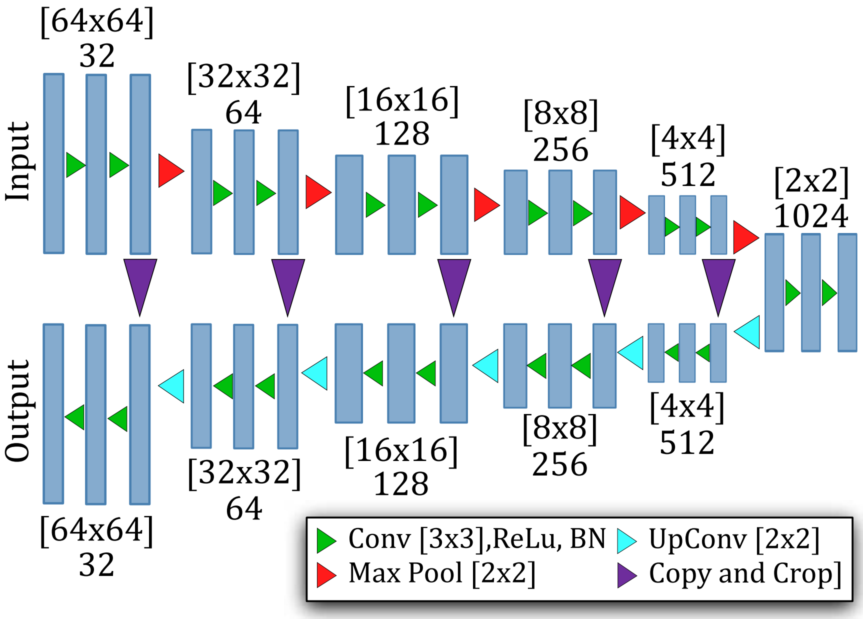


Figure 2.3: Detailed structure architecture of the U-Net model used for tissue segmentation.

a Rectified Linear Unit (ReLU) acting as activation function, outputs into a max pooling layer with pool size 2. The decoder is composed of 3 convolutional layers with batch normalisation and ReLU activation function and the feature map is expanded by a factor of 2. The output is a convolutional layer with sigmoid activation function and 1 neuron. In order to avoid overfitting, dropout is applied to the 3 encoders and decoders closer to the centre of the network. Starting from the first encoder, which includes 32 units, the following encoders are characterised by an increasing number of neurons (i.e. double at every step), to reach a maximum of 1,024 at the centre of the network. Conversely, the number of units per encoder is decreased by a factor of 2 moving from the centre to the output layer. In order to enhance the network capabilities to generalize with respect to different anatomical structures and colours, RGB depth maps (DM) are adopted as input for the neural network. DMs are images [180] in which the intensity of every pixel is associated to a defined distance from the camera lens. In this work, DMs are created base on the disparity between left and right images produced by the DVSS stereo camera. For this reason, they are robust to different lighting conditions and tissue reflections. Moreover, DMs are colour-blind, thus not varying based on the colour of different organs and tissues. As the goal of the U-Net is to detect the candidate flaps for retraction, a grayscale mask of the same size of the input DM is chosen as output, where the value of each pixel, from 0 to 1, describes the likelihood of a tissue flap appearing in that pixel. An example is shown in Figure 2.4.

2.1.2.2 Dataset Collection

In order to create a training dataset for the U-Net, video streams of surgical procedures (lobectomy) performed on a single Thiel-embalmed cadaver by experienced surgeons using a DaVinci Xi have been collected. Starting from stereo image pairs (i.e., left and right cameras), DMs are generated by means of the `stereo_img_proc` ROS package, which is based on a modified version of the Semi-Global Matching algorithm [181], available in OpenCV [182]. Under the assumption of consecutive images being very similar, as the movement of the camera is slow and discontinuous, a three-steps approach is adopted to maximise the variability between images before manual labelling.

- The 356 minutes long video file of the procedure is reduced manually to 62 minutes by selecting the most relevant parts of the procedure where one or more retraction is performed.
- One pair of images is sampled every second, resulting in a set of 3,720 pairs.
- The structural similarity index [183] is evaluated and stereo pairs with a similarity higher than 70% are discarded, thus leading to a dataset containing 368 pairs.

Cameras, with baseline $b = 5$ mm and focal length $f_c = 863$ px, are calibrated by the `camera_calibration` ROS package which uses the OpenCV calibration function, based on [184]. Subsequently, DMs are created for every pair of RGB images using the `stereo_img_proc` package in which rectification is addressed as detailed in [185]. To validate the calibration process, nine calibrations are evaluated and the re-projection error of 0.44 ± 0.06 px is estimated in the projection of the checkerboard points on the image plane. Subsequently, a checkerboard is used to detect four different points showing an error of 7.8 ± 4.4 mm in the 3D estimation.

DMs are manually labelled by means of the MATLAB 2017b Ground Truth Labeler. For a human user, DMs can be difficult to read and understand; therefore, during the labelling process, the user is shown both left and right images in addition to the DM. In order to define a general description of the target tissue flap, during the labelling process the user is asked to highlight the image region containing foreground tissues, if any are present. In every image, two separate labels are created: one representing the tissue flap to be retracted and one representing the DVSS instruments, visible in the scene. Figure 2.4 shows a sample of endoscopic image (on the left), a DM (in the centre) and a label (on

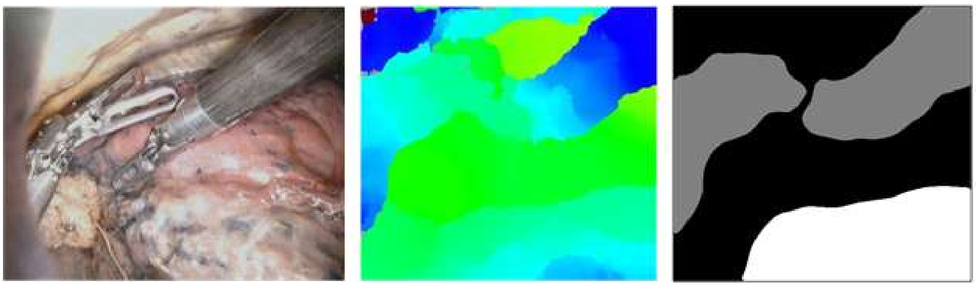


Figure 2.4: Example of tool, tissue and background labelling. The coloured DM is manually labelled to highlight the areas containing either a tool (gray) or a candidate tissue flap (white). Note that when the tool touches any anatomical structure, it disappears from the depth map and merges with the background.

the right). While the purpose of the flap label is to generate the training dataset for the U-Net network, the tool labels are only used to augment the dataset, as described in the following section. The tools' labels are not included in the U-net training set.

2.1.2.3 Dataset Augmentation

The presence of tools in laparoscopic images can obstruct the view and detection of tissue flaps. Moreover, tools introduce a significant disturbance in DMs. In order to enhance the robustness of the U-Net against disturbances generated by tools in the DM, such disturbances must be represented sufficiently in the dataset. An augmentation technique is adopted to improve the network performances. Initially, artificial DMs are generated by extracting the DMs of tools from previously labelled images. Since at the time of the development of this work technologies such as GAN and artificial dataset generation where not particularly established, portions of the DMs corresponding to the tools are overlapped on images in which no tools were originally present, as shown in Figure 2.5. With this technique, the dataset initially containing 368 images is increased to 1,080 images. In addition to this technique, random rotation, flipping and zooming are also applied to the dataset using the Keras library [186], thus obtaining a final dataset of 2,160 images.

2.1.2.4 Model Training

The resolution required to identify flaps is lower than the original RGB images produced by the endoscope. Moreover, high resolution images would unnecessarily increase the time required to train the U-Net. Consequently, size of input and target images are reduced from 506x466 (DM valid window) to 64x64, thus allowing for faster training. The network is trained for 200 epochs with a learn-

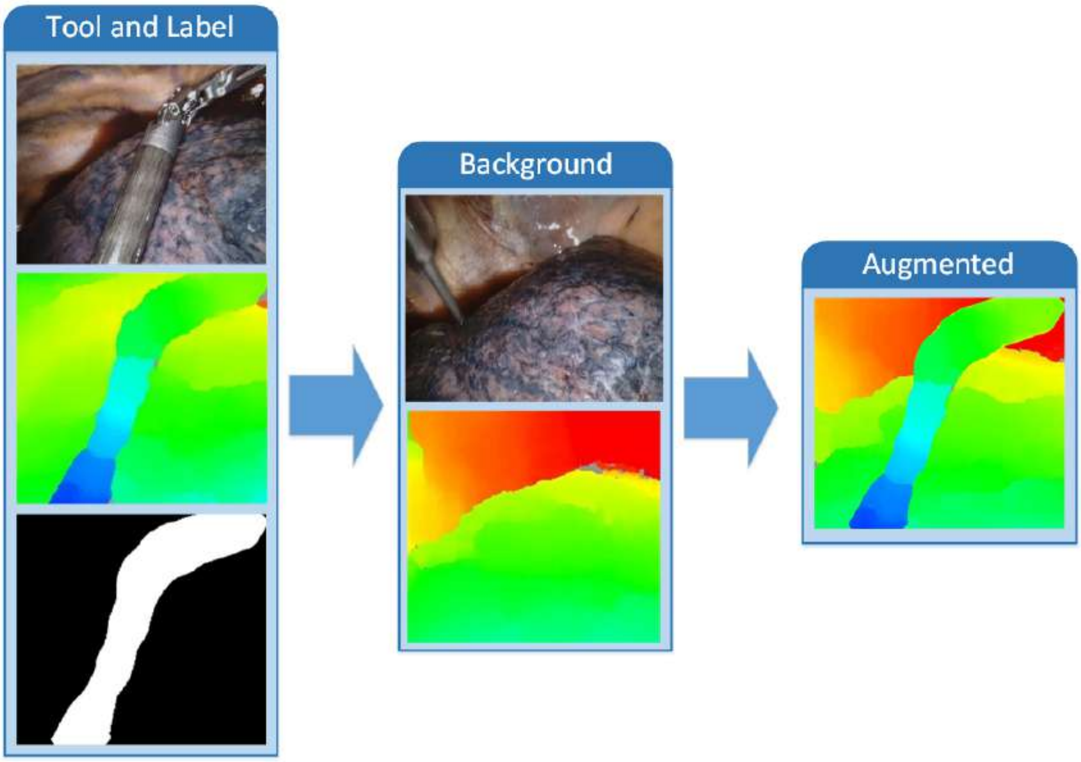


Figure 2.5: Augmentation algorithm pipeline. The tool depth map is extracted from the scene (on the left) and superimposed on a depth map where no tools are present or visible (centre). The result is a new image (on the right) which is added to the dataset.

ing rate of 0.001 and a batch size of 30 images. The Dice loss function [187] is adopted to compute accuracy and the Adam optimiser [188] is used to update the neurons' weights at every epoch.

The augmented dataset is split into a training set (90%) and a test set (10%). In order to assess the robustness of the U-Net against data variability, a training approach based on K-fold [189] cross-validation is adopted. The training process is repeated $K = 10$ times using different subsets of the dataset as training and validation sets.

In Figure 2.6, the performance of the network over the entire training process is shown for the worst (K=1), average (K=2) and best (K=3) performing model. The network accuracy, defined as the pixel-wise difference between the ground truth and the network prediction, is 80.9% ± 1.3% over the K repetitions during the validation phase. The model performance is computed by means of the precision P, defined as $P = \frac{TP}{TP+FP}$ where TP and FP are the true and false positives over the test set respectively. At the end of the training phase, an experimental value of $P = 72.6\% \pm 1.9\%$ is obtained. The network is fed with 64x64 colour depth maps and it outputs 64x64 grayscale masks, with an inference time lower than 42 ms (24 FPS), as measured during the experimental validation

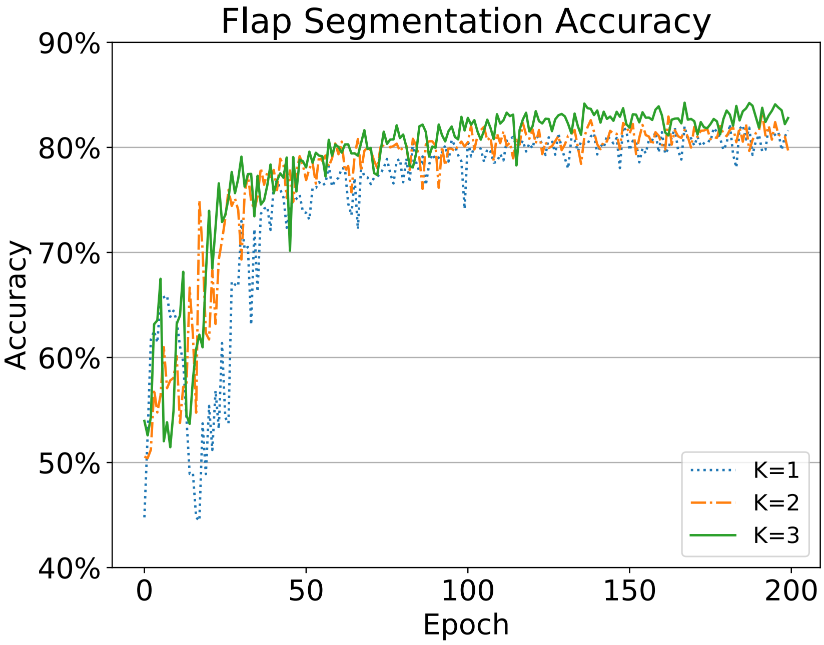


Figure 2.6: Accuracy during testing of the K=10 models considered for K-Fold cross-validation. To simplify the data visualisation, only the worst (K=1), the average (K=2) and the best (K=3) cases are shown.

phase. The pixel values in the output masks represent the confidence (between 0 and 1) used by the network to identify either the background (0) or the tissue (1). Among the possible detection errors that can affect the U-Net, false positives present the highest risk. In order to reduce the number of false positives, pixels with a confidence value below 80% are classified as background by setting their value to 0 in the tissue mask. The output mask is thus binarised, reducing the noise in the prediction.

2.1.2.5 Gesture execution and planning

After a candidate flap of tissue is identified, the retraction must be planned and subsequently executed. In order to reproduce the gesture, interviews on the standard best practice were conducted with ten experienced clinicians (4 urologists, 3 colorectal surgeons, 2 thoracic surgeons, 1 Ear, Nose and Throat (ENT) surgeon). All clinicians had performed more than 100 robotic surgeries.

From the interviews, the following guidelines emerged:

- The tissue is not grasped; rather,it is mobilised by using the rounded side of the instrument in order to minimise the risk of tissues damage and bleeding.
- The area of interest is the centre of the endoscopic image; therefore, retraction aims to clear the central area from obstructing tissue.

Algorithm 1 Retraction planning and execution

```
if tissue not detected then return
else
    $(CT, CB, tissueBorder) = readFromImage();$
    $(sl, inter) = computeLine(CT, CB);$
    $GP = intersection(tissueBorder, sl, inter);$
    $GP = get3DProjection(GP)$
    $(X, Y) = findIntermediatePoint(GP, CB, 25\%)$
    $Z = getQuote(CT) * 1.1$
    $moveTo(X, Y, Z)$
    $align(Z)$
    $OpenGripper()$
    $Z = getQuote(CB)$
    $moveTo(X, Y, Z)$
    while $toolVisible() \vee commandPressed()$ do
        $moveAlong(slope)$
```

- Instruments approach the surgical area following the direction of the endoscopic view, so to avoid unintentional contact with tissues.

- A suitable point where the instruments approach the tissue is the most central part of the flap, and the retraction is usually performed within the visible area by moving the tissue towards the border of the image.

These guidelines are formalised in the pseudocode reported in Algorithm 1. Based on the labels generated by the U-Net, a set of geometric features is defined (*readFromImage()*). Subsequently, the retraction trajectory is generated (*computeLine(), findIntermedatePoint()*. The cartesian coordinates, shown in Figure 2.7, are assumed to be in the camera frame - X and Y correspond to the width and height of the image, while the Z coordinate is the depth of the scene based on the direction of the endoscopic view. On the X-Y plane, the centroid of the background (CB, red) and the centroid of the tissue flap (CT, blue) are computed on the b/w image generated by the U-Net, as shown in Figure 2.8. The grasping point (GP, green) is computed as the intersection between the line connecting the centroids and the border of the tissue (*instersection()*). The 3D position of the aforementioned points is computed by projecting their 2D values on the depth map by applying $Z = \frac{f_c \cdot b}{d}$, where Z is the distance from the camera frame, d is the disparity value of the point, while f_c and b are the camera focal length and baseline respectively (*get3DProjection()*).

Initially, the tool is positioned as follows:

- On the X-Y plane, the tool is positioned in an intermediate position be-

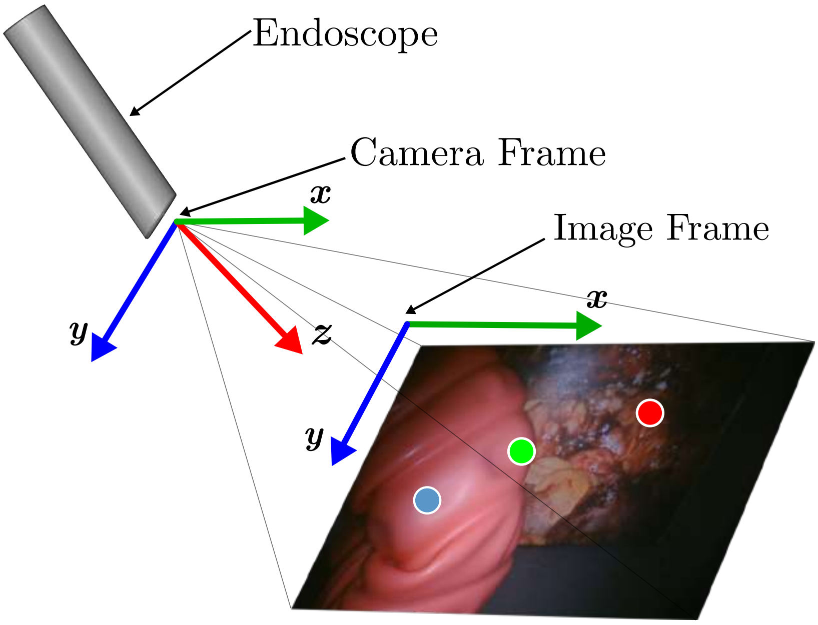


Figure 2.7: Representation of the endoscope frame: the X-Y plane of the camera is parallel to the image frame, while the Z axis represent the distance from the origin of the camera frame.

tween GP and CB, namely at 25% of the distance in the direction of the GP.

- On Z, the tool pose is set to a z-coordinate evaluated as 0.9 times the distance between z_{CT} and the camera frame origin, in such a way that it avoids contact with the tissue.
- The tool is aligned along Z.

Subsequently, the tool is moved along Z to the depth of the background and, then, along the direction defined by the line connecting CT to CB. The gesture terminates whether the surgeon releases the pedal or if the tool approaches the boundaries of the image.

2.1.3 Experimental Validation

2.1.3.1 Experimental Platform

In order to test the approach described above, an experimental platform is used, consisting of the simplified setup shown in Figure 2.1. A silicone phantom representing a colon is extracted from a training platform for colonoscopy (Kyoto Kagaku M40). A section of the phantom is placed on a background image representing the surgical scene, simulating the presence of a tissue flap (i.e. the large bowel) obstructing the surgical view. The network is fed with DMs. Hence, the difference between the surgical images of the training set and the experimental

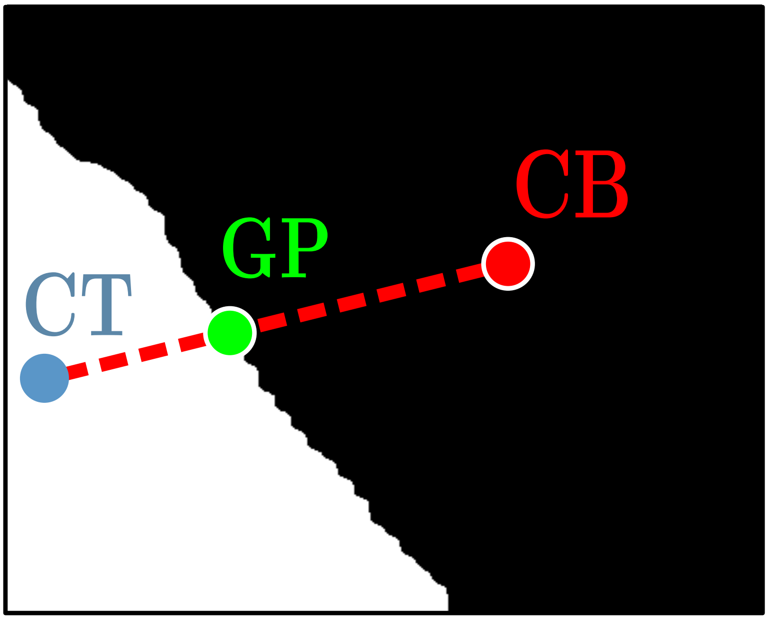


Figure 2.8: Feature extracted from the output mask of the U-Net. The tissue (CT) and background point (CB) are estimated as centroids of the areas representing the two classes: tissue flap (white) and background (black). The intersection between the line connecting CT to CB and the edges of the tissue defines the grasping point (GP).

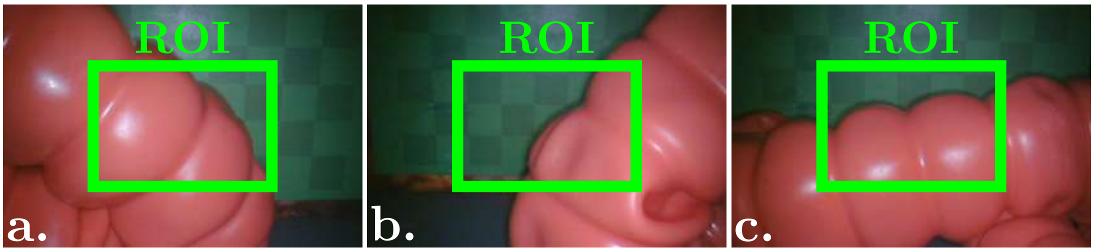


Figure 2.9: Examples of initial conditions in the three retraction cases: from the left (a), right (b) and bottom (c). The region of interest (ROI) are highlighted in green.

scene has a minimal impact in terms of tissue detection. The platform is placed into a plastic box ($36 \times 26.5 \times 11$ cm) to simulate the restricted area available in the abdominal cavity.

Three different scenarios where the artificial bowel segment is placed on the left (Figure 2.9a), right (Figure 2.9b) and bottom (Figure 2.9c) of the scene are investigated to validate the performance of the flaps detection system as well as the trajectory computation. Every test is repeated 5 times. These scene have the role of solely represent the situation and don't take into account any real case disturbance such as blood and organ reflections.

The goal of the retraction is to remove tissues obstructing the scene of interest. Hence, a quantitative approach to assess the quality of retraction consists of measuring the area of background image visible after the action is executed. In order to validate the proposed approach, a green checkerboard is superimposed on an endoscopic image of the abdominal cavity, as represented in Fig. 9. The number of visible green background pixels before and after the retraction is evaluated by adopting a Hue Saturation Value (HSV) filter, used as a metric to assess the quality of the procedure. The test is then repeated for a sixth time with the background image without the green checkerboard (i.e. Fig. 9a) to verify that results are comparable. In order to measure the visibility before and after the retraction, the number of visible pixels after the retraction is compared to the number of pixels of an optimal image where no tissues occludes the scene. The same tests are repeated with a background image representing a laparoscopic view, to demonstrate that the algorithm relying on depth maps is affected neither by the presence of the checkerboard, nor by the background.

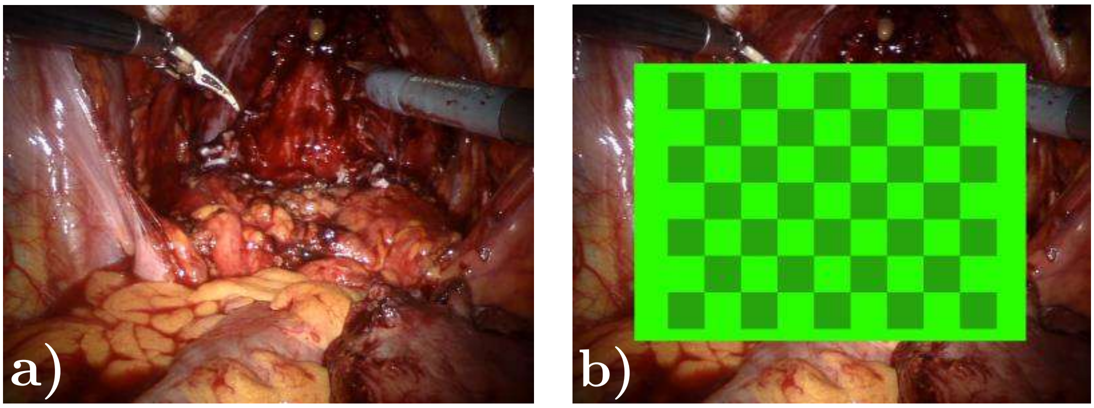


Figure 2.10: Different backgrounds used during the tests. The original endoscopic image of abdominal organs (a) and a version with a superimposed green checkerboard (b), used to quantify the amount of background visible before and after the retraction.

The hardware setup is composed of a single PSM and a stereo endoscope, as shown in Figure 2.1. Regarding the computing nodes, a Robot Operating System (ROS)-based network of two computers is used.

The DVRK low level controller, including joint control loops, is installed on a Linux PC (Control PC in Fig. 2.11) with a ROS interface. This machine is equipped with an Intel Core i5-6400 (2.70 GHz) CPU, HD Graphics 530 and 16GB DD4 (2666 MHz). The computation of the disparity map, the tissue detection U-Net, the feature extraction and the gesture controller are deployed on independent ROS nodes running on a separate machine (Graphics PC in Fig. 2.11), to prevent instability of the computer running the real-time DVRK controller. The calculator is equipped with an Intel Xeon Gold 6140 (2.30 GHz) CPU, an Nvidia Quadro P1000 GPU, and 128 GB DDR4 2666 MHz RAM. The Da Vinci endoscope used during the tests, calibrated via the procedure detailed in Section 2.1.2.2, is different from the Da Vinci Xi endoscope used for data collection. The U-Net model used in the detection phase was previously trained on a separate hardware, using the TensorFlow framework.

The surgeon's attention is usually focused on the centre of the surgical scene. For this reason, a region of interest (ROI) is defined as a rectangle placed at the image centre with width and height of half the entire frame. The percentage of visible background is computed for the entire area and for the central ROI.

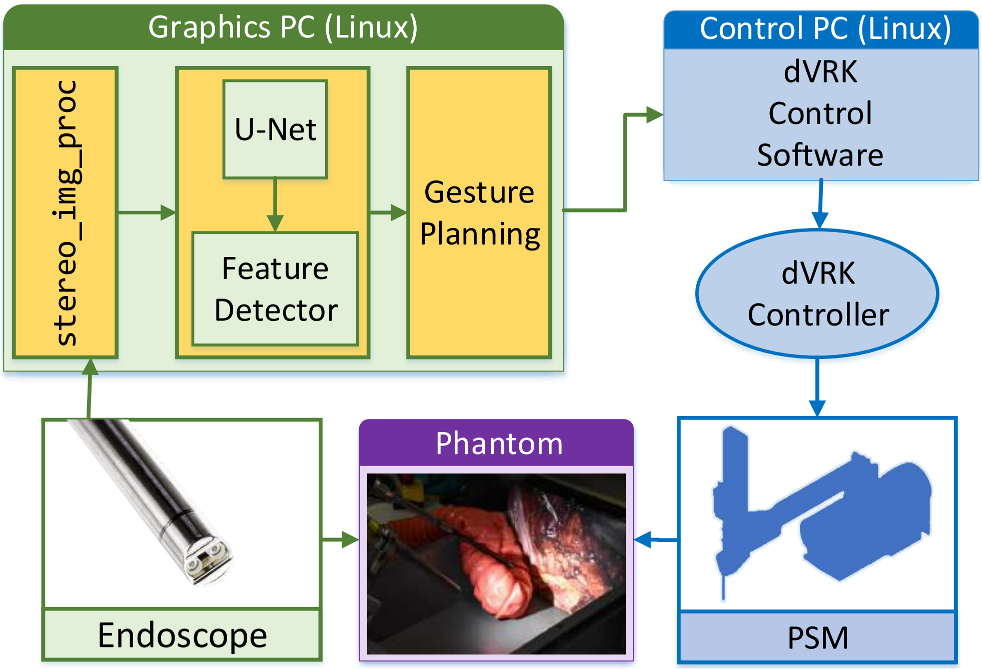


Figure 2.11: Experimental setup: DMs are evaluated from endoscopic stereo images and input to the U-Net which estimates the candidate flaps for retraction. Subsequently, features are extracted from the network outputs in order to plan the retraction gesture. Through a DVRK controller installed on a second machine the control is applied to the PSM which performs retraction on the phantom.

2.1.3.2 Experimental Results

Numerical results are summarised in Figure 2.12. Before the retraction, the visible area is 47.7% ± 4.9%, increasing to 83.4% ± 3.3% after the action takes place. On the other hand, the right retraction presents slightly lower performance, increasing from 54.2% ± 3.4% to 79.6% ± 3.3%. This different performance can be explained considering that the PSM is positioned on the left side of the surgical scene, thus performing opposite movements in the two different scenarios. This result suggests that, despite the great dexterity of the DVSS arms, the placement of the PSM with respect to the scene may influence the effectiveness of the retraction.

The worst performance is displayed by the bottom retraction, going from 41.1%±3.0% to 55%±2.8%. This performance decrease is due to the positioning

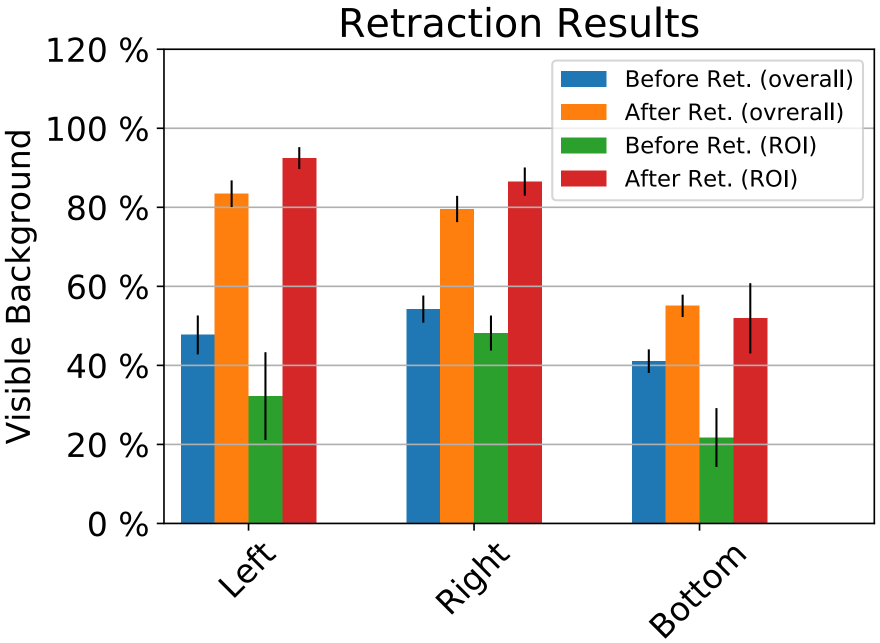


Figure 2.12: Field of view enhancement on the entire endoscopic scene expressed in percentage of visible background before and after retraction, accounting for the entire background and the ROI. The performance is calculated as the means over 5 repetition of the three different retraction cases.

of the arm, which, similarly to the right retraction case, is subject to a constrained motion. Moreover, the orientation of the arm does not allow the instrument shaft to mobilise the tissue, thus reducing the portion of tool capable of exerting force to the tool tip. The results show that this approach can lead to significant and replicable results. Since the proposed method is new, our results are not comparable with other studies.

The trajectories executed by the DaVinci instrument in the left and right retractions are reported in Figure 2.13. The solid blue and dashed red lines represent the experiments with and without the checkerboard used as background (Figure 2.10b and 2.10a), respectively. The start and end points are shown in green and cyan respectively. The red and blue trajectories are very similar, confirming that background does not significantly affect the task execution. In the trajectories, the different stages of the gesture execution are clearly visible.Although the retraction is planned and executed in separate steps, the last sections of all the trajectories (towards the cyan dot) are very similar and grouped in space, demonstrating that the approach is stable against disturbances and small variations between repetitions. All reported experiments were terminated when the tool reached the edge of the image. Moreover, the different start points (green dots) influence the initial part of the trajectory, before the contact between the tool and tissue takes place. It should be noted that the accuracy of

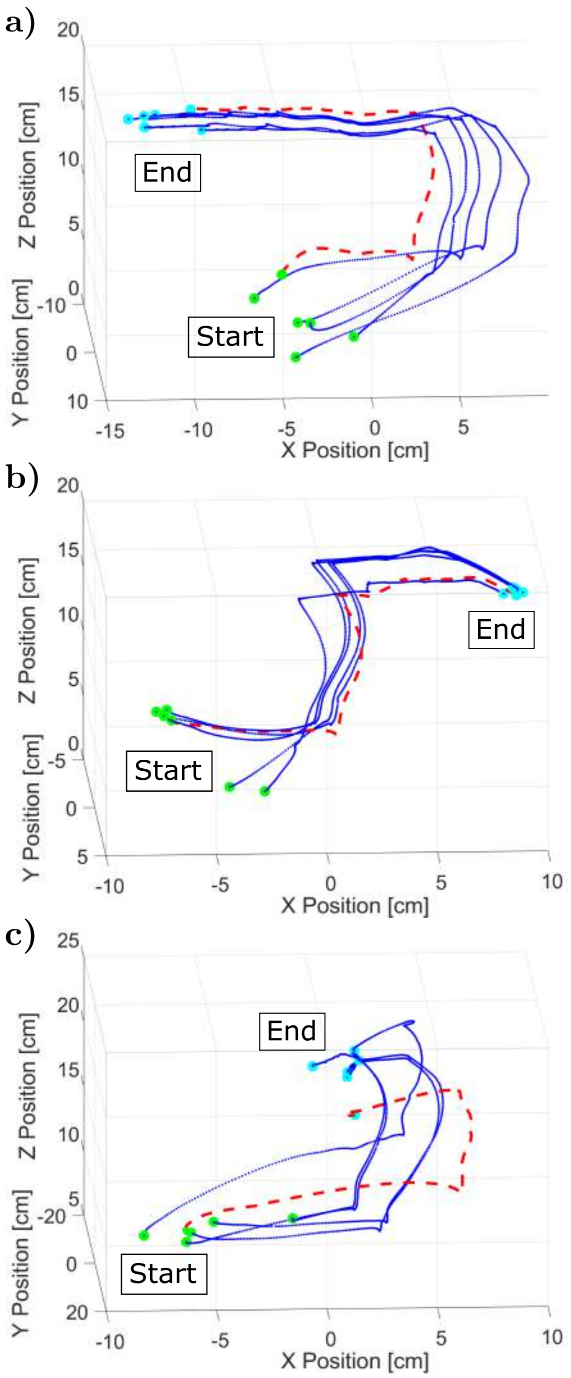


Figure 2.13: Tool trajectories during the left (a) right (b) and bottom (c) tissue retraction. The tool starts retracting the phantom tissue from a random position (green). The retraction ends when the tool reaches the edges of the field of view (cyan). Trajectories obtained using the checkerboard background (Fig. 2.10b) are plotted in blue, while the control experiments performed with the endoscopic background (Fig. 2.10a) are plotted with a dashed red line.

the trajectory execution is completely dependent on the low-level control of the DVRK and is therefore beyond the scope of this work.

2.1.4 Conclusions

A novel method for the semi-autonomous planning and execution of tissue retraction is proposed. The combined adoption of deep neural network techniques for image analysis and procedural algorithms for gesture planning is shown as a feasible approach for the execution of tissue retraction in robotic MIS. Planning and execution of the surgical gesture in the proposed approach can lead to satisfactory and replicable results in a sufficiently controlled environment. The dependability and accuracy of the robot motions offered by this approach can positively impact efficiency. Experimental results show an average increase in the visible area of 25% on the whole image and of 42.9% on the ROI. In order to conduct the flap detection stage using a deep learning algorithm, a novel dataset of labelled endoscopic images is developed and released to the community.

To ease the requirement for extensive manual labelling, future developments will concern the adoption of weak labelling [190], unsupervised learning [191] or generative adversarial networks [192] for image segmentation. Improvements in tissue detection may also include procedure-specific detection of organs and the extension of our dataset to images not containing any candidate tissue for retraction. With minor modifications, this will allow to identify when retractable tissue is present in the scene. Detection of large bowel in prostatectomy and liver in cholecystectomy may be beneficial to adjust the parameters of the retraction. Advancements to the procedural algorithm for gesture planning and execution will involve validation on ex-vivo cadaveric models performed by expert surgeons. In addition, further developments to provide a smoother interaction will involve real-time update for the gesture trajectory. The system has been designed in such a way that a clinical DVSS, including the left and right MTMs and two PSMs, can be controlled independently by the surgeon, while the third PSM can be connected to the DVRK control system. As a result, the system can be integrated into a cadaver test for further validation. The system could be combined with a manual laparoscopic procedure or other robotic platforms, where a robotic arm could be used to perform the gesture while the surgeon is operating with conventional instruments.

The main focus of this work is the removal of obstructing tissues in a static scene, which is a simplifying assumption in a realistic scenario. Consequently, future developments should address maintaining the visibility of the surgical area in a dynamic scene and achieving a more accurate depth estimation, possibly by integrating additional sensors and pre-operative analysis. In particular, the online evaluation of the visible area is a promising development and will provide an additional step towards its adoption in realistic scenarios. Although the approach described here is developed to reduce interaction with the surgeon, the user interface ergonomics should be considered in the future. A simple yet effective method for displaying the flap and retraction direction is especially required, in conjunction with a robust method for receiving the surgeon's acknowledgement.

Chapter 3

Autonomous Tissue Retraction: Perception

3.1 A Comparative Study of Spatio-Temporal U-Nets for Tissue Segmentation in Surgical Robotics

Authors: Aleks Attanasio, Chiara Alberti, Bruno Scaglioni, Nils Marahrens, Alejandro F. Frangi, Matteo Leonetti, Chandra Shekhar Biyani, Elena De Momi and Pietro Valdastri

Abstract: *In surgical robotics, the ability to achieve high levels of autonomy is often limited by the complexity of the surgical scene. Autonomous interaction with soft tissues requires machines able to examine and understand the endoscopic video streams in real-time and identify the features of interest. In this work, we show the first example of spatio-temporal neural networks, based on the U-Net, aimed at segmenting soft tissues in endoscopic images. The networks, equipped with Long Short-Term Memory and Attention Gate cells, can extract the correlation between consecutive frames in an endoscopic video stream, thus enhancing the segmentation's accuracy with respect to the standard U-Net. Initially, three configurations of the spatio-temporal layers are compared to select the best architecture. Afterwards, the parameters of the network are optimised and finally the*

results are compared with the standard U-Net. An accuracy of $83.77\% \pm 2.18\%$ *and a precision of* $78.42\% \pm 7.38\%$ *are achieved by implementing both Long Short Term Memory (LSTM) convolutional layers and Attention Gate blocks. The results, although originated in the context of surgical tissue retraction, could benefit many autonomous tasks such as ablation, suturing and debridement.*

3.1.1 Introduction

Compared to open surgery, Robotic Minimally Invasive Surgery (rMIS) provides substantial benefits to the patient, such as reduced blood loss, decreased tissue trauma and shortened post-operative recovery. Although manual laparoscopy offers similar advantages, the skills required to perform complex procedures with manually manipulated instruments demand expensive and time-consuming training for surgeons. The use of such instruments significantly increases the cognitive load, with potential negative effects on the procedure outcomes. For these reasons, rMIS became popular in surgical disciplines with limited anatomical access, such as urology, gynaecology and thoracic surgery and is gaining momentum in other practices like Ear-Nose-Throat (ENT) and gastric surgery. Significant portions of rMIS procedures consist of dissecting and mobilising healthy tissues to reach the diseased area. During this phase, the surgeon heavily relies on the assistant to clear the surgical field from obstructing tissues, facilitating the surgeon's navigation in the anatomy.

The coordination between surgeon and assistant can be difficult and requires highly specialised personnel. Immersive consoles, such as the one in the Intuitive Surgical DaVinci robot, limit the communication between members of the clinical staff. In particular scenarios such as newly formed teams or lack of adeguate training on emergency situations, the limited communication could increase the risk of adverse events. Some robotic systems, (e.g. the DaVinci robot), allow the clinician to operate three arms, thus reducing the need for external assistance, but the switching process could increase the cognitive load on the clinician [163], particularly for less experienced surgeons.

A semi-autonomous assistance system, capable of operating one arm of the surgical robot and supporting the clinician during the manipulation of soft tissues would solve many issues and open the way for a shared control paradigm, in which the clinician can rely on the robot to perform minor repetitive tasks and focus on the clinical aspects of the procedure. The first step towards the autonomous execution of surgical tasks is the analysis of the scene. The autonomous system must segment the endoscopic scene and isolate the tissue flaps

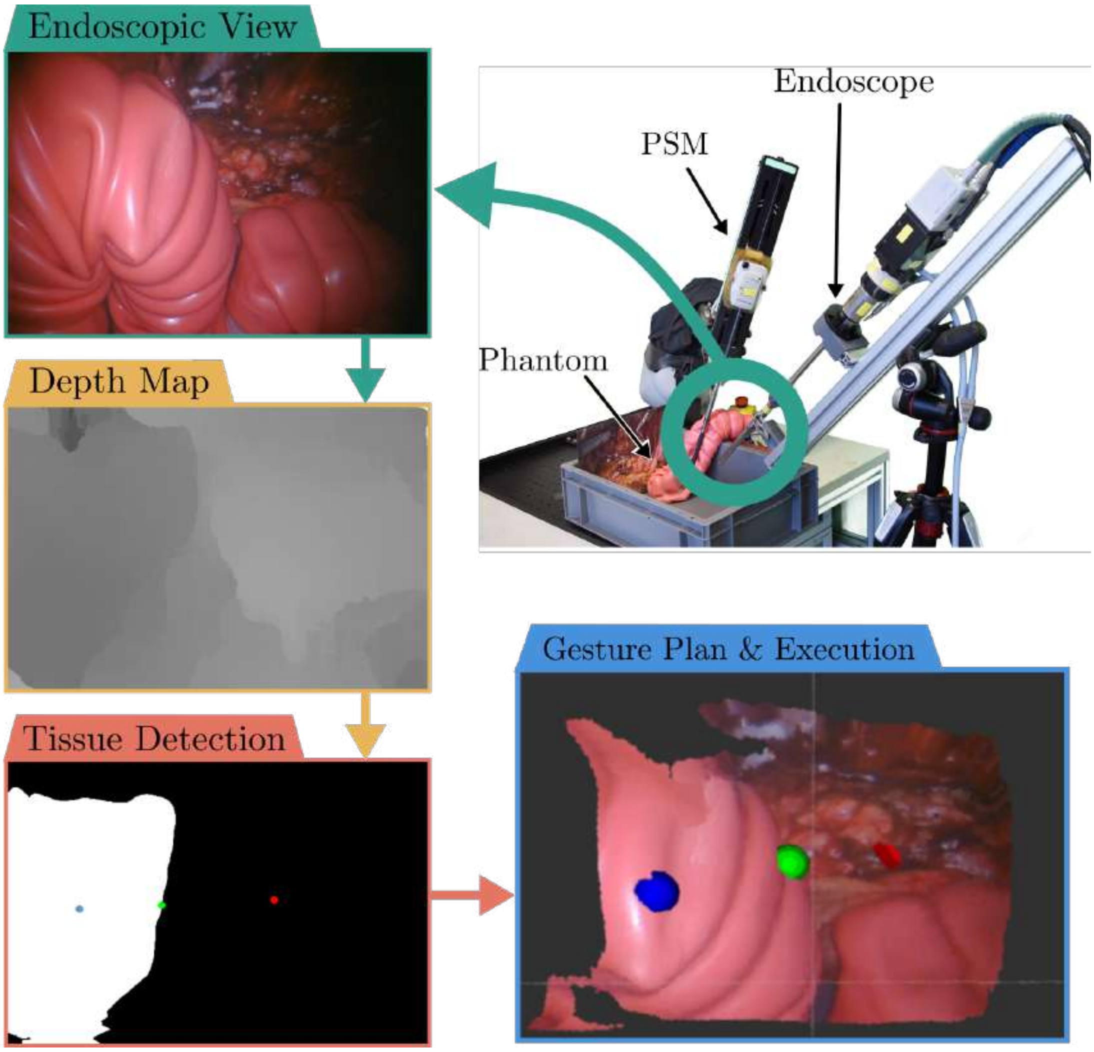


Figure 3.1: Tissue flap segmentation workflow. The stereo images acquired by the endoscope are combined to evaluate Depth Maps fed into a neural network to detect the shape and boundaries of the tissue flap. The tissue flap profile is used to define three waypoints which are used to plan the retraction gesture.

that can be manipulated to plan and execute the gesture. This is a crucial step in the accomplishment of many tasks, as any lack of accuracy at this stage could negatively affect the execution of the gesture and possibly lead to hazardous situations. For this reason, it is extremely important to provide an accurate segmentation system, capable of offering the best possible performance.

In previous work [4], we proposed a feasibility study on autonomous tissue retraction, developed on a DaVinci Research Kit (DVRK). To detect a candidate flap of tissue for the retraction, a single endoscopic Depth Map was segmented with deep-learning techniques, and the system was autonomously executing the retraction, based on the analysis of the image. The experimental setup is shown in Figure 3.1: the images captured by the endoscopic stereo-camera were segmented by means of a deep neural network (i.e. the U-Net [166]), the result of the segmentation was subsequently used to define starting and end point of the retraction. Although the images processed by the system were part of a video stream, the segmentation stage was performed on a single image, thus discarding the obvious relation between consecutive images in the stream. This approach neglects the information provided by the relation between consecutive images and therefore is sub-optimal, with negative consequences on the performance of the segmentation and of the whole task.

In this work, we propose a new approach to the segmentation of soft tissues in surgical endoscopic video streams. We take advantage of the correlation between consecutive images and demonstrate that, by considering sequences as an alternative to single images, the segmentation system outperforms our previous architecture. The main goal of the work is to provide a method to segment soft tissues in abdominal surgery. The approach, based on deep-learning, could be applied to a wide range of surgical tasks and is suitable for real-time tracking of the tissue motion. The main technical contribution of this work is the development of three deep-learning network models for video stream segmentation. Starting from a standard network architecture such as the U-Net, we combine the use of Long Short-Term Memory (LSTM) [193] and Attention Gate blocks [194, 195] to develop three network variants. Since literature lacks work performed in the same experimental setup and domain, the performances of these networks are compared to our previous work only, the process of parameters optimisation is discussed in detail and the effectiveness of a pre-training stage is evaluated. Additionally, a dataset, based on the FlapNet [4] containing real tissue images, is developed to train and verify the performances of the networks. The dataset, comprising labels and training images, and the code are publicly available for the research community at https://github.com/Stormlabuk/dvrk_ULSTM. Although the techniques described in this work originate in the context of retraction, robust segmentation of soft tissues could be used in developing many autonomous surgical tasks such as ablation [109], resection [110] and suturing [123]. The paper is organised as follows: in Section 3.1.2 the dataset processing and organisation (Section 3.1.2.1), the model architecture (3.1.2.2) and the training setup (Section 3.1.2.3) are desribed. Then, in Section 3.1.3 the performances of the three architectures are discussed. Additionally, a comparison with a pre-trained model [196] and our previous work [4] is carried out, to demonstrate the benefits in adopting LSTM layers and Attention Gate blocks in video segmentation. Section 3.1.4 concludes the paper, summarising the contribution and discussing future developments.

3.1.1.1 Technical Contribution

Despite the great interest on autonomy in surgical robotics, demonstrated by the amount of literature [1], research on soft tissues manipulation is limited. The vast majority of the literature focuses on the automation of tasks [5] involving extraneous elements such as suturing [131], [87] and interventional needle passing [110] [86]. On the other hand, automation of tasks that involve tissue manipulation are challenging due to the complex geometry and compliance of the soft tissues. Few examples of autonomous tissue manipulation are available [95, 96], mostly demonstrated in simplified scenarios with reduced complexity. The main barrier for development of realistic applications is the complexity of the scene, difficult to analyse autonomously. A significant contribution can be provided by machine learning. Techniques based on neural networks are widely adopted for medical and surgical image analysis [197]. Deep Learning models have been employed in medicine for the segmentation from MRI and CT scans [198] of either organs [143, 144] or compromised tissue such as polyps [199] and tumours [146]. The U-Net [166] is commonly used in segmentation of medical images such as the segmentation of blood vessels, brain and skin tumours [200, 201, 202]. This network consists of an encoder-decoder architecture which captures contextual information, simultaneously providing accurate detection of the image features. The main drawback of the standard U-Net is the incapacity to correlate frames in a video sequence, thus not taking advantage of the tissues motion and consequently offering limited performances in continuous tissue manipulation. To overcome this limitation, a simplistic approach could consist in linearly merging several independent U-Nets. However, literature has shown outstanding results with the adoption of recurrent neural network architectures such as the Long

Short-Term Memory (LSTM) cells [193]. LSTM provide memory to the model, thus allowing a representation of the features' evolution in time. Adding LSTMs on top of fully convolutional network proved to significantly enhance the accuracy of video segmentation [203] of street scenes. In medical imaging, LSTMs have been used to predict the growth of tumours from 4D patient's data [204] with a simple encoder/decoder model. LSTM cells have been adopted on top of a U-Net model for cell segmentation, showing a remarkable ability in discriminating both the cell's body and its boundaries from the background [196]. An alternative recurrent structure used for video segmentation is the Gated Recurrent Unit (GRU) [205]. These units, significantly simpler than LSTMs, have been implemented by means of convolutional networks to enhance the precision in prostate [206] and brain [207] segmentation. Additionally, an approach for video segmentation adopting 3D convolutional layers to extract the temporal information from image sequences was recently proposed in [31]. These blocks are particular structures that support the network's training and inference by identifying focus regions of the image where relevant information is contained. These blocks have shown effectiveness in medical image segmentation [194] for pancreas segmentation and classification [208].

3.1.2 Methods

3.1.2.1 Data Setup

The first step in the development of a tissue segmentation system is the conditioning of the input data. Since most surgical robots and advanced endoscopic systems are equipped with stereo-vision, we take advantage of the stereoscopic endoscope by considering pairs of stereo images as starting point. With a modified version of the Semi-Global Matching algorithm [181] implemented in the `stereo_img_proc` ROS package, each pair of stereo images generates a Depth Map (DM). Depth Maps are single-channel images in which pixel intensity represent the distance of each pixel from the camera frame. Distances are computed from the features' disparity in the left and right images. As DMs do not contain light and colour information, their use guarantees robustness against variations of lighting conditions and tissue colours. This aspect is particularly important in this work, as the instruments frequently cross the endoscope field of view during tissue manipulation, therefore, it is crucial to guarantee satisfactory performances in presence of the instruments. Additionally, as the colour information is represented in images with three channels (RGB), DMs allow to work on single-channel images, thus speeding up the training phase. The final model should be

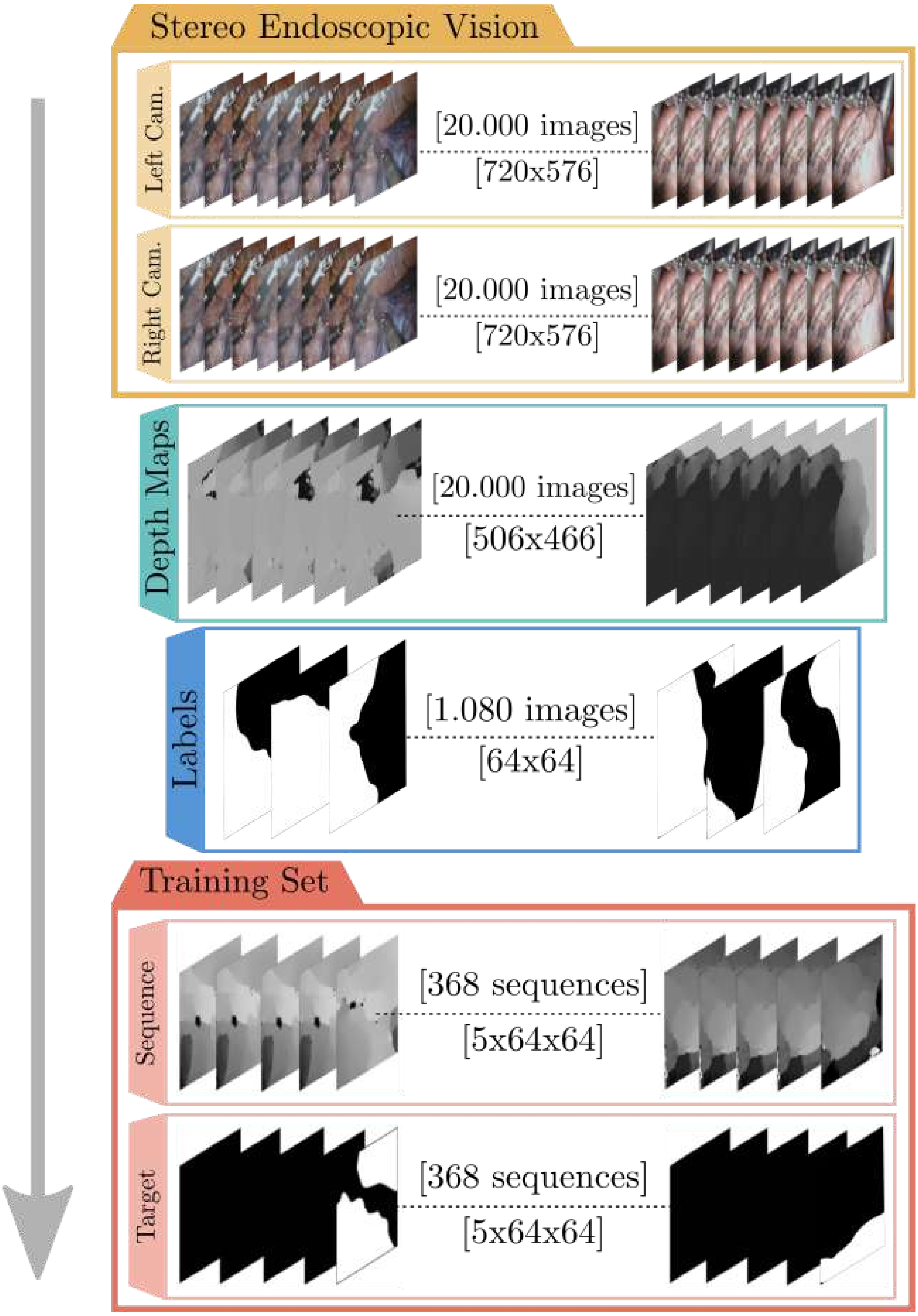


Figure 3.2: The dataset is created from images collected with a stereoscopic endoscope. Depth Maps are evaluated from the stereo pairs and manually labelled. Subsequently, sequences are created by extracting the previous 4 frames from the whole operation video and batching them with the corresponding label of the 5th frame.

able to extract the tissues of candidates tissue for retraction. These tissues are usually located closer to the camera and will present a peculiar shape in the depth maps. However, the detection of candidate tissues may presents problems while the surgeon is operating due to the appearance of tools in the scene which are in general closer to the camera than the tissues. Given the particular geometry of these tools it is easy to filter out their presence and correctly detect the background tissue. In order to properly define the specific appearance of the tissue candidate for retraction, experienced surgeon where interviewed and the dataset was labelled following their direction.

In order to train the networks, DMs must be associated with labels highlighting the areas of the image covered by tissue flaps and by the surgical tools. In a previous work [4], our group developed FlapNet: a dataset of 1080 DMs extracted from images collected during a robotic surgery course, performed with a DaVinci Xi at the University of Leeds, on Thiel-embalmed cadavers [167] by experienced surgeons. Starting from the full stereo video stream of a lobectomy, the most relevant frames of the stream are extracted and labelled: for each DM, a binary mask is created, classifying each pixel as background (0) or tissue (1). The labelling process is carried out by the authors following the direction of experienced urological and colorectal surgeons. Initially, the video sequences containing tissue flaps are identified and isolated. Subsequently, a set of single frames is manually selected. Depth Maps are generated for the identified images. The labelling process is carried out manually on the Depth Map. However, during the process, the user can visualize the RGB image to ease the label creation. Labels with Structural Similarity Index higher than 70% have been discarded to avoid similarity between the dataset entries, guaranteeing a significant variety of samples. To represent the tool's appearance in the endoscopic scene, regions of the DM containing surgical instruments are labelled, extracted from the original DM and superimposed over scenes where tools are not present. The instruments' labels are not available in the FlapNet, as the tissue flaps are the only targets for the segmentation.

The networks developed in this work require a sequence of images. To this end, the FlapNet dataset has been enriched by adding the four frames preceding every labelled image already available in the dataset. To account for this, entries of the original dataset are grouped with the four stereo-frames preceding every labelled image, thus obtaining a set of sequences, in which the last image of each set associated to a binary label (Figure 3.2). Since the majority of the samples (712 images) contained in the FlapNet are artificial images (i.e. created by the superposition of tools on the scene), no preceding frames are available for these

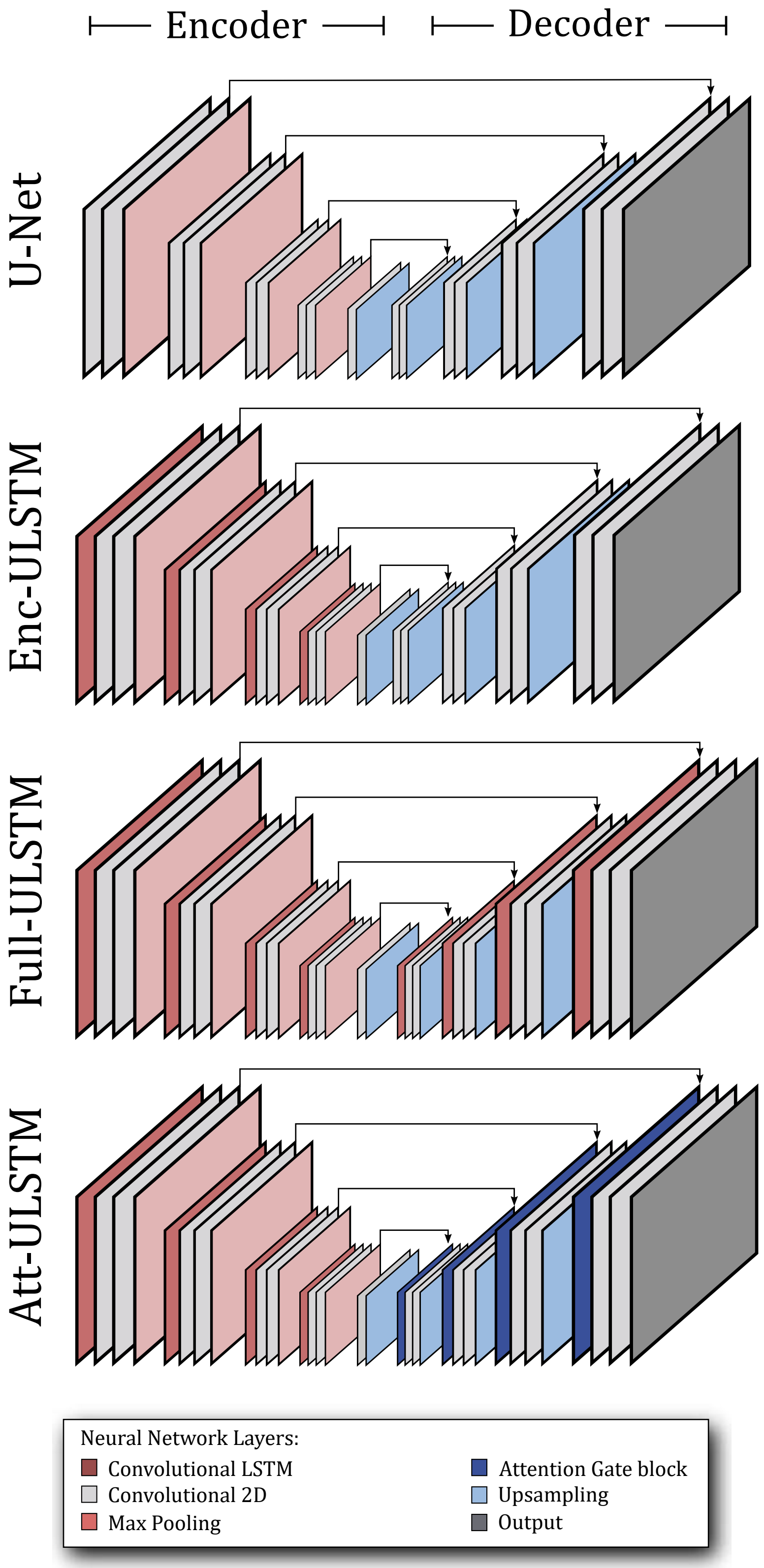


Figure 3.3: Comparison between different extensions of the U-Net[166]. The Enc-ULSTM contains LSTM cells in the encoding branch, the Full-ULSTM model incorporates LSTMs in both branches. Finally, the Att-ULSTM includes Attention Gate blocks in the decoding block.

entries, reducing the size of the dataset to 368 sequences. The images contained in the dataset are reduced to a size of 64x64. During preliminary tests this proved to be a satisfactory compromise between the amount of detail available in the image and time required to train the model. If required by a specific application, the output of the network can be up-sampled and linearised to the original size of the input image. Over the whole set of images, the pixels associated with the background are 70% of the total, leading to a slightly unbalanced dataset. Therefore, particular attention is required during the training phase to limit the amount of predicted false negative and false positive. It is well known in the literature [209] that unbalanced datasets may create issues in the modeling of the less-represented response, leading to a degradation in performance. The original dataset contains only DMs where at least one area is classified as tissue. In order to represent the case in which no foreground tissue is present in the scene, 88 new sequences associated with black mask (only background) are added to the dataset, raising the number of the total sequences to 456.

Given the limited size of the dataset, data augmentation is required. Standard augmentation computer vision techniques are adopted to enlarge the dataset, including: contrast and brightness adjustment, horizontal and vertical flipping, image shifting and rotation. These transformations, randomly selected, are equally applied to every image and label of the sequence to maintain coherence between the input and the target. Moreover, elastic deformation [210] is applied to enhance the variety among the augmented entries by distorting the input image. This technique consists of convoluting two random displacement fields Δx and Δy with a Gaussian filter having standard deviation σ, which represents the elasticity coefficient. The resulting displacement fields are scaled by a factor α that defines the deformation intensity. An additional method for video augmentation comprising the inversion of the sequences' frames to obtain new sequences is herein adopted. This technique allows to create new sequences of images with a coherent time evolution of the scene, thus doubling the number of entries while maintain the correlation within subsequent frames. By means of this augmentation, the initial 456 sequences are doubled to 912, additionally every single sequence is distorted with the aforementioned computer vision techniques up to 3 times, thus increasing the number of entries to 2736 sequences.

3.1.2.2 Neural Networks Development

One of the most common neural network architectures utilised for the segmentation of medical images is the U-Net [166]. Satisfactory performances are reported

Table 3.1: Summary of the implemented feature and architectures in the three different models proposed.

	U-Net	Enc-ULSTM	Full-ULSTM	Att-ULSTM
Conv. Layers	✔	✔	✔	✔
Encoder LSTM	✗	✔	✔	✔
Decoder LSTM	✗	✗	✔	✗
Attention Gate	✗	✗	✗	✔

in literature regarding image segmentation adopting this class of network even with limited amount of data and with high resolution images. As show in Figure 3.4, the network comprises two symmetric encoding and decoding branches, with parallel connections linking the encoders to the decoders. The standard U-Net architecture is suitable for segmenting single images in endoscopic scenarios, as demonstrated by our previous work [4], but cannot correlate consecutive frames (e.g. a video stream) and therefore has limited robustness. For this reason, we build upon the basic U-Net architecture by adding features that implement memory (i.e. recurrency) and take advantage of the relation between consecutive frames to enhance performances. We use recurrent structures such as LSTMs, proposing three network architectures. Additionally, in one of the network variants, the use of attention gates is explored. A summary of the features implemented is reported in Table 3.1.

All the U-Net variants are developed in the TensorFlow [179] framework. The basic structure, identical for all the networks, is composed of 4 encoding and 4 decoding blocks that constitute the contracting and expanding paths, respectively. The encoding blocks consist of 2 convolutional layers with batch normalisation and adopt the Rectifier Linear Unit (ReLU) activation function. Subsequently, a layer with pool size of 2 halves the output size, grouping the features detected by the previous layers to reduce over-fitting, while limiting the memory allocation required. In parallel, the decoding blocks are composed of 2 convolutional layers. The output of each block is up-sampled by a factor 2 with bilinear interpolation, to restore the original image size. The up-sampled outputs are subsequently combined with the feature maps from the encoding branch by means of parallel skip connections. The number of kernels, set to 64 for the encoding block, is doubled for every contraction step in the encoding branch and halved for every expansion step in the decoding branch, resulting in a symmetric structure. To save memory in the training phase, 128 kernels are maintained between the second and third encoder and decoder. The two branches of the network are connected by a single convolutional layer with 512 kernels. The output layer comprises a convolutional layer with a sigmoid activation function.

Starting from the basic structure, we propose three empirically defined variations implementing LSTM and attention gates starting from the structure proposed in [196] for cells segementation:

- Enc-ULSTM: the U-Net model contains convolutional LSTM layers at the beginning of each encoding block.
- Full-ULSTM: the U-Net model contains convolutional LSTM layers in the encoding and decoding branch.
- Att-ULSTM: using the Enc-USLTM as base model, attention gate blocks are added before each decoder block.

In the Enc-ULSTM and Full-ULSTM, convolutional LSTMs are used. The detailed structure of an LSTM is described in Figure 3.4. LSTMs are composed of three gates (forget, input and output) which, combined with the previous cell state c_{t-1}, the previous hidden state h_{t-1} and the input x_t, allow to extract the correlation between subsequent frames, thus rejecting lower-level responses. By means of the forgetting gate contained in the LSTM cells, the non-relevant information at time t is discarded, enhancing the accuracy of the response at time $t+1$. In this particular application, LSTM cells support the network in detecting relevant information such as the position and geometry of a tissue while ignoring and forgetting the appearance of tools. This contributes to the robustness of the network against instruments crossing the endoscopic scene.

In the Att-ULSTM, each decoding block includes a first layer composed of attention gates. In these blocks, capitalising on a gating signal g, the lower activations are discarded, thus allowing the network to autonomously find the relevant areas of the image to focus on, hence resulting in a precise segmentation. The Attention Gate unit takes x_t as input. The gating signal g is applied to every pixel in order to define the focus regions. Three linear transformation W_g, W_x and Σ define the set of parameters of the single unit and are evaluated with channel-wise [1x1x1] convolutions. These blocks contribute to the extraction of focus regions, thus helping identifying the candidate areas of the image where a flap could be found.

3.1.2.3 Models Training

The adoption of convolutional LSTM layers allows the networks to rely on both temporal and spatial features. For this recurrent architecture, a modified version of the Back Propagation Through Time algorithm has been adopted, namely the

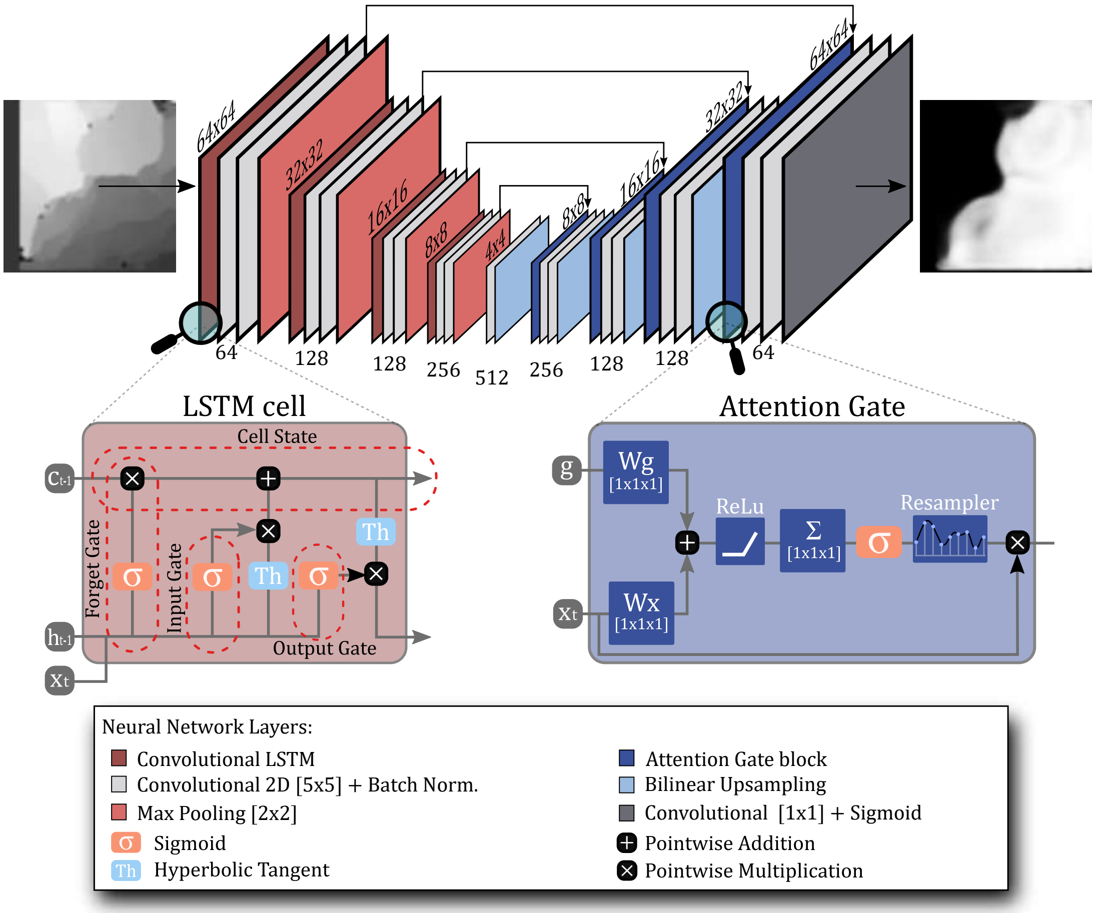


Figure 3.4: The structure of the Att-ULSTM model comprises 4 encoders, 4 decoders and the central block connecting the two branches. Each encoding block is composed of a LSTM cell, two convolutional and one max pooling layers, while the decoding blocks present an Attention Gate block, two convolutional layers and a linear upsampling layer. The output layer is a convolutional layer with sigmoid activation function.

Truncated Back Propagation Through Time [211]. This algorithm, commonly adopted for recurrent networks, periodically updates the gradient a fixed amount of times over the batch. In this work, this parameter was set to $\tau = 5$. Hence, the gradient is weighted on the previous input and hidden state, yielding a simultaneous evaluation of the temporal and spatial features in the convolutional layers. The networks are trained for 10.000 iterations over 650 epochs using the Adam [188] optimiser, capable of managing sparse gradients and preventing noise, as well as vanishing of weak gradients.

A step profile is scheduled for the value of the learning rate, decreasing from an initial value of 10^{-3} to 10^{-5} to speed-up the initial phase of the training. The kernel's weights are randomly initialised with the He uniform distribution [212] which allows to regulate the initial values depending on the preceding layers' dimension, thus reducing the time required for training. Dropout is applied in the LSTM layers and in the central block to limit over-fitting. While standard dropout is implemented for convolutional layers, the same approach is not suitable for long-term memory. As standard dropout applies a mask to the layer to randomly deactivate the neurons, if applied to LSTM cells it would resets the forget gate at each iteration, thus ereasing the cell's memory. For this reason, a recurrent of dropout [213] is applied to LSTM layers to maintain the dropout mask fixed, preventing the loss of memory of the cells. The dataset is split into 75% training set, 15% validation and 10% test set. The models are trained on a Linux (Ubuntu 18.04) machine equipped with an Intel Xeon Gold 6140 (2.30GHz) CPU, an Nvidia Quadro 5000 RTX GPU and 128 GB DDR4 2666MHz RAM.

Two loss functions are compared in this work. The Combo Loss (CL) [214] is the weighted Dice Loss $DL = \frac{2 \cdot P \cdot G}{P+G}$, where G is the ground truth and P the sigmoid output, [187] and the Weighted Cross-Entropy (WCEL) defined as:

$$WCEL = p \cdot -log(\hat{p}) \cdot \beta + (1-p) \cdot -log(1-\hat{p}) \tag{3.1}$$

where p is the ground truth label, $\hat{p}$ is the sigmoid activation of the logits and β is a trade-off factor to foster either false negatives or false positives. The CL is finally defined as:

$$CL = \alpha \cdot WCEL + (1-\alpha)DL \tag{3.2}$$

where α controls the contribution of the single DL and WCEL. Given the unbalanced dataset and considering that for the surgical application false positives must be minimised, we defined $\beta = 0.8$ and $\alpha = 0.6$ to favour the contribution of the WCEL over the DL.

The other function considered here is the Tversky Loss (TL) [215], widely used in medical image segmentation for its ability to train over highly unbalanced training sets. The TL formula is a generalisation of the DL:

$$TL = \frac{2 \cdot P \cdot G}{P + G + \gamma \cdot P \setminus G + \eta \cdot G \setminus P} \tag{3.3}$$

where G is the ground truth, P is the prediction, $P \setminus G = P \cdot (1-G)$ is the relative complement and γ, η are weights to balance false positives or false negatives.

3.1.3 Results

In this section, the performance of the three networks models is evaluated. Four metrics, all aimed at evaluating the ratio between True Positive (TP), True

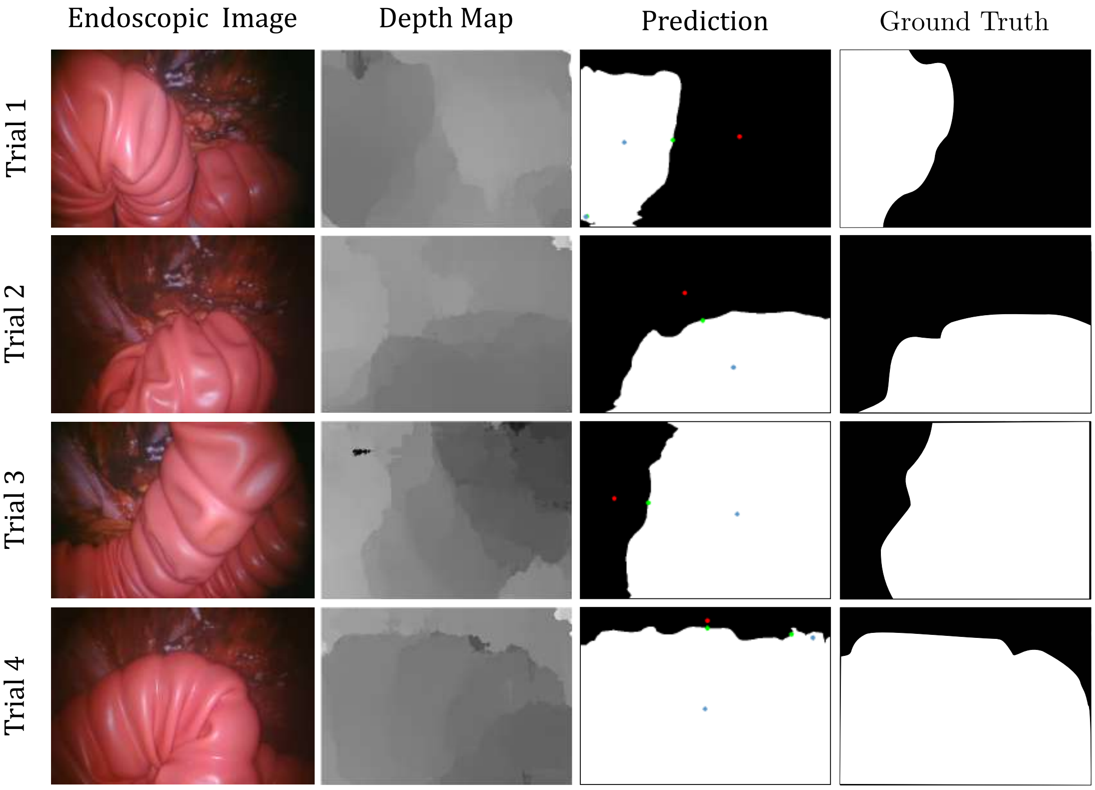


Figure 3.5: Predictions examples of the Att-ULSTM model at the end of the training phase. The tissue is placed in different regions of the endoscopic scene to verify the robust inferring of the model, independently from the tissue position.

Negative (TN) and False Positive (FP), False Negative (FN), are proposed:

- The Precision: $P = \frac{TP}{TP+FP}$, represents the capability of the algorithm to reject false postives.
- The Recall: $R = \frac{TP}{TP+FN}$ describes the sensitivity of the network in detecting TP and TN. Combined with Precision, it provides a reliable measure of the network robustness. The Recall is particularly meaningful with unbalanced datasets.
- The Accuracy: $A = \frac{TP+TN}{TP+TN+FP+FN}$, reports correct predictions over the full testing set.
- The Jaccard Index: $J = \frac{TP}{TP+FP+FN}$, estimates the similarity between the ground truth and the prediction, computing the ratio between intersection and union of the two. If used in conjunction with the accuracy, accurately predicts the quality of the segmentation.

The joint analysis of these metrics provide a comprehensive insight of the networks' performance in terms of rejection to disturbances and management of false positives/negatives. A K-fold cross-validation with $K = 10$ is adopted to

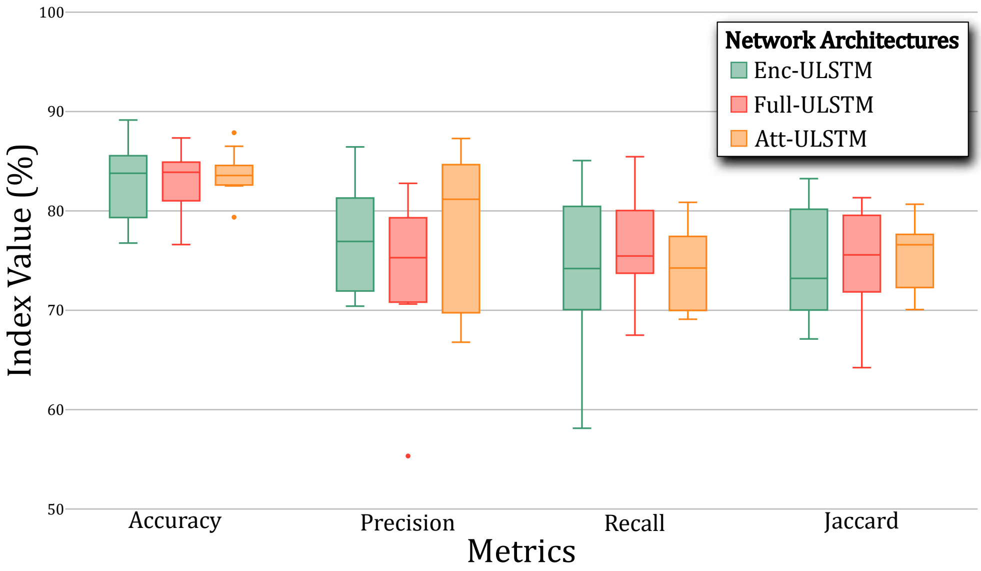


Figure 3.6: Performance comparison among the three proposed models. The metrics considered for this comparison are Accuracy, Precision, Recall and Jaccard Index. Results show that the best performance is achieved by the Att-ULSTM model in terms of accuracy, precision and Jaccard index while the Full-ULSTM show the best performance in terms of Recall.

validate the network's robustness against data variability. Initially, the models are trained and tested using the CL, discussed in Section 3.1.2.3. As in Figure 3.6, the Att-ULSTM model provides better performances in terms of accuracy, precision and Jaccard Index, while the best values of Recall is given by the Full-ULSTM network. The Att-ULSTM structure provides superior identification of the tissue flaps and a sufficient rejection to FP and FN as shown in Figure 3.5. Further analysis will be carried out only on Att-ULSTM model, comparing this architecture with the state of the art. To further improve the network performances, the Att-ULSTM is trained using the Tversky Loss instead of the Combo Loss. The results are reported in Table 3.2 and compared with the performance of the same network structure trained with the CL.

Table 3.2: Performance comparison of the model trained with both Tversky and Combo loss functions

	Tversky Loss	Combo Loss
Accuracy	82.25% ± 2.80%	83.77% ± 2.18%
Precision	74.89% ± 9.35%	78.42% ± 7.38%
Recall	70.60% ± 6.49%	74.32% ± 3.83%
Jaccard Index	72.53% ± 7.54%	75.83% ± 3.38%

The adoption of the Tversky Loss entails a slight loss of performance in the Att-ULSTM model with respect to the Combo Loss. For this reason, the combo Loss is selected. In Figure 3.7 and 3.1 the precision and accuracy during the training phase are reported for the worst $(K = 1)$, the average $(K = 2)$ and the best ($K = 3$) performing model over the K validations.

Given the restricted data available for this particular application, pre-training is evaluated, with the aim of limiting the over-fitting during training. The neural network model proposed in [196] is considered, due to its similarity with the Enc-ULSTM structure. Despite the similar structure, the pre-trained convolutional layers are characterised by an higher number of filters, thus increasing the model complexity. As shown in Table 3.3, the pre-trained model offers no performances improvement. This is motivated by the higher amount of kernels in the convolutional layers of the pre-trained model which increases the complexity of the model. Moreover, the model is pre-trained with microscopic images of cells, requiring a smaller amount of data augmentation with respect to endoscopic images, in which the geometrical constraints of the anatomy limit the image augmentation. Moreover, the amount of images contained in the pre-training dataset is limited, thus preventing the model to generalise the predictions.

Table 3.3: Performance comparison between the pre-trained model [196] and the model trained from scratch.

	P-ConvULSTM	Att-ULSTM
Accuracy	77.59% ± 2.30%	83.77% ± 2.18%
Precision	73.31% ± 5.64%	78.42% ± 7.38%
Recall	58.76% ± 5.59%	74.32% ± 3.83%
Jaccard Index	64.65% ± 4.83%	75.83% ± 3.38%

Finally, to demonstrate the increased performances provided by the approach in this work regarding the segmentation of single images, the Att-ULSTM model and the standard U-Net presented in [4] are compared. In Section 3.1.2.1, the U-Net implemented in our previous work is fed with single images from the video stream and produces a single prediction for each frame. By comparing these two networks it is possible to assert if the adoption of LSTM layers and attention gates is beneficial for tissue flap segmentation in video. The networks performances are evaluated in terms of accuracy and precision, as defined in Section 3.1.2.3.

Table 3.4: Performance comparison between the original FlapNet and the proposed Att-ULSTM

	Accuracy	Precision
U-Net [4]	80.90% ± 1.32%	72.63% ± 1.94%
Att-ULSTM	83.77% ± 2.18%	78.42% ± 7.38%
p-value	0.0173	0.0376

With the adoption of spatio-temporal layers and Attention Gates blocks in the Att-ULSTM, the model outperforms a standard feed-forward U-Net model,

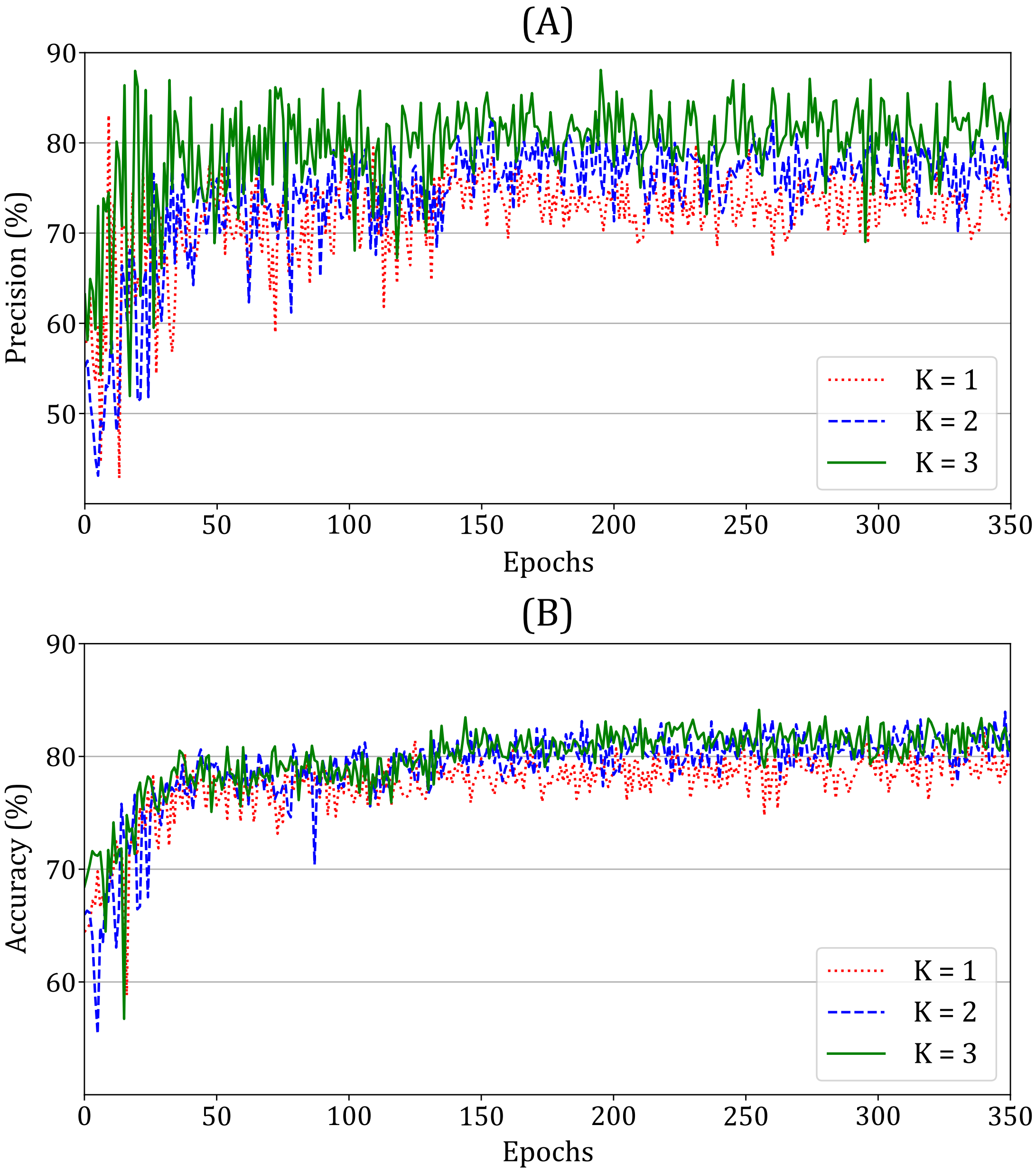


Figure 3.7: Precision (A) and accuracy (B) reported during training of the K=10 models for K-fold cross-validation. Only the best (K=3), the average (K=2) and the worst (K=1) are represent on the plot to simplify its visualisation. As the plateau is reached within the first 200 epochs, only 350 epochs are shown.

as shown in Table 3.4. In particular, the adoption of LSTM provides the ability to extract temporal information from subsequent frames, thus guaranteeing a more robust prediction. It is worth to mention that both the Att-ULSTM and U-Net models are trained over the same DMs, thus no evaluation bias is introduced in the comparison of the two models. The standard deviation of the precision is slightly higher for the Att-ULSTM. This is related to a better characterisation of the tools' presence in the augmented entries of the FlapNet, which are omitted in the training of the Att-ULSTM, as explained in Section 3.1.2.1. This enhances the robustness of the the U-Net with respect to the presence of tools, compared to the Att-ULSTM model. However, as shown by the other metrics, the segmentation of the Att-ULSTM is more reliable. Using the computer mentioned in Section 3.1.2.3, an inference time of $t_i = 0.5$ s was recorded, with a maximum speed of 2 FPS against the recorded $t_i < 42$ ms recorded for the standard feed-forward U-Net. This result is acceptable, considering that the surgeon motion are generally relatively slow to guarantee a safe interaction with the anatomy.

Given the limited training and testing data for the Att-ULSTM, a non-parametric test is required to prove the normal distribution of the two groups. A Wilcoxon rank sum test [216] is carried out for accuracy and precision to assess statistical significance of the two models' performances. This test assesses the null hypothesis that the two groups are continuous distribution with equal medians. In Table 3.4, the comparison between the models' accuracy and precision are shown. The p-value indicates a low probability for the two distribution to have equal median, thus there is a statistically significant improvement in the prediction performances using the Att-ULSTM model.

3.1.4 Conclusion

A novel approach to the segmentation of tissue in endoscopic video streams is herein discussed. Three neural network architectures for tissue segmentation in endoscopic images are proposed. The tissue detection and segmentation are considered the initial step towards intelligent interaction with the anatomy. On top of this, an estimation of the physical interaction is needed to accomplish a particular task. This evaluation however varies depending on the specific objective task to reproduce. The adoption of attention gates and recurrent structures such as LSTMs enhance the accuracy of the tissue detection, compared to a standard feed-forward network. The performances of the three variants are compared and the Att-ULSTM is selected for further investigation. For this network, different cost functions are compared, and the use of pre-training is evaluated. Experimental results show enhanced performances with respect to our previous work for what concerns the network's precision (78.42 % ± 7.38 %) and prediction stability. The adoption of LSTM and attention gates to take advantage of the time-related features, embedded in the images sequence, can improve the performances and robustness of the detection in the context of endoscopic images for surgical robotics. To achieve this result, the FlapNet dataset is enhanced to meet the requirements of the recurrent network's structure, thus resulting in a new dataset, now available to the research community.

The approach discussed in this work, demonstrating an enhanced ability to segment candidate soft tissues for retraction in the foreground of the scene, can significantly improve the implementation of autonomous tissue retraction base on the elaboration of endoscopic images only, meaning that no additional hardware is required in the da Vinci platform. Examples range from laparoscopic procedures, to non-autonomous robotic and semi-autonomous robot-assisted surgical tasks such as ablation, retraction and suturing. Localising the target tissue flap is indeed a key step towards surgical gesture automation and, given the variety and complexity of the human anatomy, this task is extremely challenging.

The major limitation of this work is the limited availability of labelled medical images. As pre-training over different dataset did not show promising results [4], weak labelling and unsupervised learning could be beneficial in dealing with such limited amount of data. As discussed above, the pre-training does not provide improved results, due to the unique characteristics of the surgical images. In conclusion, the most promising approach to increase the networks performances would consist of an increased number of entries in the dataset. However, labelling endoscopic images is time-consuming and requires specialised medical knowledge, thus hindering the process. The adoption of generative adversarial networks (GANs) could be beneficial to improve the network's ability to reject the surgical instrument, thus guaranteeing a correct and precise segmentation of the tissue flaps. Future work could include the adoption of endoscopic RGB image along with DMs to enhance the performance of the proposed model.

Chapter 4

Trajectory Planning

As anticipated in the Contribution Paragraph a significant limitation in using the dVRK platform is the impossibility to automatically register the camera held by the ECM to the tools equipped in the PSMs. This is due to fact that the SUJs information is not retrievable from the dVRK controllers and, being the SUJs the connection to the robot base, severs the kinematic chain of all the arms making impossible to use the inverse kinematic straight-forwardly. In order to recover the missing SUJs information an external frame is introduced to relate the camera to the tool in the scene. By means of a visual marker it is in fact possible to establish a transform between the camera and the SUJs thus allowing the planning and execution of trajectories in the camera frame. This is a crucial requirement for autonomous surgical task which are based on image processing of the endoscopic camera feed.

4.1 An Open Source Motion Planning Framework for Autonomous Minimally Invasive Surgical Robots

Authors: Aleks Attanasio, Nils Marahrens, Bruno Scaglioni and Pietro Valdastri

Abstract: *Planning and execution of autonomous tasks in minimally invasive sur-*

gical robotics are significantly more complex with respect to generic manipulators. Narrow abdominal cavities and limited entry points restrain the use of external vision systems and specialized kinematics prevent the straightforward use of standard planning algorithms. In this work, we present a novel implementation of a motion planning framework for minimally invasive surgical robots, composed of two subsystems: An arm-camera registration method only requiring the endoscopic camera and a graspable device, compatible with a 12mm trocar port, and a specialized trajectory planning algorithm, designed to generate smooth, non piecewise trajectories. The approach is tested on a DaVinci Research Kit obtaining an accuracy of 2.71 ± 0.89 *cm in the arm-camera registration and of* 1.30 ± 0.39 *cm during trajectory execution. The code is organised into STORM Motion Library (STOR-MoLib), an open source library, publicly available for the research community.*

4.1.1 Introduction

Trajectory planning lies at the heart of most robotic manipulation tasks and is crucial to enable high levels of autonomy [217]. While tasks usually define a set of different poses to be achieved, how the robot should move in between these poses is often left to motion planning algorithms. Common motion planners integrate a plethora of robot models, but surgical minimally invasive surgical systems are not well represented. This may attributed to their complex kinematic structures, often including parallel chains that are not supported by most inverse kinematics solvers and can be numerically challenging. Moreover, the software frameworks used to control surgical robots such as the Collaborative Robot Toolkit (CRTK) [218] and the DaVinci Research Kit (dVRK) [3] only provide the ability to reach a final pose with zero velocity, thus not supporting the execution of complex trajectories.

In the particular case of the dVRK, one of the most popular surgical robotics research platform [219], a point to point trajectory in the joint space is generated from the current end effector pose to the goal by means of the Reflexxes RML II [220] library. The resulting trajectory might be optimized in joint space but is generally neither smooth nor optimal in Cartesian space. The available literature on motion planning for surgical robots is scarce. In [221] the problem is addressed for the dVRK platform using the MoveIt![222] motion platform. However, the extended abstract is silent on how the problem of parallel kinematics, not supported in MoveIt!, is solved, nor is their code publicly available to the community. Recent works have focused on employing machine learning techniques, such as

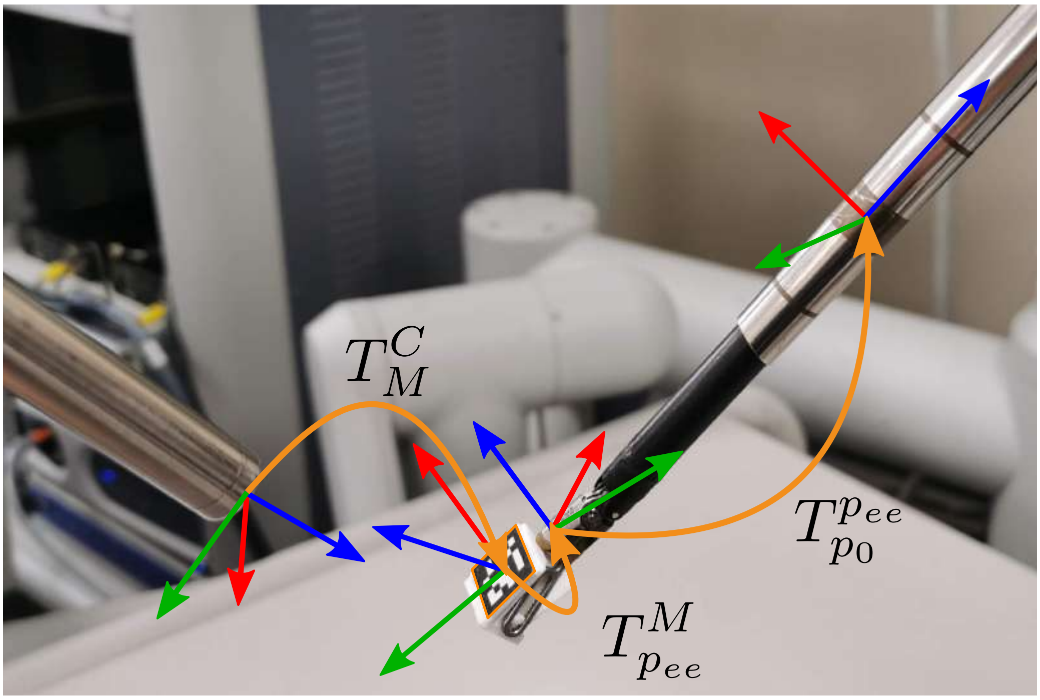


Figure 4.1: Transformations of the different frames considered for the registration of the arm to the camera frame.

Pyramid Stereo Matching Network (PSMNet) [223] and reinforcement learning [224]. While these methods show impressive results on specific tasks, they are not generally applicable and easily adaptable. Moreover, they are highly dependent on large amounts of labelled data, obtained via computationally and time-intensive simulations. Another common problem limiting the development of autonomous tasks in MIS robotics platforms is the co-registration between the camera and the robotic arms, since the two subsystems are usually connected to different bases. This issue is commonly solved for generic manipulators using external optical trackers [225]. This approach has been adopted for surgical robots [223, 226] by attaching markers on the tip of the surgical instruments. Although accurate, this method requires the use of an external camera, which is a major limitation in a small and delicate environment such as the abdominal cavity, and is prone to inaccuracies due to the presence blood or debris in the surgical scene. In this work, we: (1) Present a software framework aimed at solving the problem of co-registration for robotic platforms specific to MIS, focused on the ease of use and the potential transferability of the application to a clinical environment. (2) Present an approach to the planning and execution of complex trajectories on surgical robots, integrated with ROS and easily adaptable to any platform. (3) Provide public and documented code in a web repository to benefit the surgical robotics research community.

4.1.2 Co-registration algorithm

This section describes the approach adopted to determine the transformation between the endoscopic camera and the surgical instrument held by the robot. This step is crucial to plan and execute autonomous tasks based on visual servoing in scenarios where the endoscope and the robotic arm do not share the same reference frame. This is the case with robots such as the dVRK, the Raven [227] and modular robots like CMR Versyus or Medtronic's Hugo RAS. The goal is to compute the transformation from the camera frame to the origin of the robotic arm. This can be solved by evaluating a sequence of transformations that start from the pose of the robot end-effector with respect to the camera. In robots equipped with cameras, this can be achieved by adopting a computer vision algorithm to detect one or more visual markers mounted on the end-effector. To this end, we adopt the ArUco markers [228] and mount them on a custom 3D printed pick-up device, designed to be held by standard surgical instruments and be inserted through standard 12mm trocar ports. These markers have been adopted for their simple implementation, however, other forms of data matrix markers or passive lighting markers can be adopted to reject the disturbance of blood in the scene. Once the pick-up device with ArUco marker is grasped by the robotic instrument (Fenestrated Bipolar Forceps), exposed to the camera and recognized by the vision algorithm, the transformation T_C^{p0} between the PSM's base frame T_{p0} and the endoscope's base frame T_C is calculated as follows:

$$T_C^{p0} = T_C^M T_M^{p_{ee}} T_{p_{ee}}^{p_0} \tag{4.1}$$

where T_C^M is the transformation between camera and a visual marker held by the end-effector, $T_M^{p_{ee}}$ is the transformation between the marker and the end-effector reference frame, and finally $T_{p_{ee}}^{p_0}$ is the pose of the end-effector with respect to the robot base frame. The transformations are shown in Figure 4.1 on a DaVinci Patient Side Manipulator (PSM), in which the base frame is placed in the remote centre of motion, on the trocar. Assuming that $T_{p_{ee}}^{p_0}$ can be extracted from the robot kinematics and that $T_M^{p_{ee}}$ is known by design of the marker holder, T_C^M can be estimated by using the endoscope in conjunction with software packages like tuw_marker_detection [229] available on GitHub. Finally, the transformation $T_M^{p_{ee}}$ is applied to align the marker frame with the tool tip frame of the robot. To increase robustness of the results, we combine both detected transformations from the left and right endoscopic camera and average the results over 100 frames, each 100ms apart.

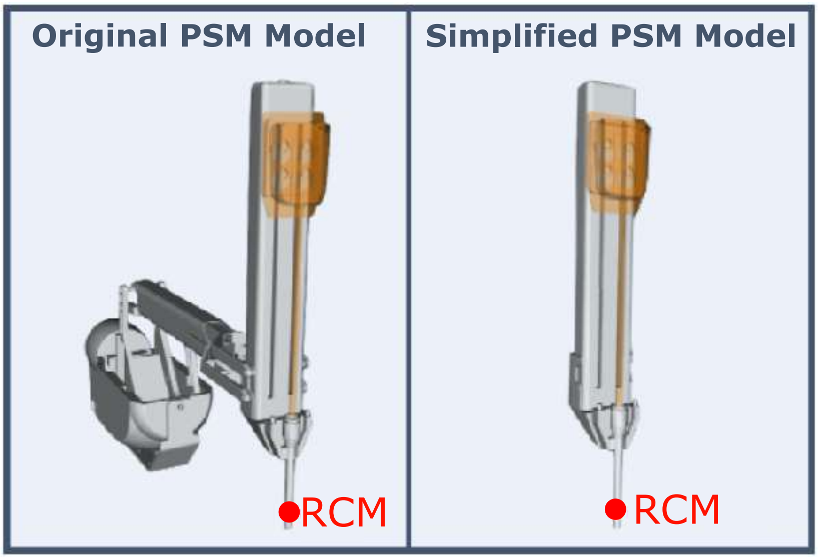


Figure 4.2: Original PSM model and the simplified model used in this work. In our simplified model, the base of the robot is omitted, thus removing the parallel kinematic chain and allowing the usage of the MoveIt! package without any loss of generality in the trajectory planning.

4.1.3 Trajectory planning

The co-registration algorithm enables to evaluate and control the position of the robot end-effector in the camera workspace. This feature facilitates the definition of points of interest based on computer vision or deep-learning algorithms and to relate them to the position of the end-effector. In many autonomous tasks, it is required to generate a trajectory based on the points identified in this step, and to execute it smoothly. One goal of this paper is to provide a framework for planning and smoothing of the trajectory dedicated of surgical robotic tools. For this purpose, the MoveIt! [230] framework has been used, due to the wide adoption in the research community. MoveIt! is based on the widely used Open Motion Planning Library (OMPL) [231] that includes state-of-the-art algorithms for trajectory planning, manipulation and navigation and is integrated into ROS [232]. In order to plan a trajectory for a specific robot, and therefore produce a feasible trajectory in joint and Cartesian spaces, MoveIt! gathers information about the robot layout from two files: the Unified Robot Description Format file (URDF), used in the ROS ecosystem to define robots kinematics, and the Semantic Robot Description Format file (SRDF), which includes additional information to the URDF such as default robot configuration and collision checking. The trajectory planning is carried out in four steps: (1) The robot URDF and SRDF are loaded onto Moveit!. (2) The robot starting position, way-points and goal of the trajectory are defined. (3) The MoveIt! function computeCartesianPath() is used to evaluate a sequence of points on straight lines from the starting position, through the way-points, to the final goal. (4) The Stochastic Trajectory Optimization for Motion Planning (STOMP) [233] is used to plan trajectory in the joint space using the previously generated points as seeds and produce the final trajectory, represented as a set of points in the 3D workspace. STOMP is adopted for its capability of avoiding local minima while allowing a faster convergence to the solution if compared to other planners such as Covariant Hamiltonian Optimization for Motion Planning (CHOMP) [234].

A C++ library, STORM Motion Library (STOR-MoLib) is developed to provide the code to the community. The library requires minimal user input and can be utilized by means of the following methods: *compileMotionPlanRequest(waypoints_constraint, trajectory_seed)* and *transformTrajectory(trajectory, base_frame)*. The first populates the MoveIt! motion request constraining the passage through the desired way-points. The trajectory seeds are the output of the computeCartesianPath function included in MoveIt!. The second function transforms the trajectory points from the robot frame to the user-defined base frame, in our case the camera frame. The MoveIt! motion request is then solved by the STOMP Planner which returns a smoothed trajectory. In summary, these functions allow a straight forward implementation of trajectory planning based on the previous co-registration of the ECM and PSMs exploiting the capabilities of the MoveIt! framework.

4.1.4 Experimental validation

The validation of our approach is composed of two steps: the evaluation of the accuracy for the camera-arm registration and the assessment of the trajectories planning and execution. Although the application of the framework could be generalized to any robot, in this work we focus on the dVRK due to its ubiquity and the availability of an open source simulation software, thus circumventing the need for a physical platform, to replicate the results described here. In particular, we adopt a subset of the full DaVinci system composed of one PSM and one stereoscopic endoscope mounted on an independent base. A Linux (Ubuntu 18.04) machine equipped with an Intel Xeon Gold 6140 (2.30GHz) CPU, an Nvidia Quadro 5000 RTX GPU and 128 GB DDR4 2666MHz RAM was adopted to carry out the planning. While the use of a specific robot is transparent to the co-registration algorithm, the trajectory planning depends on the features of each robotic arm through the URDF and SRDF files. Initially, the PSM description files provided with the dVRK library [3] are used. However, the PSM adopts a parallel mechanism to ensure a fixed remote centre of motion.

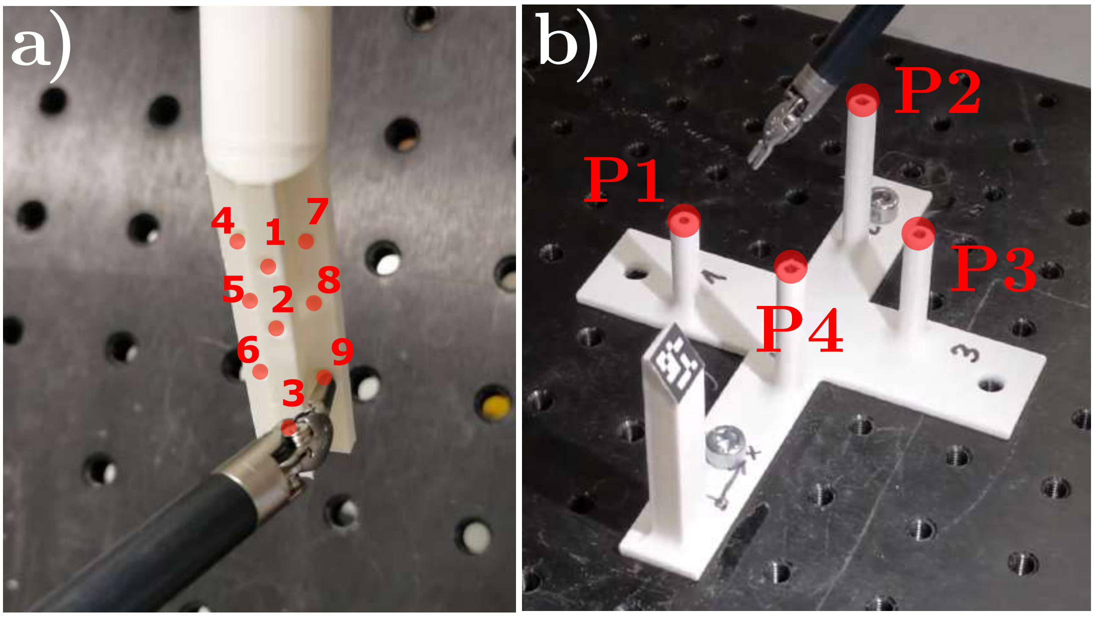


Figure 4.3: 3D-printed rigid body used for the validation of the marker-based co-registration (a). 3D-printed rigid body used to validate the precision during the trajectory execution (b). A marker has been attached to the body to allow the registration of the points via the camera.

This type of kinematics is not supported in MoveIt!. In order to overcome this issue, a modified version of the PSM excluding the parallel link is developed (Figure 4.2). Despite the different physical layout, the kinematics of the robot is correctly reproduced by maintaining the Remote Centre of Mass fixed and eliminating the parallel link and the preceding links in the kinematic chain.

To quantify the registration error, a 3D-printed calibration body attachable to the endoscope's tip was designed. The calibration body contains nine landmark points (p_C^1 - p_C^9) with known distance with respect to the camera's base frame T^C (Figure 4.3a). By touching the landmarks with the tip of the surgical instrument, we acquired the location of these positions in the PSM's base frame T_{p0}.By performing several registrations ($n = 5$) and averaging the position of each of the nine points over all runs we obtain p_{p0}^1 - p_{p0}^9. With a confidence interval of 0.0734 mm ($c = 0.95$), we assume the robot's positional accuracy to be fairly high and consistent compared to the camera. In order to assess the accuracy of the co-registration approach on our surgical setup, five registrations are performed using the ArUco marker with differing tool positions and thus different placements of the marker with respect to the camera. With the acquired transformations T_C^{po} from the visual marker registrations, we transform the points p_C^1 - p_C^9 on the calibration body from the camera's base frame T_C to the PSM's base frame T_{p0} and calculate the euclidean distance to the respective points obtained via landmark registration. Our results indicate a mean positional error of 2.71 ± 0.89 cm ($c = 0.95$) over all registered points and registration runs compared to the position obtained via the camera calibration body. We believe the main source of inaccuracy to be the camera distortion. Despite a thorough calibration, the fish-eye lenses of the endoscope produce a significant distortion that negatively affects the accuracy of the marker detection, particular when the marker is not place directly at the center of the image. Additionally, the small distance between the two cameras limits the usage of further information from the 3D scene via stereo matching or similar techniques.

In order to evaluate the accuracy of the trajectory planning and execution, a 3D-printed reference body with four vertical pegs was designed. The tip of each peg represents either a way-point or the goal of the trajectory (Figure 4.3b). The reference body also integrates an ArUco marker, added to obtain a transformation from its local reference frame to the camera frame T_C^{RB}. The coordinates of each way-point are transformed into the PSM's base frame T_{p0} by combining the two previously obtained transformations ($T_{p0}^{RB} = T_{p0}^{C} T_C^{RB}$). The planner evaluates a trajectory starting from the current position of the instrument, passing along the way-points and ending in the goal position. Two different trajectory scenarios have been considered with three and four way-points, respectively. Each trajectory has been repeated 8 times and, for each repetition, the surgical instrument was initially manually placed in a varying position around the starting point. Although the planner can consider variable instrument orientations, we maintained a constant, randomly selected, orientation during the whole trajectory.

The planner's output consists of a trajectory defined as an array of joint values, one set for every trajectory point. These are converted to the Cartesian space by means of forward kinematics and eventually organised in a vector of poses sent to the dVRK software. The dVRK only allows a point to point trajectory, constraining the initial and goal velocity to zero. To perform a smooth trajectory, we published the new poses at a rate of 20Hz, sending a new command before the robot had reached the previous goal and thus avoiding the condition of zero velocity. Before executing each trajectory, the position of each way-point with respect to the robot's base frame T_{p0} was collected by manually positioning the surgical instrument (large needle driver) onto a landmark on each peg's tip and recording its position. Figure 4.4 shows the 8 trajectories for both the three and four point case. The start and end point of the trajectory are represented in blue and green, respectively. The way-points are represented in red. It must be pointed out that the sequence of the way-points is different for the two trajectories. The sequence chosen in the four point case is aimed at demonstrating the ability of the planner to find a solution in the even in the case of more involved trajectories, containing a indirect path with back and forth motion.

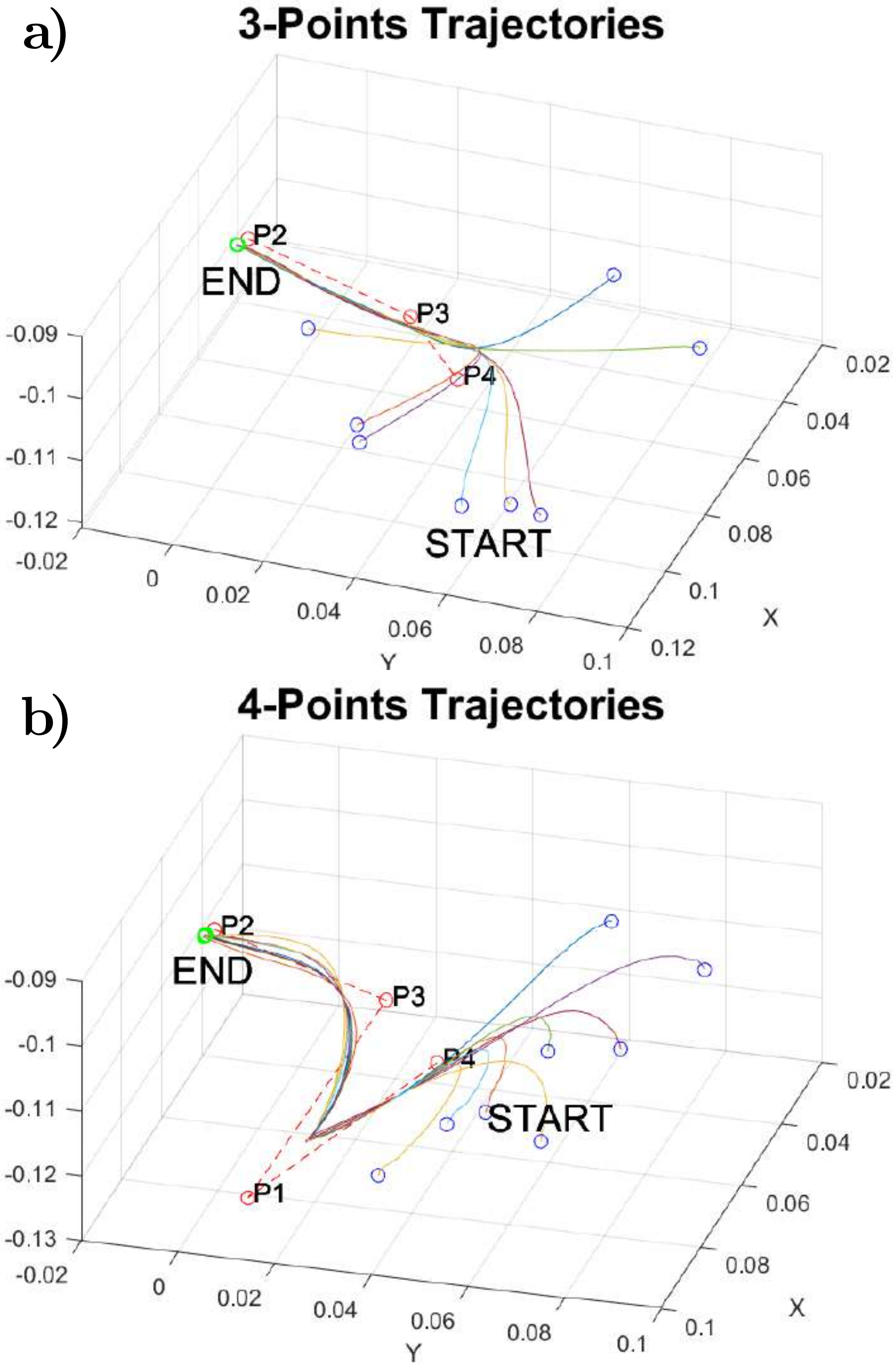


Figure 4.4: Repetitions for the trajectory planning and execution for three point (a) and four point (b) case. The initial point is shown in blue, the goal point in green and the way-points in red. The red dashed lines depict the seeds used by the STOMP planner.

The evaluation of the trajectories is carried out by considering the minimum distance between the path executed by the robot and each way-point measured before the trajectory execution via the robots tool tip. With this reference, the average error amounts to 1.09 ± 0.59 cm ($c = 0.95$) for the three point and

1.30 ± 0.39 cm ($c = 0.95$) in the four point case.

4.1.5 Conclusions

In this paper, we presented a comprehensive library to manage the trajectory planning of surgical robots with the specific aim of developing a method that does not require dedicated hardware such as optical trackers or external cameras, thus applicable in the context of minimally invasive surgery. Initially, we presented a method for arm-to-camera registration based on the ArUco markers. We showed the method to be a feasible approach in robotic systems where the arms and the camera do not share the same kinematic base. Subsequently, we demonstrated an approach for planning and executing trajectories based on Moveit! and integrated with ROS. For our evaluation, we applied our framework and approach to the dVRK platform. The registration makes it possible to plan trajectories with respect to the camera frame, thus supporting the execution of vision-based autonomous surgical gestures. Moreover, the registration algorithm can be useful in setups, such as the dVRK, in which teleoperation is challenging due to the lack of a simple built-in co-registration protocol. Although the dVRK Setup Joints controller will be available in the future, not all the research groups have access to the full platform. We believe that this library could significantly benefit the research community. STOR-MoLib code is open source and publicly available [1].

Further development of this library, currently under investigation, include the implementation of a collision avoidance algorithm, useful in collaboration scenarios in which a human operator is controlling one arm, while the other arm is autonomously operated. Other improvements, particularly regarding the registration accuracy, might be obtained by further investigations on the distortion of the cameras' lenses which majorly contributes to the registration error.

[1] https://github.com/Stormlabuk/dvrk_stormolib

Chapter 5

Conclusions

The results presented in this work define a new paradigm of autonomy in surgical robotics and validate the feasibility of an approach for autonomous tissue retraction. The formulation of the hypothesis of an autonomous surgical gesture is driven by a thorough survey of the state-of-the-art. Starting from an established definition of autonomy for medical robotics [1], our analysis consisted in extending this paradigm and deepening its aspects for what concerns surgical robotics. Focusing on this technology branch revealed both strong and weak points of the research reported to date, further helping to define the scope of this thesis. Additionally, gathering knowledge from different case studies allowed implementing the most trending approaches and moving towards their adoption for the work conducted in this thesis.

Inspired by these trending technologies, this thesis is founded on data-based approaches to guarantee adaptability and transferability to different application domains. In order to train different machine learning models for tissue segmentation, data were collected under the guidance of experienced surgeons. With their support it was possible to both gather data from a realistic scenario and properly categorise and label the activity carried out during collection. The images gathered during this initial phase and collected in the FlapNet dataset were afterwards used to train neural network models in order to detect tissue flaps candidates for retraction in the surgical scene. The FlapNet dataset, which is publicly available at STORM Lab GitHub repositories for the research community, constitutes the foundation and first contribution of this thesis. The dataset, collected and labelled alongside with experienced surgeon, is fundamental to tackle the main technical issues related to the perception, planning and execution described in the introduction of this work. In order to understand and detect the surgical workspace it is fundamental to reconstruct its 3D features.

To this end, the dVRK stereo cameras proved to be an effective instrument to extract depth maps. Although the high distortion of the endoscopic lenses prevents an accurate 3D reconstruction of the workspace, the fusion of cameras' feeds and neural network models trained with FlapNet proved to be an efficient way to define 3D features in the workspace allowing a subsequent interaction with the anatomy.

The 3D features detected by the model trained on our collected dataset are fundamental to plan a safe and precise interaction with anatomical structures. Since the dVRK kinematic chain results severed without the Set Up Joint controllers, planning motion and interaction within the camera space proves to be impossible. This thesis presented an image-based method using visual markers to retrieve the missing joint information and restore the otherwise severed kinematic chain, thus allowing the planning and execution of controlled motions within the camera space. Moreover, as the dVRK does not provide any motion planner for complex trajectories, an additional contribution of this work consisted in the development of a user-friendly C++ library to plan and execute a trajectory with the PSM arms in an arbitrarily chosen reference frame, in our case the camera one. The library is based on the well-known MoveIt! environment and presents functions with the purpose of minimising the user input while planning a trajectory on the dVRK, simplifying the task. The code, as well as examples on how to use it, is publicly available on the STORM Lab repository knowing that this will support research groups in defining future approaches and methods to tackle autonomous tasks.

The final contribution of in this thesis was the definition of the tissue retraction gesture, achieved by merging both the previous elements with the knowledge gathered from experienced clinicians. In fact, by interviewing them about the details on when and how a tissue flap should be retracted and moved away from the scene, it was possible to characterise and replicate a pattern in their motions, defining a gesture model that can be applied to automate the task. This knowledge, along with a perception model capable of extracting tissue flaps border from the camera view and a motion planner able to generate smooth and controlled robotic arm trajectories, allowed the replication of the surgeon's or an assistant's activity during a procedure. The effectiveness and robustness of this approach has been tested and validated in a controlled environment proving the feasibility of this approach.

Although the presented method proved to be effective under specific circumstances, major limitations have been recorded. The first issue faced during development was given by the data-driven aspect of this work. In fact, however advantageous for dynamic environments such as the human anatomy, machine learning models such as neural networks demand a high quantity of data to be properly trained. In the surgical domain these data are either hard to collect or difficult to label, requiring the expertise of competent clinicians. This constitutes a noticeable obstacle in training the models and requires advanced techniques and powerful computers to compensate for. Future work may include the enrichment of the FlapNet dataset as well as other available datasets for training in the surgical domain. This can only be achieved with the collaboration between engineers and surgeons moving towards a more data-driven technology environment. An alternative approach could also consist in the generation of image through GAN or using 3D simulated environment in Unity. This would significantly help the data gathering process, removing the labelling procedure bottleneck. Additionally, the generation of dataset base on specific anatomical structure could guarantee extended generality of the herein proposed method. Additionally to a more extended dataset, another approach worth of testing regards the flap detection part. Recently, vision transformers have been used in computer vision. These models have been increasingly adopted in the last 2 to 3 years and showed a remarkable capability in generalizing, thus allowing a more convenient and accurate classification. The adoption of models such has XCiT could provide enhanced performance during the flap detection process, thus allowing a more precise gesture. Another impediment reported during the development of this thesis was the remarkably high distortion of the dVRK stereo endoscopic lenses. These were designed to guarantee a wide field of view to the surgeon who can easily spot salient aspects of the anatomy during an intervention. From a robotic perspective, this distortion is hard to map on a camera and even a small error in the estimation of the model causes a noticeable error during 3D scene reconstruction. This generates errors in the order of magnitude (1 cm to 2 cm) which can be easily ignored in everyday applications, but that are not neglectable in a surgical environment. Usually problems related to the camera distortion is tackled using other cameras from the one provided with the dVRK. However, further studies must include possible solution to either reduce the camera distortion or to increase the accuracy of the camera model estimation. This can be done either with conventional computer vision algorithm or with more recent data based approaches to extract the intrinsic camera matrix. An alternative approach, could be to use different cameras which would present a reduced distortion guaranteeing a more robust 3D reconstruction. This has not been done in this work since one of the objective of this thesis was to demonstrate the feasibility of autonomous task based on visual feedback without additional equipment other than the one provided with the Da Vinci platform. Concluding, an additional feature necessary for a compliant implementation of autonomous tissue retraction in a real OR consists in the obstacle avoidance for the PSM motion planner. Since, ideally, this autonomous system is supposed to work along with a human operator, being able to prevent collisions with the human driven tools becomes mandatory to avoid continuous workflow disruptions. This can be achieved expanding the work of the already available StorMoLib library with the obstacle avoidance approaches already implemented in the MoveIt! framework. This part of work has not been addressed in the thesis due to lack of time, however this can significantly enhance the performance of autonomous tissue retraction.

References

[1] G.-Z. Yang, J. Cambias, K. Cleary, E. Daimler, J. Drake, P. E. Dupont, N. Hata, P. Kazanzides, S. Martel, R. V. Patel, V. J. Santos, and R. H. Taylor, "Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2, no. 4, p. eaam8638, 2017. 1, 4, 11, 12, 64, 90

[2] A. Mitsala, C. Tsalikidis, M. Pitiakoudis, C. Simopoulos, and A. K. Tsaroucha, "Artificial intelligence in colorectal cancer screening, diagnosis and treatment. a new era," *Current Oncology*, vol. 28, no. 3, pp. 1581–1607, 2021. 2

[3] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci® Surgical System," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6434–6439, IEEE, may 2014. 2, 43, 81, 85

[4] A. Attanasio, B. Scaglioni, M. Leonetti, A. Frangi, W. Cross, C. S. Biyani, and P. Valdastri, "Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery - A Feasibility Study," *IEEE Robotics and Automation Letters*, 2020. 3, 63, 64, 67, 70, 76, 79

[5] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, and P. Valdastri, "Autonomy in surgical robotics," *Annual Review of Control Robotics and Autonomous Systems*, 2020. 3, 64

[6] A. Attanasio, C. Alberti, B. Scaglioni, N. Marahrens, A. F. Frangi, M. Leonetti, C. S. Biyani, E. De Momi, and P. Valdastri, "A comparative study of spatio-temporal u-nets for tissue segmentation in surgical robotics," *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 1, pp. 53–63, 2021. 3

[7] A. Attanasio, N. Marahrens, B. Scaglioni, and P. Valdastri, "An open source motion planning framework for autonomous minimally invasive sur-

gical robots," *IEEE International Conference on Autonomous Systems*, 2021. 3

[8] "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," standard, SAE International, Warrendale, Pensylvenia, US, June 2018. 11, 35

[9] N. Simaan, R. M. Yasin, and L. Wang, "Medical Technologies and Challenges of Robot-Assisted Minimally Invasive Intervention and Diagnostics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 465–490, may 2018. 13

[10] M. Hoeckelmann, I. J. Rudas, P. Fiorini, F. Kirchner, and T. Haidegger, "Current capabilities and development potential in surgical robotics," *International Journal of Advanced Robotic Systems*, vol. 12, 2015. 14

[11] B. J. Nelson, I. K. Kaliakatsos, and J. J. Abbott, "Microrobots for Minimally Invasive Medicine," *Annual Review of Biomedical Engineering*, vol. 12, no. 1, pp. 55–85, 2010. 14

[12] C. Bergeles and G. Z. Yang, "From passive tool holders to microsurgeons: Safer, smaller, smarter surgical robots," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1565–1576, 2014. 14

[13] V. Vitiello, Su-Lin Lee, T. P. Cundy, and Guang-Zhong Yang, "Emerging Robotic Platforms for Minimally Invasive Surgery," *IEEE Reviews in Biomedical Engineering*, vol. 6, no. 1, pp. 111–126, 2013. 14

[14] B. Dahroug, B. Tamadazte, L. Tavernier, S. Weber, and N. Andreff, "Review on Otological Robotic Systems: Toward Micro-Robot Assisted Cholesteatoma Surgery," *IEEE Reviews in Biomedical Engineering*, vol. XXX, no. XXX, pp. 1–19, 2018. 14

[15] J. A. Smith, J. Jivraj, R. Wong, and V. Yang, "30 Years of Neurosurgical Robots: Review and Trends for Manipulators and Associated Navigational Systems," *Annals of Biomedical Engineering*, vol. 44, pp. 836–846, apr 2016. 14

[16] C. Faria, W. Erlhagen, M. Rito, E. De Momi, G. Ferrigno, and E. Bicho, "Review of Robotic Technology for Stereotactic Neurosurgery," *IEEE Reviews in Biomedical Engineering*, vol. 8, pp. 125–137, 2015. 14

[17] F. Pugin, P. Bucher, and P. Morel, "History of robotic surgery : From AESOP® and ZEUS® to da Vinci®," *Journal of Visceral Surgery*, vol. 148, no. 5, pp. e3–e8, 2011. 14

[18] B. P. M. Yeung and T. Gourlay, "A technical review of flexible endoscopic multitasking platforms," *International Journal of Surgery*, vol. 10, no. 7, pp. 345–354, 2012. 14

[19] MMI S.P.A., "MMI - Medical Micro Instruments." `http://www.mmimicro.com/`, 2020. Accessed: 04-05-2020. 14

[20] Auris Health inc., "The Monarch platform." `https://www.aurishealth.com/`, 2020. Accessed: 04-05-2020. 14

[21] V. Groenhuis, F. J. Siepel, J. Veltman, J. K. van Zandwijk, and S. Stramigioli, "Stormram 4: An MR Safe Robotic System for Breast Biopsy," *Annals of Biomedical Engineering*, vol. 46, pp. 1686–1696, oct 2018. 14

[22] R. H. Taylor, A. Menciassi, G. Fichtinger, P. Fiorini, and P. Dario, "Medical Robotics and Computer-Integrated Surgery," in *Springer Handbook of Robotics*, pp. 1657–1684, Cham: Springer International Publishing, 2016. 16

[23] S. J. Mckenna, H. N. Charif, and T. Frank, "Towards Video Understanding of Laparoscopic Surgery : Instrument Tracking," in *Proc. of Image and Vision Computing*, no. November, pp. 2–6, 2005. 16

[24] Intuitive Surgical, "DaVinci Surgical System." `www.intuitivesurgical.com`, 2020. Accessed: 04-05-2020. 16

[25] B. Hannaford, J. Rosen, D. W. Friedman, H. King, P. Roan, Lei Cheng, D. Glozman, Ji Ma, S. N. Kosari, and L. White, "Raven-II: An Open Platform for Surgical Robotics Research," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 954–959, apr 2013. 16

[26] S. Voros, J. A. Long, and P. Cinquin, "Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders," *International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1173–1190, 2007. 16

[27] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros, "3D Tracking of Laparoscopic Instruments Using Statistical and Geometric Modeling," *Medical*

Image Computing and Computer-Assisted Intervention, vol. 6891, pp. 203–210, 2011. 16

[28] L. C. Garcia-Peraza-Herrera, W. Li, L. Fidon, C. Gruijthuijsen, A. Devreker, G. Attilakos, J. Deprest, E. V. Poorten, D. Stoyanov, T. Vercauteren, and S. Ourselin, "ToolNet: Holistically-nested real-time segmentation of robotic surgical tools," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5717–5722, IEEE, sep 2017. 16

[29] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, and D. Stoyanov, "Toward detection and localization of instruments in minimally invasive surgery," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1050–1058, 2013. 16

[30] A. Reiter, P. K. Allen, and T. Zhao, "Feature Classification for Tracking Articulated Surgical Tools," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, pp. 592–600, 2012. 16

[31] E. Colleoni, S. Moccia, X. Du, E. De Momi, and D. Stoyanov, "Deep Learning Based Robotic Tool Detection and Articulation Estimation With Spatio-Temporal Layers," *IEEE Robotics and Automation Letters*, vol. 4, pp. 2714–2721, Jul 2019. 16, 44, 65

[32] D. Bouget, R. Benenson, M. Omran, L. Riffaud, B. Schiele, and P. Jannin, "Detecting Surgical Tools by Modelling Local Appearance and Global Shape," *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2603–2617, 2015. 16

[33] M. S. Atkins, G. Tien, R. S. Khan, A. Meneghetti, and B. Zheng, "What do surgeons see: Capturing and synchronizing eye gaze for surgery applications," *Surgical Innovation*, vol. 20, no. 3, pp. 241–248, 2013. 17

[34] D. García-Mato, A. Lasso, A. Szulewski, J. Pascau, and G. Fichtinger, "3D Gaze Tracking based on Eye and Head Pose Tracking," *10th Hamlyn Symposium on Medical Robotics*, pp. 87–88, 2017. 17

[35] "Head-Mounted Display Use in Surgery: A Systematic Review," *Surgical Innovation*, vol. 27, pp. 88–100, feb 2020. 17

[36] I. Tong, O. Mohareri, S. Tatasurya, C. Hennessey, and S. Salcudean, "A retrofit eye gaze tracker for the da Vinci and its integration in task execu-

tion using the da Vinci Research Kit," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Decem, pp. 2043–2050, 2015. 17

[37] H. M. Yip, D. Navarro-Alarcon, and Y. H. Liu, "Development of an eye-gaze controlled interface for surgical manipulators using eye-tracking glasses," *2016 IEEE International Conference on Robotics and Biomimetics, ROBIO 2016*, pp. 1900–1905, 2016. 17

[38] Tobii Pro, "Eye tracking for research." `www.tobiipro.com`, 2020. Accessed: 04-05-2020. 17

[39] J. Konstantinova, A. Jiang, K. Althoefer, P. Dasgupta, and T. Nanayakkara, "Implementation of Tactile Sensing for Palpation in Robot-Assisted Minimally Invasive Surgery: A Review," *IEEE Sensors Journal*, vol. 14, pp. 2490–2501, aug 2014. 17

[40] D. G. Black, A. H. H. Hosseinabadi, and S. E. Salcudean, "6-DOF Force Sensing for the Master Tool Manipulator of the da Vinci Surgical System," *IEEE Robotics and Automation Letters*, vol. 5, pp. 2264–2271, apr 2020. 17

[41] F. Piqué, M. N. Boushaki, M. Brancadoro, E. De Momi, and A. Menciassi, "Dynamic Modeling of the da Vinci Research Kit Arm for the Estimation of Interaction Wrench," *2019 International Symposium on Medical Robotics, ISMR 2019*, 2019. 17

[42] A. Marban, V. Srinivasan, W. Samek, J. Fernández, and A. Casals, "A recurrent convolutional neural network approach for sensorless force estimation in robotic surgery," *Biomedical Signal Processing and Control*, vol. 50, pp. 134–150, 2019. 17

[43] Y. Wang, R. Gondokaryono, A. Munawar, and G. S. Fischer, "A Convex Optimization-Based Dynamic Model Identification Package for the da Vinci Research Kit," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3657–3664, 2019. 17

[44] H. Sang, J. Yun, R. Monfaredi, E. Wilson, H. Fooladi, and K. Cleary, "External force estimation and implementation in robotically assisted minimally invasive surgery," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 13, no. 2, pp. 1–15, 2017. 17

[45] A. Krupa, J. Gangloff, C. Doignon, M. F. De Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux, "Autonomous 3-D Positioning of Surgical Instruments in Robotized Laparoscopic Surgery Using Visual Servoing," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 5, pp. 842–853, 2003. 19

[46] C. Suárez, B. Acha, C. Serrano, C. Parra, and T. Gómez, "VirSSPA- A virtual reality tool for surgical planning workflow," *International Journal of Computer Assisted Radiology and Surgery*, vol. 4, no. 2, pp. 133–139, 2009. 19

[47] N. Das and M. C. Yip, "Forward Kinematics Kernel for Improved Proxy Collision Checking," *IEEE Robotics and Automation Letters*, vol. 5, pp. 2349–2356, apr 2020. 19

[48] G. Sys, H. Eykens, G. Lenaerts, F. Shumelinsky, C. Robbrecht, and B. Poffyn, "Accuracy assessment of surgical planning and three-dimensional-printed patient-specific guides for orthopaedic osteotomies," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 231, no. 6, pp. 499–508, 2017. 19

[49] S. L. Lee, M. Lerotic, V. Vitiello, S. Giannarou, K. W. Kwok, M. Visentini-Scarzanella, and G. Z. Yang, "From medical images to minimally invasive intervention: Computer assistance for robotic surgery," *Computerized Medical Imaging and Graphics*, vol. 34, no. 1, pp. 33–45, 2010. 19

[50] D. W. Roberts, J. W. Strohbehn, J. F. Hatch, W. Murray, and H. Kettenberger, "A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope," *Journal of Neurosurgery*, vol. 65, no. 4, pp. 545–549, 1986. 19

[51] Y. Mochizuki, A. Hosaka, H. Kamiuchi, J. X. Nie, K. Masamune, K. Hoshina, T. Miyata, and T. Watanabe, "New simple image overlay system using a tablet PC for pinpoint identification of the appropriate site for anastomosis in peripheral arterial reconstruction," *Surgery Today*, vol. 46, no. 12, pp. 1387–1393, 2016. 19

[52] S.-H. Kong, N. Haouchine, R. Soares, A. Klymchenko, B. Andreiuk, B. Marques, G. Shabat, T. Piechaud, M. Diana, S. Cotin, and J. Marescaux, "Robust augmented reality registration method for localization of solid organs' tumors using CT-derived virtual biomechanical model

and fluorescent fiducials," *Surgical Endoscopy*, vol. 31, pp. 2863–2871, jul 2017. 19

[53] G. Samei, K. Tsang, C. Kesch, J. Lobo, S. Hor, O. Mohareri, S. Chang, S. L. Goldenberg, P. C. Black, and S. Salcudean, "A partial augmented reality system with live ultrasound and registered preoperative MRI for guiding robot-assisted radical prostatectomy," *Medical Image Analysis*, vol. 60, p. 101588, Feb 2020. 19

[54] D. Katić, A. L. Wekerle, J. Görtler, P. Spengler, S. Bodenstedt, S. Röhl, S. Suwelack, H. G. Kenngott, M. Wagner, B. P. Müller-Stich, R. Dillmann, and S. Speidel, "Context-aware Augmented Reality in laparoscopic surgery," *Computerized Medical Imaging and Graphics*, vol. 37, no. 2, pp. 174–182, 2013. 19

[55] L. Qian, A. Deguet, and P. Kazanzides, "ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery," *Healthcare Technology Letters*, vol. 5, pp. 194–200, oct 2018. 19

[56] O. Sgarbura and C. Vasilescu, "The decisive role of the patient-side surgeon in robotic surgery," *Surgical Endoscopy*, vol. 24, pp. 3149–3155, dec 2010. 19, 23

[57] M. A. Williams, J. McVeigh, A. I. Handa, and R. Lee, "Augmented reality in surgical training: a systematic review," *Postgraduate Medical Journal*, pp. postgradmedj–2020–137600, mar 2020. 20

[58] Osso Virtual Reality, "The OSSO virtual training system." https://ossovr.com/, 2020. Accessed: 04-05-2020. 20

[59] S. Wang, M. Parsons, J. Stone-McLean, P. Rogers, S. Boyd, K. Hoover, O. Meruvia-Pastor, M. Gong, and A. Smith, "Augmented Reality as a Telemedicine Platform for Remote Procedural Training," *Sensors*, vol. 17, p. 2294, oct 2017. 20

[60] M. Bowthorpe, M. Tavakoli, H. Becher, and R. Howe, "Smith predictor based control in teleoperated image-guided beating-heart surgery," in *2013 IEEE International Conference on Robotics and Automation*, pp. 5825–5830, may 2013. 20

[61] N. A. Wood, D. Schwartzman, M. J. Passineau, R. J. Moraca, M. A. Zenati, and C. N. Riviere, "Beating-heart registration for organ-mounted robots,"

International Journal of Medical Robotics and Computer Assisted Surgery, vol. 14, no. 4, pp. 1–9, 2018. 20

[62] A. Ruszkowski, O. Mohareri, S. Lichtenstein, R. Cook, and S. Salcudean, "On the feasibility of heart motion compensation on the daVinci® surgical robot for coronary artery bypass surgery: Implementation and user studies," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015-June, no. June, pp. 4432–4439, 2015. 20

[63] S. A. Bowyer, B. L. Davies, and F. Rodriguez y Baena, "Active Constraints/Virtual Fixtures: A Survey," *IEEE Transactions on Robotics*, vol. 30, pp. 138–157, feb 2014. 20

[64] A. Al Nooryani and W. Aboushokka, "Rotate-on-Retract Procedural Automation for Robotic-Assisted Percutaneous Coronary Intervention: First Clinical Experience," *Case Reports in Cardiology*, vol. 2018, pp. 1–3, dec 2018. 21

[65] H. Naghibi, W. B. Hoitzing, S. Stramigioli, and M. Abayazid, "A Flexible Endoscopic Sensing Module for Force Haptic Feedback Integration," in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, pp. 158–161, IEEE, dec 2018. 21

[66] S. Hodgson, M. Tavakoli, A. Lelevé, and M. Tu Pham, "High-fidelity sliding mode control of a pneumatic haptic teleoperation system," *Advanced Robotics*, vol. 28, no. 10, pp. 659–671, 2014. 21

[67] K. Ogawa, K. Ohnishi, and Y. Ibrahim, "Development of Flexible Haptic Forceps Based on the Electro-Hydraulic Transmission System," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 1–1, 2018. 21

[68] M. B. Molinero, G. Dagnino, J. Liu, W. Chi, M. E. M. K. Abdelaziz, T. Kwok, C. Riga, and G. Yang, "Haptic Guidance for Robot-Assisted Endovascular Procedures: Implementation and Evaluation on Surgical Simulator," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5398–5403, IEEE, nov 2019. 21

[69] R. Moccia, M. Selvaggio, L. Villani, B. Siciliano, and F. Ficuciello, "Vision-based Virtual Fixtures Generation for Robotic-Assisted Polyp Dissection Procedures," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7934–7939, IEEE, nov 2019. 21

[70] 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," *Surgical Endoscopy*, vol. 23, pp. 1180–1190, jun 2009. 21

[71] A. Spinelli, G. David, S. Gidaro, M. Carvello, M. Sacchi, M. Montorsi, and I. Montroni, "First experience in colorectal surgery with a new robotic platform with haptic feedback," *Colorectal Disease*, vol. 20, pp. 228–235, mar 2018. 21

[72] P. R. Slawinski, A. Z. Taddese, K. B. Musto, K. L. Obstein, and P. Valdastri, "Autonomous Retroflexion of a Magnetic Flexible Endoscope," *IEEE Robotics and Automation Letters*, vol. 2, pp. 1352–1359, jul 2017. 22

[73] B. B. Haro, L. Zappella, and R. Vidal, "Surgical gesture classification from video data.," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 15, no. Pt 1, pp. 34–41, 2012. 23

[74] B. Van Amsterdam, H. Nakawala, E. D. Momi, and D. Stoyanov, "Weakly Supervised Recognition of Surgical Gestures," in *2019 International Conference on Robotics and Automation (ICRA)*, vol. 2019-May, pp. 9565–9571, IEEE, may 2019. 23

[75] C. Loukas and E. Georgiou, "Surgical workflow analysis with Gaussian mixture multivariate autoregressive (GMMAR) models: A simulation study," *Computer Aided Surgery*, vol. 18, no. 3-4, pp. 47–62, 2013. 23

[76] T. Beyl, P. Nicolai, M. D. Comparetti, J. Raczkowsky, E. De Momi, and H. Wörn, "Time-of-flight-assisted Kinect camera-based people detection for intuitive human robot cooperation in the surgical operating room," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 7, pp. 1329–1345, 2016. 23

[77] N. Ahmidi, L. Tao, S. Sefati, Y. Gao, C. Lea, B. B. Haro, L. Zappella, S. Khudanpur, R. Vidal, and G. D. Hager, "A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2025–2041, 2017. 23

[78] R. DiPietro, N. Ahmidi, A. Malpani, M. Waldram, G. I. Lee, M. R. Lee, S. S. Vedula, and G. D. Hager, "Segmenting and classifying activities in

robot-assisted surgery with recurrent neural networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 11, pp. 2005–2020, 2019. 23

[79] F. Nageotte, P. Zanne, C. Doignon, and M. De Mathelin, "Stitching planning in laparoscopic surgery: Towards robot-assisted suturing," *International Journal of Robotics Research*, vol. 28, no. 10, pp. 1303–1321, 2009. 24

[80] R. C. Jackson, V. Desai, J. P. Castillo, and M. C. Çavuşoğlu, "Needle-tissue interaction force state estimation for robotic surgical suturing," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-November, pp. 3659–3664, 2016. 24

[81] F. Zhong, Y. Wang, Z. Wang, and Y. H. Liu, "Dual-Arm Robotic Needle Insertion with Active Tissue Deformation for Autonomous Suturing," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2669–2676, 2019. 24

[82] S. A. Pedram, P. Ferguson, J. Ma, E. Dutson, and J. Rosen, "Autonomous suturing via surgical robot: An algorithm for optimal selection of needle diameter, shape, and path," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2391–2398, 2017. 24

[83] C. Staub, T. Osa, A. Knoll, and R. Bauernschmitt, "Automation of tissue piercing using circular needles and vision guidance for computer aided laparoscopic surgery," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 4585–4590, 2010. 24

[84] K. Watanabe, T. Kanno, K. Ito, and K. Kawashima, "Single-Master Dual-Slave Surgical Robot with Automated Relay of Suture Needle," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6343–6351, 2018. 24

[85] J. Schulman and M. Tayson-frederick, "A Case Study of Trajectory Transfer Through Non-Rigid Registration for a Simplified Suturing Scenario," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4111–4117, 2013. 24

[86] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical

needle guide and sequential convex optimization," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4178–4185, IEEE, May 2016. 24, 31, 43, 64

[87] D. L. Chow and W. Newman, "Improved knot-tying methods for autonomous robot surgery," *IEEE International Conference on Automation Science and Engineering*, pp. 461–465, 2013. 25, 64

[88] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber, "A system for robotic heart surgery that learns to tie knots using recurrent neural networks," *Advanced Robotics*, vol. 22, no. 13-14, pp. 1521–1537, 2008. 25

[89] A. Knoll, H. Mayer, C. Staub, and R. Bauernschmitt, "Selective automation and skill transfer in medical robotics: a demonstration on surgical knot-tying," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 8, pp. 384–397, dec 2012. 25

[90] D. L. Chow and W. Newman, "Trajectory optimization of robotic suturing," *IEEE Conference on Technologies for Practical Robot Applications, TePRA*, vol. 2015-Augus, pp. 1–6, 2015. 25

[91] S. Leonard, K. L. Wu, Y. Kim, A. Krieger, and P. C. Kim, "Smart tissue anastomosis robot (STAR): A vision-guided robotics system for laparoscopic suturing," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1305–1317, 2014. 25

[92] A. Krieger *et al.*, "Development and Feasibility of a Robotic Laparoscopic Clipping Tool for Wound Closure and Anastomosis," *Journal of Medical Devices*, vol. 12, no. 1, 2017. 25, 43

[93] R. Jansen, K. Hauser, N. Chentanez, F. Van Der Stappen, and K. Goldberg, "Surgical retraction of non-uniform deformable layers of tissue: 2D robot grasping and path planning," *IEEE International Conference on Intelligent Robots and Systems, IROS 2009*, pp. 4092–4097. 26

[94] S. Patil and R. Alterovitz, "Toward automated tissue retraction in robot-assisted surgery," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2088–2094, 2010. 26, 43

[95] R. Elek, T. D. Nagy, D. Nagy, T. Garamvölgyi, B. Takács, P. Galambos, J. K. Tar, I. J. Rudas, and T. Haidegger, "Towards surgical subtask

automation-blunt dissection," in *INES 2017 - IEEE 21st International Conference on Intelligent Engineering Systems, Proceedings*, vol. 2017-Janua, pp. 253–257, 2017. 26, 44, 64

[96] T. D. Nagy, M. Takacs, I. J. Rudas, and T. Haidegger, "Surgical subtask automation — Soft tissue retraction," in *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 000055–000060, IEEE, Feb 2018. 26, 44, 64

[97] A. L. Trejos, J. Jayender, M. T. Perri, M. D. Naish, R. V. Patel, and R. A. Malthaner, "Robot-assisted tactile sensing for minimally invasive tumor localization," *International Journal of Robotics Research*, vol. 28, no. 9, pp. 1118–1133, 2009. 27

[98] J. Back, P. Dasgupta, L. Seneviratne, K. Althoefer, and H. Liu, "Feasibility study- novel optical soft tactile array sensing for minimally invasive surgery," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Decem, no. c, pp. 1528–1533, 2015. 27

[99] S. McKinley, A. Garg, S. Sen, R. Kapadia, A. Murali, K. Nichols, S. Lim, S. Patil, P. Abbeel, A. M. Okamura, and K. Goldberg, "A single-use haptic palpation probe for locating subcutaneous blood vessels in robot-assisted minimally invasive surgery," *IEEE International Conference on Automation Science and Engineering*, pp. 1151–1158, 2015. ix, 27, 34, 35

[100] F. Campisano, S. Ozel, A. Ramakrishnan, A. Dwivedi, N. Gkotsis, C. D. Onal, and P. Valdastri, "Towards a soft robotic skin for autonomous tissue palpation," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 6150–6155, 2017. 27

[101] A. Bajo and N. Simaan, "Hybrid motion/force control of multi-backbone continuum robots," *The International Journal of Robotics Research*, vol. 35, pp. 422–434, apr 2016. 27

[102] E. Ayvali, A. Ansari, L. Wang, N. Simaan, and H. Choset, "Utility-Guided Palpation for Locating Tissue Abnormalities," *IEEE Robotics and Automation Letters*, vol. 2, pp. 864–871, apr 2017. 27

[103] K. A. Nichols and A. M. Okamura, "Methods to Segment Hard Inclusions in Soft Tissue During Autonomous Robotic Palpation," *IEEE Transactions on Robotics*, vol. 31, pp. 344–354, apr 2015. 27

[104] E. Ayvali, R. A. Srivatsan, L. Wang, R. Roy, N. Simaan, and H. Choset, "Using Bayesian optimization to guide probing of a flexible environment for simultaneous registration and stiffness mapping," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 931–936, IEEE, may 2016. 28

[105] P. Chalasani, L. Wang, R. Roy, N. Simaan, R. H. Taylor, and M. Kobilarov, "Concurrent nonparametric estimation of organ geometry and tissue stiffness using continuous adaptive palpation," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4164–4171, IEEE, may 2016. 28

[106] E. Constanciel, W. A. N'Djin, F. Bessiere, F. Chavrier, D. Grinberg, A. Vignot, P. Chevalier, J. Y. Chapelon, and C. Lafon, "Design and evaluation of a transesophageal HIFU probe for ultrasound-guided cardiac ablation: simulation of a HIFU mini-maze procedure and preliminary ex vivo trials," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, pp. 1868–1883, sep 2013. 28

[107] H. Wang, W. Kang, T. Carrigan, A. Bishop, N. Rosenthal, M. Arruda, and A. M. Rollins, "In vivo intracardiac optical coherence tomography imaging through percutaneous access: toward image-guided radio-frequency ablation," *Journal of Biomedical Optics*, vol. 16, no. 11, p. 110505, 2011. 28

[108] L. Yang, R. Wen, J. Qin, C. K. Chui, K. B. Lim, and S. K. Y. Chang, "A robotic system for overlapping radiofrequency ablation in large tumor treatment," *IEEE/ASME Transactions on Mechatronics*, vol. 15, no. 6, pp. 887–897, 2010. 28

[109] B. Su, J. Tang, and H. Liao, "Automatic laser ablation control algorithm for an novel endoscopic laser ablation end effector for precision neurosurgery," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-December, pp. 4362–4367, 2015. 28, 64

[110] N. Sarli, G. Del Giudice, S. De, M. S. Dietrich, S. D. Herrell, and N. Simaan, " Preliminary Porcine In Vivo Evaluation of a Telerobotic System for Transurethral Bladder Tumor Resection and Surveillance ," *Journal of Endourology*, vol. 32, no. 6, pp. 516–522, 2018. 28, 64

[111] F. Alambeigi, Z. Wang, Y. hui Liu, R. H. Taylor, and M. Armand, "Toward Semi-autonomous Cryoablation of Kidney Tumors via Model-Independent

Deformable Tissue Manipulation Technique," *Annals of Biomedical Engineering*, vol. 46, no. 10, pp. 1650–1662, 2018. 28

[112] S. Taktak, P. Jones, A. Haq, B. P. Rai, and B. K. Somani, "Aquablation: a novel and minimally invasive surgery for benign prostate enlargement," *Therapeutic Advances in Urology*, vol. 10, pp. 183–188, jun 2018. 28

[113] P. Gilling, R. Reuther, A. Kahokehr, and M. Fraundorfer, "Aquablation - Image-guided robot-assisted waterjet ablation of the prostate: Initial clinical experience," *BJU International*, vol. 117, no. 6, pp. 923–929, 2016. 28

[114] J. W. Martin, P. R. Slawinski, B. Scaglioni, J. C. Norton, P. Valdastri, and K. L. Obstein, "Assistive autonomoy in colonoscopy: Propulsion of a magnetic flexible endoscope," *Gastrointestinal Endoscopy*, vol. 89, no. 6, pp. AB76–AB77, 2019. 29

[115] A. M. Okamura, C. Simone, and M. D. O'Leary, "Force modeling for needle insertion into soft tissue," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 10, pp. 1707–1716, 2004. 30

[116] T. Osa, C. F. Abawi, N. Sugita, H. Chikuda, S. Sugita, H. Ito, T. Moro, Y. Takatori, S. Tanaka, and M. Mitsuishi, "Autonomous penetration detection for bone cutting tool using demonstration-based learning," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 290–296, 2014. 30

[117] M. C. Yip, D. G. Lowe, S. E. Salcudean, R. N. Rohling, and C. Y. Nguan, "Tissue tracking and registration for image-guided surgery," *IEEE Transactions on Medical Imaging*, vol. 31, no. 11, pp. 2169–2182, 2012. 30

[118] I. Peterlík, H. Courtecuisse, R. Rohling, P. Abolmaesumi, C. Nguan, S. Cotin, and S. Salcudean, "Fast elastic registration of soft tissues under large deformations," *Medical Image Analysis*, vol. 45, pp. 24–40, 2018. 30

[119] D. Navarro-Alarcon, H. M. Yip, Z. Wang, Y. H. Liu, F. Zhong, T. Zhang, and P. Li, "Automatic 3-D Manipulation of Soft Objects by Robotic Arms with an Adaptive Deformation Model," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 429–441, 2016. 30

[120] F. Alambeigi, Z. Wang, R. Hegeman, Y. H. Liu, and M. Armand, "Autonomous data-driven manipulation of unknown anisotropic deformable tissues using unmodelled continuum manipulators," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 254–261, 2019. 30

[121] C. Pappone, G. Ciconte, G. Vicedomini, J. O. Mangual, W. Li, M. Conti, L. Giannelli, F. Lipartiti, L. McSpadden, K. Ryu, M. Guazzi, L. Menicanti, and V. Santinelli, "Clinical Outcome of Electrophysiologically Guided Ablation for Nonparoxysmal Atrial Fibrillation Using a Novel Real-Time 3-Dimensional Mapping Technique," *Circulation: Arrhythmia and Electrophysiology*, vol. 11, mar 2018. 30

[122] R. Decker, A. Shademan, J. Opfermann, S. Leonard, P. C. W. Kim, and A. Krieger, "Performance evaluation and clinical applications of 3D plenoptic cameras," in *Next-Generation Robotics II; and Machine Intelligence and Bio-inspired Computation: Theory and Applications IX*, no. June, p. 94940B, jun 2015. 30

[123] A. Shademan, R. S. Decker, J. Opfermann, S. Leonard, P. C. Kim, and A. Krieger, "Plenoptic cameras in surgical robotics: Calibration, registration, and evaluation," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2016-June, pp. 708–714, 2016. 30, 32, 64

[124] E. Bloch, B. Thurin, P. Keane, S. Nousias, C. Bergeles, and S. Ourselin, "Retinal fundus imaging with a plenoptic sensor," in *Ophthalmic Technologies XXVIII*, vol. 1047429, p. 81, SPIE, feb 2018. 30

[125] N. T. Clancy, G. Jones, L. Maier-Hein, D. S. Elson, and D. Stoyanov, "Surgical spectral imaging," *Medical Image Analysis*, vol. 63, p. 101699, jul 2020. 30

[126] L. Yu, L. Hao, T. Meiqiong, H. Jiaoqi, L. Wei, D. Jinying, C. Xueping, F. Weiling, and Z. Yang, "The medical application of terahertz technology in non-invasive detection of cells and tissues: opportunities and challenges," *RSC Advances*, vol. 9, no. 17, pp. 9354–9363, 2019. 30

[127] S. Speidel, A. Kroehnert, S. Bodenstedt, H. Kenngott, B. Müller-Stich, and R. Dillmann, "Image-based tracking of the suturing needle during laparoscopic interventions," in *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling* (R. J. Webster and Z. R. Yaniv, eds.), vol. 9415, p. 94150B, mar 2015. 31

[128] Y. Gu, Y. Hu, L. Zhang, J. Yang, and G.-Z. Yang, "Cross-Scene Suture Thread Parsing for Robot Assisted Anastomosis based on Joint Feature Learning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 769–776, oct 2018. 31

[129] R. C. Jackson, R. Yuan, D.-L. Chow, W. S. Newman, and M. C. Cavusoglu, "Real-Time Visual Tracking of Dynamic Surgical Suture Threads," *IEEE Transactions on Automation Science and Engineering*, vol. 15, pp. 1078–1090, jul 2018. 31

[130] N. Padoy and G. D. Hager, "3D thread tracking for robotic assistance in tele-surgery," *IEEE International Conference on Intelligent Robots and Systems*, pp. 2102–2107, 2011. 31

[131] C. D'Ettorre, G. Dwyer, X. Du, F. Chadebecq, F. Vasconcelos, E. De Momi, and D. Stoyanov, "Automated Pick-Up of Suturing Needles for Robotic Surgical Assistance," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, pp. 1370–1377, May 2018. 31, 64

[132] P. Beigi, R. Rohling, T. Salcudean, V. A. Lessoway, and G. C. Ng, "Detection of an invisible needle in ultrasound using a probabilistic SVM and time-domain features," *Ultrasonics*, vol. 78, pp. 18–22, 2017. 31

[133] K. Mathiassen, D. Dall'Alba, R. Muradore, P. Fiorini, and O. J. Elle, "Robust Real-Time Needle Tracking in 2-D Ultrasound Images Using Statistical Filtering," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 966–978, 2017. 31

[134] F. Zhong and Y. Liu, "Image-based 3D pose reconstruction of surgical needle for robot-assisted laparoscopic suturing," *Chinese Journal of Electronics*, vol. 27, no. 3, pp. 476–482, 2018. 31

[135] M. Abayazid, R. J. Roesthuis, R. Reilink, and S. Misra, "Integrating deflection models and image feedback for real-time flexible needle steering," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 542–553, 2013. 32

[136] G. J. Vrooijink, M. Abayazid, S. Patil, R. Alterovitz, and S. Misra, "Needle path planning and steering in a three-dimensional non-static environment using two-dimensional ultrasound images," *International Journal of Robotics Research*, vol. 33, no. 10, pp. 1361–1374, 2014. 32

[137] N. A. Patel, T. van Katwijk, Gang Li, P. Moreira, Weijian Shang, S. Misra, and G. S. Fischer, "Closed-loop asymmetric-tip needle steering under continuous intraoperative MRI guidance," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4869–4874, IEEE, aug 2015. 32

[138] P. Moreira, K. J. Boskma, and S. Misra, "Towards MRI-guided flexible needle steering using fiber Bragg grating-based tip tracking," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4849–4854, IEEE, may 2017. 32

[139] N. Shahriari, J. R. Georgiadis, M. Oudkerk, and S. Misra, "Hybrid control algorithm for flexible needle steering: Demonstration in phantom and human cadaver," *PLOS ONE*, vol. 13, p. e0210052, dec 2018. 32

[140] M. Fu, A. Kuntz, R. J. Webster, and R. Alterovitz, "Safe Motion Planning for Steerable Needles Using Cost Maps Automatically Extracted from Pulmonary Images," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4942–4949, IEEE, oct 2018. 32

[141] G. Fagogenis, M. Mencattelli, Z. Machaidze, B. Rosa, K. Price, F. Wu, V. Weixler, M. Saeed, J. E. Mayer, and P. E. Dupont, "Autonomous robotic intracardiac catheter navigation using haptic vision," *Science Robotics*, vol. 4, p. eaaw1977, apr 2019. 32

[142] H. Saeidi, H. N. D. Le, J. D. Opfermann, S. Leonard, A. Kim, M. H. Hsieh, J. U. Kang, and A. Krieger, "Autonomous Laparoscopic Robotic Suturing with a Novel Actuated Suturing Tool and 3D Endoscope," in *2019 International Conference on Robotics and Automation (ICRA)*, vol. 2019-May, pp. 1541–1547, IEEE, may 2019. 32

[143] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, and H. Larochelle, "Brain tumor segmentation with Deep Neural Networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017. 34, 64

[144] P. Hu, F. Wu, J. Peng, P. Liang, and D. Kong, "Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution," *Physics in Medicine and Biology*, vol. 61, pp. 8676–8698, Dec 2016. 34, 64

[145] W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Prostate segmentation: an efficient convex optimization approach with axial symmetry using 3-D TRUS and MR images.," *IEEE transactions on medical imaging*, vol. 33, pp. 947–60, Apr 2014. 34

[146] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 2663–2674, Dec 2018. 34, 44, 64

[147] BrainLab, "iPlan RT." `https://www.brainlab.com/radiosurgery-products/iplan-rt-treatment-planning-software/`, 2020. Accessed: 04-05-2020. 34

[148] J. D. Opfermann, S. Leonard, R. S. Decker, N. A. Uebele, C. E. Bayne, A. S. Joshi, and A. Krieger, "Semi-autonomous electrosurgery for tumor resection using a multi-degree of freedom electrosurgical tool and visual servoing," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3653–3660, sep 2017. 35

[149] K. A. Nichols and A. M. Okamura, "Autonomous robotic palpation: Machine learning techniques to identify hard inclusions in soft tissues," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 4384–4389, 2013. 35

[150] S. McKinley *et al.*, "An interchangeable surgical instrument system with application to supervised automation of multilateral tumor resection," *IEEE International Conference on Automation Science and Engineering*, vol. nov 2016, pp. 821–826, 2016. 35, 43

[151] USA, Department of Defense, "Toward a next-generation trauma care capability: Foundational research for autonomous, unmanned, and robotics development of medical technologies forward award," 2017. 36

[152] S. O'Sullivan, N. Nevejans, C. Allen, A. Blyth, S. Leonard, U. Pagallo, K. Holzinger, A. Holzinger, M. I. Sajid, and H. Ashrafian, "Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 15, no. 1, p. e1968, 2019. 36

[153] European Parliament, “"European Parliament resolution of 16 February 2017 number:P8 TA 2017 0051",” 2017. 36

[154] European Parliament, “"Comprehensive European industrial policy on artificial intelligence and robotics",” 2018. 36

[155] S. O’Sullivan, S. Leonard, A. Holzinger, C. Allen, F. Battaglia, N. Nevejans, F. W. Leeuwen, M. I. Sajid, M. Friebe, H. Ashrafian, H. Heinsen, D. Wichmann, and M. Hartnett, “Anatomy 101 for AI-driven robotics: Explanatory, ethical and legal frameworks for development of cadaveric skills training standards in autonomous robotic surgery/autopsy,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, p. e2020, may 2019. 36

[156] A. A. B. Jamjoom, A. M. A. Jamjoom, and H. J. Marcus, “Exploring public opinion about liability and responsibility in surgical robotics,” *Nature Machine Intelligence*, vol. 2, pp. 194–196, apr 2020. 36

[157] R. Shah and S. Nagaraja, “Privacy with surgical robotics: challenges in applying contextual privacy theory,” 2019. 36

[158] T. Haidegger, “Autonomy for surgical robots: Concepts and paradigms,” *IEEE Transactions on Medical Robotics and Bionics*, vol. 1, no. 2, pp. 65–76, 2019. 36

[159] E. Datteri, “Predicting the long-term effects of human-robot interaction: A reflection on responsibility in medical robotics,” *Science and engineering ethics*, vol. 19, no. 1, pp. 139–160, 2013. 36

[160] B. C. Stahl and M. Coeckelbergh, “Ethics of healthcare robotics: Towards responsible research and innovation,” *Robotics and Autonomous Systems*, vol. 86, pp. 152–161, 2016. 37

[161] Z. Moghadamyeghaneh, M. H. Hanna, J. C. Carmichael, A. Pigazzi, M. J. Stamos, and S. Mills, “Comparison of open, laparoscopic, and robotic approaches for total abdominal colectomy,” *Surgical Endoscopy*, vol. 30, pp. 2792–2798, jul 2016. 42

[162] P. Steele *et al.*, “Current and future practices in surgical retraction,” *The Surgeon*, vol. 11, pp. 330–337, dec 2013. 42

[163] M. Liu and M. Curet, "A Review of Training Research and Virtual Reality Simulators for the da Vinci Surgical System," *Teaching and Learning in Medicine*, vol. 27, no. 1, pp. 12–26, 2015. 43, 61

[164] K. Catchpole *et al.*, "Safety, efficiency and learning curves in robotic surgery: a human factors analysis," *Surgical Endoscopy*, vol. 30, pp. 3749–3761, Sep 2016. 43

[165] J. C. Hu *et al.*, "Perioperative complications of laparoscopic and robotic assisted laparoscopic radical prostatectomy," *The Journal of urology*, vol. 175, no. 2, pp. 541–546, 2006. 43

[166] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015. xi, 43, 44, 63, 64, 68, 69

[167] M. Benkhadra *et al.*, "Flexibility of thiel's embalmed cadavers: the explanation is probably in the muscles," *Surgical and Radiologic Anatomy*, vol. 33, pp. 365–368, May 2011. 43, 67

[168] B. S. Peters, P. R. Armijo, C. Krause, S. A. Choudhury, and D. Oleynikov, "Review of emerging surgical robotic technology," *Surgical Endoscopy*, vol. 32, pp. 1636–1655, Apr 2018. 43

[169] Muradore *et al.*, "Development of a cognitive robotic system for simple surgical tasks," *International Journal of Advanced Robotic Systems*, vol. 12, 2015. 43

[170] H. Nakawala, R. Bianchi, L. E. Pescatori, O. De Cobelli, G. Ferrigno, and E. De Momi, ""Deep-Onto" network for surgical workflow and context recognition," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 4, pp. 685–696, 2019. 43

[171] C. D'Ettorre *et al.*, "Automated pick-up of suturing needles for robotic surgical assistance," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 1, no. c, pp. 1370–1377, 2018. 43

[172] Y. Gao *et al.*, "JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling," *Modeling and Monitoring of Computer Assisted Interventions*, pp. 1–10, 2014. 43

[173] A. Murali *et al.*, "Learning by observation for surgical subtasks: Multilateral cutting of 3D viscoelastic and 2D Orthotropic Tissue Phantoms," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015-June, no. June, pp. 1202–1209, 2015. 43

[174] Y. Li, F. Richter, J. Lu, E. K. Funk, R. K. Orosco, J. Zhu, and M. C. Yip, "Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2294–2301, 2020. 43

[175] D. Á. Nagy, T. D. Nagy, R. Elek, I. J. Rudas, and T. Haidegger, "Ontology-based surgical subtask automation, automating blunt dissection," *Journal of Medical Robotics Research*, vol. 3, no. 03n04, p. 1841005, 2018. 44

[176] F. Isensee *et al.*, "Brain Tumor Segmentation Using Large Receptive Field Deep Convolutional Neural Networks," in *Brain Tumor Segmentation Using Large Receptive Field Deep Convolutional Neural Networks*, pp. 86–91, 2017. 44

[177] Fedorov *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging*, vol. 30, pp. 1323–1341, nov 2012. 44

[178] B. Kehoe *et al.*, "Autonomous multilateral debridement with the Raven surgical robot," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1432–1439, IEEE, may 2014. 44

[179] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org. 45, 70

[180] J. Ko, M. Kim, and C. Kim, "2d-to-3d stereoscopic conversion: depth-map estimation in a 2d single-view image," in *Proc. SPIE 6696, Applications of Digital Image Processing XXX*, sep 2007. 46

[181] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 328–341, 2007. 47, 65

[182] G. Bradski, "The OpenCV Library," 2000. 47

[183] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. 47

[184] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, 2000. 47

[185] R. I. Hartley, "Theory and practice of projective rectification," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 115–127, 1999. 47

[186] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015. 48

[187] F. Milletari *et al.*, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, 2016. 49, 73

[188] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the 3rd International Conference on Learning representations*, pp. 1–15, Dec 2015. 49, 72

[189] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974. 49

[190] F. Fuentes-Hurtado, A. Kadkhodamohammadi, E. Flouty, S. Barbarisi, I. Luengo, and D. Stoyanov, "EasyLabels: weak labels for scene segmentation in laparoscopic videos," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 7, pp. 1247–1257, 2019. 58

[191] T. Moriya *et al.*, "Unsupervised segmentation of 3D medical images based on clustering and deep representation learning," in *Medical Imaging*, pp. 71–78, 2018. 58

[192] A. Rau *et al.*, "Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 7, pp. 1167–1176, 2019. 58

[193] F. Gers, "Learning to forget: continual prediction with LSTM," in *9th International Conference on Artificial Neural Networks: ICANN '99*, vol. 1999, pp. 850–855, IEE, 1999. 63, 65

[194] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," Apr 2018. 63, 65

[195] M.-T. Luong, H. Pham, and C. D. Manning, "Effective Approaches to Attention-based Neural Machine Translation," Aug 2015. 63

[196] A. Arbelle and T. R. Raviv, "Microscopy Cell Segmentation Via Convolutional LSTM Networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1008–1012, IEEE, Apr 2019. xiv, 64, 65, 71, 76

[197] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec 2017. 64

[198] A. Norouzi, M. S. M. Rahim, A. Altameem, T. Saba, A. E. Rad, A. Rehman, and M. Uddin, "Medical Image Segmentation Methods, Algorithms, and Applications," *IETE Technical Review*, vol. 31, pp. 199–213, May 2014. 64

[199] J. Bernal, N. Tajkbaksh, F. J. Sanchez, B. J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, K. Pogorelov, S. Choi, Q. Debard, L. Maier-Hein, S. Speidel, D. Stoyanov, P. Brandao, H. Cordova, C. Sanchez-Montes, S. R. Gurudu, G. Fernandez-Esparrach, X. Dray, J. Liang, and A. Histace, "Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 1231–1249, Jun 2017. 64

[200] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, p. 1, Mar 2019. 64

[201] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2019. 64

[202] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation," pp. 358–368, 2019. 64

[203] M. Fayyaz, M. H. Saffar, M. Sabokrou, M. Fathy, F. Huang, and R. Klette, "STFCN: Spatio-Temporal Fully Convolutional Neural Network for Semantic Segmentation of Street Scenes," pp. 493–509, 2017. 65

[204] L. Zhang, L. Lu, X. Wang, R. M. Zhu, M. Bagheri, R. M. Summers, and J. Yao, “Spatio-Temporal Convolutional LSTMs for Tumor Growth Prediction by Learning 4D Longitudinal Patient Data,” *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1114–1126, Apr 2020. 65

[205] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” Dec 2014. 65

[206] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches,” Sep 2014. 65

[207] S. Andermatt, S. Pezold, and P. Cattin, “Multi-dimensional Gated Recurrent Units for the Segmentation of Biomedical 3D-Data,” pp. 142–151, 2016. 65

[208] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, “Attention gated networks: Learning to leverage salient regions in medical images,” *Medical Image Analysis*, vol. 53, pp. 197–207, Apr 2019. 65

[209] A. More, “Survey of resampling techniques for improving classification performance in unbalanced datasets,” Aug 2016. 69

[210] P. Simard, D. Steinkraus, and J. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, vol. 1, pp. 958–963, IEEE Comput. Soc, 2003. 69

[211] R. J. Williams and J. Peng, “An Efficient Gradient-Based Algorithm for On-Line Training of Recurrent Network Trajectories,” *Neural Computation*, vol. 2, pp. 490–501, Dec 1990. 72

[212] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,” Feb 2015. 72

[213] S. Semeniuta, A. Severyn, and E. Barth, “Recurrent Dropout without Memory Loss,” Mar 2016. 73

[214] S. A. Taghanaki, Y. Zheng, S. Kevin Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, “Combo loss: Handling input

and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24–33, Jul 2019. 73

[215] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks," pp. 379–387, 2017. 73

[216] F. Wilcoxon, S. Katti, and R. A. Wilcox, "Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test," *Selected tables in mathematical statistics*, vol. 1, pp. 171–259, 1970. 78

[217] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, and P. Valdastri, "Autonomy in surgical robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, 2020. 81

[218] A. D. Peter Kazanzides, "CRTK - Collaborative Toolkit," 2021. [Online; accessed 21-April-2021]. 81

[219] C. D'Ettorre, A. Mariani, A. Stilli, F. R. y. Baena, P. Valdastri, A. Deguet, P. Kazanzides, R. H. Taylor, G. S. Fischer, S. P. DiMaio, A. Menciassi, and D. Stoyanov, "Accelerating Surgical Robotics Research: Reviewing 10 Years of Research with the dVRK," apr 2021. 81

[220] T. Kroeger, "Opening the door to new sensor-based robot applications - The reflexxes motion libraries," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 6–9, 2011. 81

[221] Z. Zhang, A. Munawar, and G. S. Fischer, "Implementation of a motion planning framework for the davinci surgical system research kit," in *The Hamlyn Symposium on Medical Robotics*, p. 43, 2014. 81

[222] S. Chitta, I. Sucan, and S. Cousins, "Moveit![ros topics]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012. 81

[223] F. Richter, S. Shen, F. Liu, J. Huang, E. K. Funk, R. K. Orosco, and M. C. Yip, "Autonomous robotic suction to clear the surgical field for hemostasis using image-based blood flow detection," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1383–1390, 2021. 82

[224] F. Richter, R. K. Orosco, and M. C. Yip, "Open-sourced reinforcement learning environments for surgical robotics," *arXiv preprint arXiv:1903.02090*, 2019. 82

[225] C. M. Heunis, B. F. Barata, G. Phillips Furtado, and S. Misra, “Collaborative Surgical Robots: Optical Tracking During Endovascular Operations,” *IEEE Robotics & Automation Magazine*, vol. 27, pp. 29–44, sep 2020. 82

[226] C. Schneider, C. Nguan, R. Rohling, and S. Salcudean, “Tracked "pick-Up" ultrasound for robot-assisted minimally invasive surgery,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 2, pp. 260–268, 2016. 82

[227] B. Hannaford, J. Rosen, D. W. Friedman, H. King, P. Roan, L. Cheng, D. Glozman, J. Ma, S. N. Kosari, and L. White, “Raven-ii: an open platform for surgical robotics research,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 954–959, 2012. 83

[228] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” *Pattern Recognition*, vol. 47, pp. 2280–2292, jun 2014. 83

[229] M. B. Lukas Pfeifhofer, “ROS package tuw_aruco,” 2019. [Online; accessed 14-April-2021]. 83

[230] S. Chitta, I. Sucan, and S. Cousins, “MoveIt!,” *IEEE Robotics and Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012. 84

[231] I. A. Sucan, M. Moll, and L. E. Kavraki, “The open motion planning library,” *IEEE Robotics & Automation Magazine*, vol. 19, no. 4, pp. 72–82, 2012. 84

[232] H. Yoshida, H. Fujimoto, D. Kawano, Y. Goto, M. Tsuchimoto, and K. Sato, “Range extension autonomous driving for electric vehicles based on optimal velocity trajectory and driving braking force distribution considering road gradient information,” *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, pp. 4754–4759, 2015. 84

[233] M. Kalakrishnan, S. Chitta, E. Theodorou, P. Pastor, and S. Schaal, “STOMP: Stochastic trajectory optimization for motion planning,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 4569–4574, 2011. 85

[234] N. Ratliff, M. Zucker, J. A. Bagnell, and S. Srinivasa, “CHOMP: Gradient optimization techniques for efficient motion planning,” in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 489–494, 2009. 85