



UNIVERSITY OF GENOVA

PHD PROGRAM IN BIOENGINEERING AND ROBOTICS

Registration Methods for Surgical Augmented Reality

by

Alberto Neri

Thesis submitted for the degree of *Doctor of Philosophy* (38° cycle)

February 2026

Veronica Penza, Leonardo S. Mattos
Paolo Massobrio

Supervisors
Head of the PhD program

Thesis Jury:

Sharib Ali, *University of Leeds*
Cristian A. Linte, *Rochester Institute of Technology*

External examiner
External examiner

Dibris

Department of Informatics, Bioengineering, Robotics and Systems Engineering

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Alberto Neri
February 2026

Acknowledgements

I would like to use this space to say something personal, to share who Alberto is beyond the engineer. Three years ago, at the beginning of my PhD, I believed I knew who I was. However, I could not have imagine how much I would change along the path that brought me here. It is surprising how a single moment, a single choice, can change the course of a life. I am grateful.

In February 2022, I arrived at IIT with three certainties: I wanted to work in robotic surgery; I did not want to pursue a PhD; I did not want to work with machine learning. Then, in the summer of that same year, just before a match in IIT's football tournament, Enrico asked me, naturally: "Now that you've been here for a few months, would you like to do a PhD?" It was only curiosity (he had nothing to offer me), but that question opened a door. My instinctive answer was: "Yes. I feel good here." "Then check the university website: around this time, there are often positions opening for November, and the deadlines are closing." Out of curiosity, I looked. There was an open call in the very same lab where I was already working. **The deadline was the next day.** A friend, Livio, had taught me that important decisions sometimes have to be made quickly. Deep down, I had already chosen. And yet I called my mother: "Mum, what should I do?" "Do what you feel," she said. And that is what I did. The latins said: *Faber est suae quisque fortunae*; I agree with them. I am grateful.

Over these three years, I have had unforgettable experiences. Among them, the strongest and most transformative was certainly the five months in **Boston**. Alone, on the other side of the world, with my dream of submitting to MICCAI. I was also fortunate to travel a lot and visit countless cities around the world. I tore my ACL playing football. I started skiing again after ten years. I laughed and was sad; I loved and got angry; I cried with joy and with pain. I am grateful.

In these years, I learned what **courage** means. It is not a gift you possess once and for all; it is something you train. If you practice it, it grows; if you let it go, it fades. I also understood that courage finds strength in **passion**: when they work together, they become a force that can carry you farther than you believed possible. Alongside them, I learned **awareness**: to look at myself honestly, to understand what I do and why I do it, to recognise progress and accept falls. Without awareness, there is no growth; without awareness, there is no authentic gratitude.

I am grateful.

And then there are **people**. Without them, nothing of what we are would have meaning. I thank my family, my mother and my father, who have supported me from the very first day, with a constant and quiet presence worth more than a thousand words. I owe everything to them. I thank my grandparents for what they gave me and for the gaze that still accompanies me: I wish I could hold them tight. I thank my friends, the ones who have always been there and the ones I met along the way: with them I have laughed, and I will keep laughing. I thank my colleagues, who endured my temperament and stubbornness for more than three years and made this journey a shared one. I thank Leonardo and the University of Genoa for making this journey possible.

I am grateful.

Finally, I thank myself for perseverance, effort, difficult choices, and the desire to improve. I feel fortunate to be Alberto Neri. I am grateful for my health, for my passions, for my growth, for who I have become and who I will become. Life has given me a lot, and I feel a responsibility to do my best *to create good for myself and others*. My **mission**.

I am infinitely grateful.

Abstract

Objective. Surgical safety remains a major challenge: even with the advent of minimally invasive surgery (MIS) and robotic systems, surgeons still rely heavily on their own experience to interpret preoperative images and mentally translate them into the operative field. In many procedures, particularly in laparoscopy, locating critical structures (e.g. vessels, tumours) is difficult, especially in altered anatomy, leading to longer operative times, increased cognitive load, and a higher risk of complications. Surgical augmented reality (AR) aims to alleviate these issues by overlaying preoperative 3D information onto the laparoscopic view, but this requires accurate (*Target Registration Error (TRE) < 5-6mm*) and robust **registration** under soft-tissue motion and deformation. This thesis addresses this need by developing and evaluating registration methods for surgical AR in laparoscopy, using only data available in a typical operating room setup (preoperative 3D models + laparoscopic video). It follows a *stepwise path*: establishing practical **rigid** baselines in realistic settings, building the **foundations** (benchmarks, datasets, and reconstruction tools) needed for deformable guidance, and finally proposing a **patient-specific non-rigid** approach for soft-tissue registration.

Approach. We proceed in three parts. *Part 1 — Traditional rigid methods:* We first investigate how far classic rigid methods can go in realistic settings. (i) A comparative study of human–computer interfaces for **manual** alignment aimed at selecting the most effective user interface and understanding the intrinsic limitations of the approaches. (ii) We then embed this interface into **EVA**, a surgical navigator combining **point-based** initialization, **manual** refinement, and optical **tracking**, and evaluate it *in vivo* on a porcine model for ureter identification to show how rigid AR guidance affects task performance. Finally, (iii) we extend EVA to **robot-assisted** laparoscopy, replacing external tracking with robotic **kinematics** to explore how existing robotic platforms can maintain rigid overlays without additional hardware.

Part 2 — Foundations for learning-based registration: we then turn to deep learning, aiming to understand its limits and to create the necessary resources for deformable methods. (iv) A structured **review** of surgical AR registration methods clarifies the landscape and gaps. (v) A **benchmark** of rigid complete-to-partial point cloud methods under controlled

deformation and visibility characterises when rigid deep models succeed and when they fail. (vi) Two **dataset** pipelines, including a curated set of **688** kidney meshes and point clouds with synthetic deformations and partial views, provide training and evaluation data tailored to surgical settings. In parallel, (vii) a **single-image NeRF** pipeline for intraoperative reconstruction in mostly rigid neurosurgical cases shows how 3D surfaces can be recovered under strict data constraints and used as front-ends for registration.

Part 3 — Non-rigid registration: building on these insights and tools, (viii) we finally propose a deep-learning pipeline for non-rigid point cloud registration that predicts dense correspondences with a **Transformer**-based architecture and applies a **physics-based** deformation model. Here we change the training paradigm from anatomy-agnostic to **patient-specific**, training and testing on the same anatomy.

Main results. *Rigid methods:* (i) Manual registration can yield accurate overlays but is time-consuming and operator-dependent; (ii) in the in vivo study, **EVA reduced ureter-identification time** versus a conventional view, highlighting the practical benefit of the proposed registration pipeline; (iii) the robot-assisted variant maintained accurate overlay in simulation using arm kinematics. *Foundations:* (v) The rigid point cloud **benchmark** shows **sub-centimetre TRE** when deformation is limited and the visible surface is sufficient, with marked degradation under large deformations/low visibility; (vi) the **kidney dataset** (688 models) and deformation/visibility pipelines enable reproducible training and tests; (vii) the **single-image NeRF** achieves plausible reconstructions compared with multi-view baselines in mostly rigid scenarios. *Non-rigid registration:* (viii) On synthetic soft-deformation data with limited visibility, the patient-specific approach **improves TRE by $> 4\times$** over the best rigid baseline (e.g., $22.36\text{mm} \rightarrow 4.82 \pm 3.33\text{ mm}$ at 5% partiality) and reaches **$< 3\text{ mm}$ TRE** at higher partialities; under very large deformations with anatomical variability, it outperforms rigid methods but remains above surgical accuracy targets.

Significance. The thesis provides **practical baselines**, shows that well-integrated AR can **shorten task time** and be usable in surgical settings, and demonstrates that robotic kinematics can sustain rigid overlays without external trackers. It delivers the **foundational assets**: benchmarks, datasets, and a constrained-reconstruction pipeline, needed for fair comparison and faster iteration. Finally, it introduces a **patient-specific non-rigid** pathway that substantially narrows the gap to clinically meaningful accuracy in laparoscopy, while clarifying the remaining challenges (large deformations, partial views) and pointing to next steps such as uncertainty-aware overlays and multi-organ constraints.

Table of contents

List of figures	x
List of tables	xvi
Introduction	1
1 Principles of surgical AR for minimally invasive procedures	2
1.1 Clinical context and motivation	2
1.1.1 Advantages in minimally invasive surgery	3
1.2 Background of AR in surgery	3
1.2.1 Input data	4
1.2.2 Registration	4
1.2.3 Metrics for registration	5
1.2.4 AR display technologies	8
1.2.5 Challenges of AR registration in surgery	9
2 Aim of this thesis	11
2.1 Contributions	11
2.2 Structure of this thesis	12
State-of-the-art	14
3 Surgical augmented reality registration methods: a review from traditional to deep learning approaches	15
3.1 Traditional registration methods	16
3.1.1 Rigid registration methods	16
3.1.2 Non-rigid registration methods	21

3.2	Deep Learning methods	26
3.2.1	Hybrid DL-enhanced methods	27
3.2.2	DL-based point cloud registration methods	32
I	Traditional rigid registration methods	44
	Introduction	45
4	Comparative usability study of human-computer interfaces for 3D model manipulation in surgical augmented reality application	46
4.1	Introduction	46
4.2	Materials and Methods	47
4.2.1	Experimental protocol	48
4.3	Results	48
4.4	Discussion	49
4.5	Conclusion	50
5	Evaluating the EVA surgical navigation system for ureteral identification in an in vivo porcine model	51
5.1	Introduction	51
5.2	Materials and Methods	54
5.2.1	EVA-AR surgical navigation system	54
5.2.2	Study protocol	55
5.3	Results and Discussion	59
5.3.1	Technical evaluation results	59
5.3.2	Clinical evaluation results	61
5.4	Conclusion	65
6	Augmented Reality navigation in robot-assisted surgery with a teleoperated robotic endoscope	67
6.1	Introduction	67
6.2	Material and Methods	69
6.2.1	Surgeon remote console	70
6.2.2	Robotic endoscopic camera	70
6.2.3	Teleoperation with remote center of motion	70
6.2.4	Hand-eye calibration	72
6.2.5	Augmented Reality navigation system	72

6.3	Experimental validation	74
6.3.1	System architecture	74
6.3.2	Experimental protocol	75
6.4	Results	76
6.5	Discussion and Conclusions	77
II	Foundations for learning-based registration	79
	Introduction	80
7	Benchmarking complete-to-Partial point cloud registration techniques for laparoscopic surgery	82
7.1	Introduction	82
7.2	Benchmarking protocol	85
7.2.1	Problem formulation	85
7.2.2	Competing methods	85
7.2.3	Dataset and pre-processing	90
7.2.4	Metrics	92
7.2.5	Experimental setup	93
7.3	Results	94
7.3.1	Evaluation on the IRCAD deformed dataset	94
7.3.2	Evaluation on the DePoll dataset	96
7.3.3	Discussion	97
7.4	Conclusion	99
8	Kidney point clouds dataset from KITS segmentation challenge	100
8.1	Introduction	100
8.2	Methods	101
8.2.1	Filtering on 3DSlicer	102
8.2.2	STL split	103
8.2.3	Manual check	104
8.2.4	Conversion to point cloud	106
8.3	Conclusion	106
9	Surgical neural radiance fields from one image	107
9.1	Introduction	107
9.2	Methods	108

9.3	Experiments and Results	111
9.4	Conclusions	114
III	Deep Learning non-rigid point cloud registration method	115
	Introduction	116
10	Towards patient-specific deformable registration in laparoscopic surgery	117
10.1	Introduction	117
10.2	Methods	118
10.3	Results	123
10.4	Conclusion	126
	Conclusion	127
10.5	Thesis contributions	128
10.6	Impact	129
10.7	Future perspective	130
	References	133

List of figures

1.1	Visual interpretation of metrics. (a) TRE; the registration algorithm uses the dotted blue and red points as input. The dark blue and dark red stars indicate internal (beneath the surface) landmarks used to compute the TRE. In practice, these points are accessed via minimally invasive fiducial markers placed before surgery and tracked during the procedure. (b) FRE is measured as the Euclidean distance between corresponding dark blue and dark red stars on the kidney surface. Unlike the subsurface landmarks in (a), these surface points are directly used by the registration algorithm during its optimization. (c) IoU is defined as the ratio of overlapping pixels (intersection) to the total pixels covered by either mask (union). A perfect registration achieves an IoU of 1.	8
1.2	AR display technologies described in this chapter. (a) Display monitor in the operating room. It allows reproduction of 2D or 3D images. (b) Video see-through approach with tablet, proposed by Yasuda et al. (2018). (c) Microsoft Hololens in an example of AR assisted spinal surgery.	8
3.1	Overview of surgical AR registration methods. Inputs: a preoperative 3D organ model and intraoperative 2D or stereoscopic (3D) video. Registration estimates a rigid/non-rigid transform that aligns the model with the scene; the transformed model is then overlaid to produce the AR view.	15
3.2	Some of the most commonly used HCI for manual registration in recent years. (a) Mouse and keyboard. (b) 3D mouse. (c) UltraLeap.	18
3.3	Schematic overview of traditional registration methods. Each horizontal lane corresponds to a method category, while the vertical columns represent the common pipeline stages: input data, preprocessing, registration core, and registration output.	26

3.4	Popularity of deep learning in medical imaging papers and registration papers. As observed in Haskins et al. (2020), registration papers emerged slightly later than medical imaging papers and are now following a similar trend. . .	27
3.5	Types of matching used in registration approaches. (a) Point matching focuses only on the point coordinates (ICP). (b) Feature matching, in addition to the coordinates, uses other information typically extracted using deep learning methods. Features could be local or global, depending on the scale of the information they capture.	29
3.6	Schematic overview of Hybrid DL-Enhanced methods. Each horizontal lane corresponds to a method category, while the vertical columns represent the common pipeline stages: input data, preprocessing, registration core, and registration output.	32
3.7	Two examples of kidney registration in different conditions. (a) Complete-to-complete registration with exact one-to-one correspondences. In the registered plot, each point of the <i>src</i> (red) falls within the diamond of the <i>tgt</i> (blue). (b) Partial-to-partial registration with noise and density variation. The input point clouds partially describe the kidney, sharing an overlap area. In this case, there are no exact one-to-one correspondences not even in the overlap area, as shown in the registered plot.	37
3.8	Distribution-to-distribution matching. GMMs represent point clouds as a combination of Gaussians, depicted in the figure as coloured circles. Each point is assigned to one of these Gaussians. The algorithm searches for matches between the Gaussians of each GMM.	40
3.9	Schematic overview of DL-based point cloud registration methods.	43
4.1	Overview of the AR system architecture and the HCI used to perform the manual registration.	47
4.2	(a) IOU and execution time box plots (blue background). (b) SURG-TLX dimensions (yellow background).	49
5.1	An example of AR visualization of EVA Surgical Navigation System during the in-vivo study on animal model.	53
5.2	The EVA surgical navigation system: on the left, the 3D models of the pig's selected organs generated with HA3D™ (Medics, Italy) and the intraoperative endoscopic image; on the right, the system medical cart, composed of a screen, an optical tracking system and a computer.	55

5.3	Positioning of EVA Surgical Navigation System in a OR-like setting.	60
5.4	An instance of identifying anatomical landmarks during the calibration process between the preoperative 3D models and the animal model.	61
5.5	Sample frames from ureter identification task with the support of EVA-AR. In these pictures, the surgeon was indicating with the instrument the location of the ureter.	62
5.6	Face Validity Questionnaire Results.	62
5.7	System Usability Scale (SUS) Results.	64
5.8	UEQ questionnaire results	64
5.9	Surg-TLX questionnaire results. The y-axis was truncated at 50 for better visibility of the results.	65
6.1	Overview of the AR navigation system integrated with the Robotic Endoscopic Camera teleoperated through the Surgeon Remote Console.	68
6.2	Robot-integrated Augmented Reality Navigation System Architecture.	69
6.3	Reference frames of all the individual components of the teleoperation console and the robotic endoscope respectively namely: $\{C\}$: camera frame, $\{H\}$: haptic interface frame, $\{R\}$: robot's end-effector, $\{M\}$: RCM, $\{W\}$: robot's base, ${}^{\{R\}}T_{\{C\}}$: transformation between the robot end-effector and the camera link, ${}^{\{C\}}T_{\{W\}}$: transformation between the robot's basis and the camera link and $\{T\}$: the calibration target.	69
6.4	On the left, 3D design of the adapter to attach the endoscopic camera to the robot end-effector. On the right, an assembly of the adapter and the endoscope, with their reference systems.	70
6.5	2D and 3D illustration of the displacement Δx and Δy applied by the haptic interface. l is the distance between the RCM and end-effector $\{C\}$. θ and ϕ are the resulting angles.	71
6.6	Steps to reproduce rotation around RCM. (1) Compute the RCM pose starting from EE pose and applying $T1$ (rigid translation). (2) Apply ΔQ to find the new RCM orientation. (3) Find final EE pose applying $T2$ (rigid translation) to RCM.	72
6.7	Plot of the $\{R\}$ (light green), the $\{M\}$ (red) and the $\{C\}$ (blue) during a cross-shape movement of the Phantom Omni of 42cm around x and y axis and 6cm insertion depth. The dark green lights represent the endoscope length.	76

6.8	Example of initial manual registration between the 3D virtual organ model and the image of the corresponding kidney phantom. As an example, a representation of a lesion is visible thanks to the transparency of the virtual model.	77
6.9	Sample images of the AR guidance visualization in four different positions of the teleoperated robotic endoscopic camera during the experiments. The kidney virtual 3D model remains overlaid with high accuracy on the image of the real organ independently of the position.	77
7.1	DL-based point cloud registration methods use organ point clouds extracted from the preoperative model and intraoperative stereo-video as inputs and estimate the transformation that aligns the two point clouds, allowing the creation of the AR image for surgical navigation.	83
7.2	The network leverages KPConv to extract keypoints and their associated features. These features are further refined using self- and cross-attention mechanisms and then passed to the decoder, which predicts overlap scores and the corresponding keypoints. This information is ultimately used to estimate the final transformation.	88
7.3	An example of synthetic deformation applied to the complete preoperative liver (red), where the right lobe is shifted upward (blue). The final intraoperative cloud is then cropped from the deformed model (blue).	91
7.4	(a) The complete preoperative pig liver point cloud with segmented lobes. (b) The intraoperative video reconstructed point cloud (partial), relative to case 1. (c) The intraoperative CT scan point cloud (partial), relative to case 1. (d) Point cloud (a) and (b) registered with the ground truth rigid transformation. The dark red stars represent the surface marker on the preoperative anatomy; the blue stars represent the same surface markers on the intraoperative anatomy (after non-rigid deformations). The black arrows show the displacements between the corresponding markers, representing the effect of non-rigid deformation. The GT TRE is computed as the sum of the distances indicated by the black arrows. For simplicity, the figure shows only surface markers and the two clouds with a lower resolution.	93

7.5	Each model’s qualitative results were evaluated under four distinct levels of partiality (i.e., 5%, 10%, 15% and 25%). The first column illustrates the initial conditions (i.e. after pre-processing), the following shows the predicted registration, and the last provides the ground truth registration. . .	94
7.6	Qualitative registration results on the DePoll dataset, each row represents a different case. The inputs used were the preoperative complete point cloud and the intraoperative partial point cloud reconstructed from the video. . . .	96
8.1	The semi-automatic pipeline proposed in this work. The only manual step is the Manual Check, during which the user must inspect the STL models to determine if they are good, bad or need further filtering. All the other steps are automated.	102
8.2	Pipeline of the 3DSlicer macro. Initially, the segmentation is loaded into the environment. Then, a threshold filter is applied, but the surface is still wrinkled and irregular. Thus, a second smoothing filter is applied. Finally, the 3D models are saved as an STL file. The pipeline is iterated over every segmentation in the original dataset.	103
8.3	STL split logic. In both cases, a vertical plane is drawn at the origin, which serves as a cut region to split the STL. We proceed by calculating the spatial coordinates of the centroids of the two splits and the distance d between them. (a) If $d > threshold$, the splits are saved in two separate STLs. (b) If $d < threshold$, we have only one model and save it in a single STL. The threshold was found empirically.	104
8.4	(a) Examples of discarded kidneys. It includes kidneys with a very irregular shape due to massive tumours, kidneys in which the tumour is missing, and kidneys resulting from poor segmentation. (b) Examples of refinable kidneys. They are promising as their shape resembles traditional kidneys, but we need to improve their surface quality through an additional smoothing filter. Staircase-like pattern are visible.	105
9.1	Method overview: the key aspect of our approach is to retain the preoperative poses $\{\mathbf{P}_i\}_i$ and transfer the intraoperative appearance $\mathbf{J}$ to the preoperative images $\{\mathbf{I}_i\}_i$ using S , an unpaired image-to-image translation. This process allows for the supervision of the training of a NeRF F_θ using only one image, circumventing the error-prone multi-view intraoperative acquisition. . . .	109

9.2	Tests on four real cases, one per row with pairs of synthesized images from three random viewpoints using MV-NeRF and our method, respectively. . .	112
9.3	Comparison against ground-truth image after registration.	113
9.4	Comparison of various NST methods. Our hybrid approach strikes an optimal balance between style and structure. For readability, PSNR and GMS are scaled to the range [0,1]: PSNR is normalized with a maximum value of 40dB, and GMS is computed using cosine similarity.	113
9.5	Interactive, semi-automated segmentation on synthesized images.	114
9.6	Proposed pipeline for deformable surgical augmented reality. A preoperative 3D organ model and an intraoperative laparoscopic image are first converted to a common representation (point clouds). A deep learning-based point cloud matching model estimates dense correspondences between the complete and partial point clouds. These correspondences are then used within a physics-based framework (SOFA) to perform non-rigid registration.	116
10.1	The network uses KPConv to detect keypoints and extract their features, refines these features through self- and cross-attention, and then feeds them into the decoder and matching module to predict dense point correspondences.	119
10.2	Qualitative results of predicted matches. Blue points denote the preoperative point cloud, while orange points represent the partial and deformed intraoperative cloud. Red lines indicate incorrect matches, and green lines highlight exact matches relative to the ground truth.	124
10.3	Non-rigid registration results on IRCAD-Liver1: Source point cloud is represented with gray points while target point clouds are represented in black wireframe. Green lines represent the matching correspondences and FEM models are represented with grid cells. Registered point clouds are presented at $t=0$, $t=T/2$ and $t=T$ are shown in 4 different views.	124
10.4	Qualitative results of non-registration on six cases from the DePoLL dataset are presented, showing the endoscopic image, the matching, and the registration from left to right. Preoperative point clouds are represented in gray, intraoperative point clouds in black, and correspondences with green lines.	125

List of tables

3.1	Overview of traditional rigid registration methods	16
3.2	Overview of traditional non-rigid registration methods	22
3.3	Overview of Hybrid DL-Enhanced Methods	28
3.4	Overview of DL-based Point Cloud Registration. Methods in bold are applied to the medical field.	34
5.1	Execution times for all trials with and without EVA-AR.	61
6.1	AR navigation evaluation	77
7.1	Summary of the methods compared	86
7.2	Performance Metrics at different levels of partiality. Legend: best , <u>second best</u>	95
7.3	Average Performance Metrics on DePoll dataset. Legend: best , <u>second best</u>	97
9.1	Average synthesize agreement with MV-NeRF across all poses.	112
10.1	Matching Metrics	123
10.2	Registration results on DePoLL dataset (results in mm).	125

Introduction

Chapter 1

Principles of surgical AR for minimally invasive procedures

1.1 Clinical context and motivation

The popularity of Minimally Invasive Surgery (MIS) over open surgery has expanded in recent decades, primarily due to its superior clinical advantages, such as reduced postoperative pain, shorter hospital stays, and faster recovery (Pang et al., 2023). However, MIS also introduces challenges, including a limited field of view, reduced depth perception, and loss of direct tactile feedback. Within MIS, Robotic-Assisted Surgery (RAS) has profoundly transformed practice by improving instrument dexterity, precision, and ergonomics, and by facilitating access to anatomically constrained regions (Taylor et al., 2016). Despite these benefits, RAS high acquisition and maintenance costs restrict its use to selected procedures, while standard laparoscopy remains the reference technique for most interventions (Leal Ghezzi and Campos Corleta, 2016). Moreover, RAS still shares the same fundamental limitations as conventional laparoscopy in terms of visualisation. In this context, surgical Augmented Reality (AR) has the potential to benefit *both* laparoscopic and robotic MIS by enhancing spatial awareness, alleviating depth-perception issues, and reducing the cognitive effort required to mentally fuse pre- and intraoperative data (Seetohul et al., 2023).

Following Milgram and Kishino's reality-virtuality continuum (Milgram and Kishino, 1994), mixed reality (MR) spans the range between a purely real environment and a purely virtual environment. AR lies toward the real end of this continuum, where the real world remains primary and virtual content is superimposed in a registered manner. In other words, AR is a technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view. Besides healthcare, AR is gaining popularity in

different contexts, from entertainment and gaming to automotive, manufacturing, construction and tourism. In surgery, AR was gradually introduced in different experimental medical settings, such as neurosurgery, orthopaedics and laparoscopy. While in orthopaedy the bones are treated as a rigid body, thereby benefiting the registration between the augmentation and the scene, in laparoscopy, effectively enhancing the scene remains challenging due to the lack of rigidity in abdominal tissues and organs. Consequently, surgical AR has gained significant interest within the research community, as evidenced by a substantial volume of related publications spanning the past two decades.

1.1.1 Advantages in minimally invasive surgery

AR technologies provide surgeons with a more informative and efficient environment by superimposing virtual elements onto the surgical field. For instance, these virtual elements may represent critical structures (i.e. organs, tumours, vessels), patient data, or navigation guidance. AR boosts the surgeon's situational awareness, allowing him to achieve more accurate incisions and interventions by exploiting navigation guidance based on preoperative imaging (Guha et al., 2017). Another benefit is reducing their cognitive load (Seetohul et al., 2023); in fact, by presenting relevant information within the surgeon's field of view, AR minimizes the need to shift attention between different displays. Moreover, AR fosters improved communication among surgical teams by offering a shared visual of the procedure. Lastly, it also serves as a potent tool for training and education, helping surgical trainees to understand and recognize the anatomy, and become aware of hidden anatomical structures.

Overall, AR contributes to safer, more precise, and more efficient surgical interventions, making it a valuable technology in the medical field.

1.2 Background of AR in surgery

Traditionally, surgeons have relied on imaging modalities such as X-rays, CT scans, and MRI for preoperative planning and intraoperative guidance. However, these modalities often require the surgeon to divert attention from the patient to external displays. For example, a situation that can occur in the operating room is locating an internal tumour not visible on the organ's surface. Currently, surgeons rely on preoperative 2D images or additional intraoperative imaging, if available. In both cases, the 2D images require the surgeon to mentally translate the 2D information into the real 3D anatomy, which can sometimes be

challenging. Moreover, intraoperative imaging, such as fluoroscopy or intraoperative CT, exposes the patient to additional radiation.

AR revolutionizes the surgical field by overlaying critical information directly onto the surgeon's view, minimizing the need to shift focus and providing a more natural and integrated approach to surgical navigation.

1.2.1 Input data

Various approaches to AR exist, often determined by the type of input data available. We will explore the conventional data encountered in laparoscopic surgery:

- **preoperative imaging.** Typically, patient-specific imaging data is acquired from traditional scanners (CT or MRI) in the preoperative phase, providing high resolution and good image quality. They are frequently subjected to digital processing, whether manual or automatic, to refine and/or isolate structures of interest, like organs, tumours, bones, or vessels. From these images it is possible to obtain a 3D model of the structure of interest, which is generally exploited to augment the surgical image. In addition to this, the acquisition of preoperative data is a widespread practice since it provides excellent information for diagnosis and surgical planning before the actual procedure.
- **intraoperative imaging.** It is provided by a real-time imaging device used during the surgery, such as a laparoscope or endoscopic camera. These devices allow for magnification of the surgical site, providing a detailed and close-up view of organs. Furthermore, stereoscopic images grant depth perception to surgeons. While the laparoscopic image is implicitly always present in the operating room, there are many other types of image that may be acquired depending on the specific surgical procedure or anatomical structure addressed. Some examples include fluoroscopy images, which are acquired using X-ray and used for guidance to visualize structures in real-time; ultrasound images, typically used for tumour localization; and intraoperative MRI or CT scans, used for guidance in specific settings.

Combining the preoperative 3D model with the intraoperative laparoscopic image and solving the registration problem allows us to obtain an AR image.

1.2.2 Registration

Image registration is a fundamental process in computer vision and medical imaging, which implies aligning multiple images of the same scene or object captured from varying per-

spectives, times, or sensors. The principal purpose of image registration is to identify a transformation that maps one image (the source or reference) with another image (the target), ensuring geometrical alignment. Registration can be divided into two categories: rigid and non-rigid. In rigid registration, it is assumed that mapping between the images can be described by affine transformations, which include translation, rotation and scaling. On the contrary, we speak of non-rigid registration when considering deformations, such as bending, stretching, or twisting. In such cases, the number of degrees of freedom (DoF) in the mapping is undefined, allowing for more complex transformations.

Besides the transformation model (rigid / non-rigid), there are other ways to classify medical image registration:

- **Input data:** unimodal / multimodal; inter-patient / intra-patient.
- **Dimensionality:** 3D–3D, 3D–2D, 2D–2D, or 2D–3D.
- **Anatomical site:** grouped by target organ or region (e.g., brain, lung, liver).

We study a specific surgical AR scenario. *A 3D model obtained from preoperative CT/MRI is aligned (registered) to the intraoperative laparoscopic video for guidance.* This particular type of medical image registration is multimodal, intra-patient, deformable, and 3D-2D or 3D-3D (based on stereoscopy). The key difference between medical image registration and AR registration lies in their purposes and applications. Image registration primarily deals with aligning static medical images for analysis and diagnosis. In contrast, AR registration aligns virtual information with real-world scenes for enhanced visualization and guidance during medical procedures.

1.2.3 Metrics for registration

In surgical augmented reality, the evaluation process involves a range of metrics that address the system's technical performance and clinical utility.

Target Registration Error

One of the primary metrics is the Target Registration Error (TRE), which quantifies the accuracy of the spatial alignment between virtual content and real-world anatomical structures. TRE is determined by identifying one or more target points with known ground truth coordinates, often acquired from external references such as fiducial markers placed explicitly for this purpose (Fig. 1.1(a)). The Euclidean distances between these known positions and

the positions computed by the registration algorithm are then calculated, with the average or maximum error commonly reported as the TRE. Eq. 1.1 describes the TRE:

$$\text{TRE}_i = \|\mathbf{p}_i^{\text{reg}} - \mathbf{p}_i^{\text{gt}}\|_2, \quad i = 1, \dots, N, \quad \overline{\text{TRE}} = \frac{1}{N} \sum_{i=1}^N \text{TRE}_i. \quad (1.1)$$

where $\mathbf{p}_i^{\text{gt}}$ is a set of N ground-truth target landmarks and $\mathbf{p}_i^{\text{reg}}$ their registered counterparts. Specifically, each $\mathbf{p}_i$ should not be included in the input point clouds used by the registration algorithm. This metric is particularly important because even minor misalignments can lead to significant clinical risks, such as inaccurate instrument guidance or misinterpretation of critical anatomical boundaries during surgery.

TRE Thresholds In current surgical practice, acceptable registration accuracy depends on the procedure and the rigidity of the anatomy.

In **neurosurgery**, millimetric precision is required because targets (e.g. small tumours) are often located close to crucial structures, where even a few millimetres of error can lead to major neurological deficits. For this reason, neuronavigation and AR systems for cranial surgery are generally expected to keep target errors on the order of ~ 2 mm or less (Hey et al., 2023; Kim and Kim, 2021). A similar constraint applies in **spinal and orthopaedic** navigation, where tasks such as pedicle screw placement are performed close to the spinal cord and nearby nerves. Here, guidance systems must achieve high spatial accuracy to avoid cortical damage and neural injury. Clinical studies of AR spine navigation report mean screw-placement errors of around 2 mm (Molina et al., 2021). In laparoscopic soft-tissue surgery, tasks such as tumour resection or identification of vessels are performed on relatively large organs, with safety margins of a few millimetres (Östrand et al., 2024) and additional intraoperative checks (visual cues, ultrasound, etc.) helping to compensate for small localisation errors. Within this practical context, several works indicate that registration errors up to about **5–6 mm** are still considered clinically acceptable for AR guidance (Doornbos et al., 2024; Köhler et al., 2022; Teatini et al., 2021). Taken together, these findings suggest that ~ 5 mm TRE is a *reasonable* threshold for deployable laparoscopic AR today, but that approaching the ~ 2 mm range already used in rigid domains (neurosurgery and spine) would be an **optimal** long-term target for soft-tissue laparoscopy as well.

Fiducial Registration Error

Fiducial Registration Error (FRE) is closely related to TRE because it quantifies error using fiducial points. The key distinction lies in their evaluation: FRE computes errors based on the same points used in the registration algorithm, while TRE measures errors at target points not included in the registration process (Fig. 1.1(b)); thus, the FRE formula is the same as Eq. 1.1, but with the $\mathbf{p}_i$ used by the registration algorithm. FRE is effective for assessing the registration quality of the organ's visible regions during surgery, while TRE offers a more comprehensive evaluation of the overall registration performance.

Intersection over Union

A standard way to measure AR overlay quality is the Intersection over Union (IoU). This metric is frequently used to judge AR systems, particularly when they are required to recognize anatomical features and project them in real-time. IoU quantifies how well the 2D projection of the 3D model overlaps the anatomy visible in the laparoscopic image. In practice, it quantifies the agreement between the projected model contour (from the AR system) and the reference organ region in the 2D image. It is computed as (Eq. 1.2):

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}, \quad (1.2)$$

where A and B are the binary masks for the projected 3D model and the 2D reference organ region, respectively. A higher IoU score implies better correspondence and reliability of the AR overlay (Fig. 1.1(c)).

Usability metrics

Other key metrics include processing time and frames per second (FPS). Processing time, which covers the duration needed for registration, rendering, and visualization, directly influences the FPS ratio, indicating the smoothness of the visual overlay.

Beyond these technical measures, usability metrics play an important role. User studies, typically involving surgeons and other clinical staff, provide qualitative assessments of the system's intuitiveness and ease of integration into the surgical workflow. These studies may include task completion rates, user satisfaction scores, and error rates during simulated or actual surgical procedures. Moreover, qualitative assessments through visual inspection of the AR overlays and feedback from clinical trials contribute to understanding the practical implications of the technology. To systematically capture these insights, researchers commonly

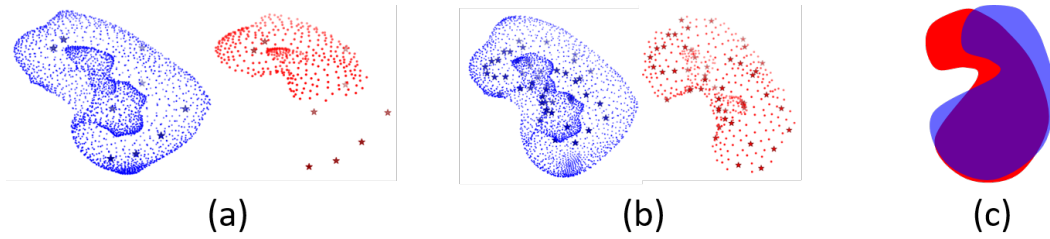


Figure 1.1 Visual interpretation of metrics. (a) TRE; the registration algorithm uses the dotted blue and red points as input. The dark blue and dark red stars indicate internal (beneath the surface) landmarks used to compute the TRE. In practice, these points are accessed via minimally invasive fiducial markers placed before surgery and tracked during the procedure. (b) FRE is measured as the Euclidean distance between corresponding dark blue and dark red stars on the kidney surface. Unlike the subsurface landmarks in (a), these surface points are directly used by the registration algorithm during its optimization. (c) IoU is defined as the ratio of overlapping pixels (intersection) to the total pixels covered by either mask (union). A perfect registration achieves an IoU of 1.

employ standardized questionnaires such as the System Usability Scale (SUS) (Brooke et al., 1996), the User Experience Questionnaire (UEQ) (Laugwitz et al., 2008), and SURG-TLX (Wilson et al., 2011).

1.2.4 AR display technologies

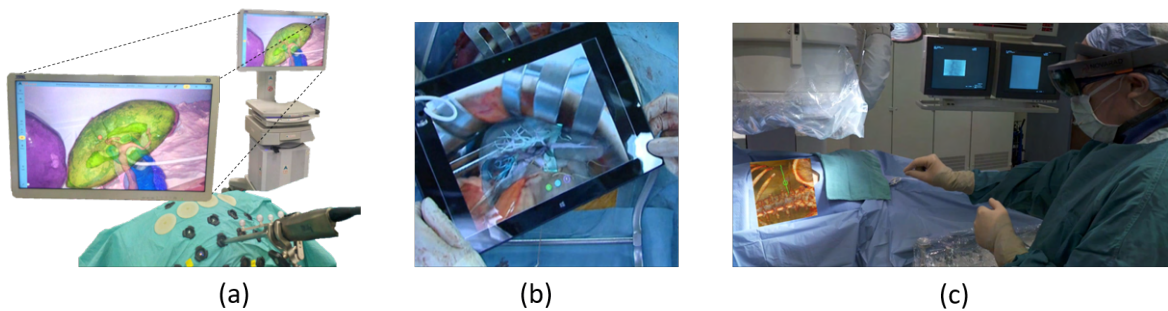


Figure 1.2 AR display technologies described in this chapter. (a) Display monitor in the operating room. It allows reproduction of 2D or 3D images. (b) Video see-through approach with tablet, proposed by Yasuda et al. (2018). (c) Microsoft HoloLens in an example of AR assisted spinal surgery.

The effectiveness of AR depends on advanced display technologies that integrate computer-generated data with the surgeon's view. In surgical applications, the choice of display plays a pivotal role in providing accurate, high-quality, and real-time visual feedback to the surgeon.

In the context of MIS, the main alternative is AR with monitor (Fig. 1.2a). This is achieved by superimposing the preoperative 3D models onto the endoscopic image shown by the display monitor in the operating room. Monitor-based displays are a core component of computer-integrated surgery and are one of the most common platforms for AR. Because the same workstation controls both the live video and the virtual overlay, they can be shown with minimal relative latency (Qian et al., 2019). The growing use of 3D monitors, which provide stereoscopic viewing with dedicated glasses, further improves depth perception. This display mode is particularly suitable for MIS, where the surgeon already operates by looking at the endoscopic image on a screen rather than directly at the surgical field.

In other scenarios, such as open surgery, there are more alternatives, including Optical and Video see-through, which have become less popular recently. One recent endeavour involving video see-through is presented in Yasuda et al. (2018), where a tablet PC equipped with an infrared position marker is used (Fig. 1.2b). This setup enables the superimposition of a 3D organ model onto the tablet PC display that captures real-time surgical field images. On the contrary, compared to the analysis done by Bernhardt et al. (2017a) in 2016, methods involving projection onto the patient evolved with the Head Mounted Display (HMD) technology. HMDs are wearable devices (headsets) that allow surgeons to visualize augmented information directly within their field of view. As highlighted by Barcali et al. (2022), the most used devices are the Microsoft HoloLens 1 and 2 (Fig. 1.2c). For instance, Gouveia et al. (2021) exploited the HoloLens in breast surgery, while Cofano et al. (2021) in spine surgery (Fig. 1.2c). Although these technologies are fascinating, there are papers like Condino et al. (2019) and Carbone et al. (2020) that recognize them as inadequate to perform high-accuracy tasks, such as surgery. Indeed, the HMD-assisted surgery field has recently experienced a notable increase in publications and diverse surgical applications. However, only some systems have been used clinically, and demonstration of utility is rare, meaning that further research is needed to achieve satisfactory results, as concluded in Birlo et al. (2022). A final remark concerns a potential disadvantage of AR in surgery, i.e., the fact that the overlaid AR can cause inattention blindness, where the augmented view distracts from essential events in the real view. Nonetheless, this potential safety issue can be effectively managed and the benefits of AR, as outlined in the next section, may largely outweigh the risks.

1.2.5 Challenges of AR registration in surgery

The surgical context implicitly introduces significant challenges due to the patient's physiology, which generates natural motions such as breathing and heartbeat. Furthermore, the

preoperative image represents the structure of interest at the moment in time in which it was acquired. Thus, notable alterations in the patient's anatomy may occur between the preoperative acquisition and the intervention, making it challenging to find correspondences. For example, during laparoscopic surgery, some possible causes of shifts in abdominal structures are the patient's body positioning and pneumoperitoneum insufflation. Finally, the interaction of surgical instruments with organs must also be considered. All this causes the scene to be dynamic; therefore the problem to be addressed is the non-rigid registration. Research on both rigid and non-rigid registration is still needed, as there is still no commercially available surgical AR technology capable of solving the problem satisfactorily for surgeons.

Chapter 2

Aim of this thesis

This thesis focuses on developing and evaluating **non-rigid registration** methods for **surgical augmented reality** under typical operating-room constraints. We target the routine operating room scenario in which the inputs are a **3D model derived from preoperative CT/MRI** and an **intraoperative laparoscopic video**. This choice reflects typical practice: intraoperative CT / MRI is often unavailable and involves additional costs and radiations, while extracting preoperative 3D models and using a stereovision laparoscope are standard practices.

With many approaches available, we investigated alternatives to understand their strengths and limits. Therefore, the main objectives of this PhD thesis are summarized in:

- **Develop and evaluate traditional rigid baselines** (manual and point-based) to characterize task and failure modes.
- **Develop and evaluate deep-learning methods for rigid complete-to-partial point cloud registration.**
- **Develop and evaluate complementary components for surgical AR**, including dataset generation strategies and intraoperative surface reconstruction.
- **Develop and evaluate a deep-learning method for non-rigid** point cloud registration tailored to soft-tissue deformation.

2.1 Contributions

At a high level, this thesis contributes to surgical augmented reality for minimally invasive surgery along three main axes:

- **Traditional rigid registration for surgical AR.** We design and evaluate three rigid registration workflows in realistic settings: a comparative study of human–computer interfaces for manual 3D model alignment; an *in vivo* porcine study with our surgical navigation system (EVA Navi) combining point-based initialisation, manual refinement, and optical tracking for ureter identification; and an extension of EVA Navi to robot-assisted surgery, where the AR overlay is updated in real time from robotic kinematics.
- **Building blocks for learning-based point cloud registration.** We lay the groundwork for learning-based non-rigid registration by (i) providing a structured review of registration and surgical AR methods, (ii) benchmarking traditional and deep-learning rigid complete-to-partial point cloud registration methods under controlled deformation and visibility, and (iii) proposing dataset-generation strategies, including a curated set of 688 kidney 3D models and a pipeline for creating realistic complete-to-partial training pairs. In addition, (iv) we introduce a single-image NeRF-based intraoperative surface reconstruction method for rigid or near-rigid settings, illustrating how 3D surfaces can be recovered under strict data constraints.
- **Patient-specific non-rigid point cloud registration with deep learning.** Finally, we propose, to the best of our knowledge, the first *patient-specific* deep-learning framework for non-rigid point cloud registration in surgical AR. The method combines a Transformer-based correspondence network with a physics-based deformation model. It is trained and tested on the same patient-specific anatomy using on-the-fly deformations and partial views. On synthetic data with soft deformations, it improves TRE by more than a factor of four over the best rigid baseline and achieves sub-millimetre-range accuracy under favourable visibility conditions. Experiments on highly deformable liver data highlight both its potential and the remaining challenges for robust non-rigid AR in extreme deformation scenarios.

2.2 Structure of this thesis

The dissertation is organized into:

Introduction sets the clinical context and technical foundations of surgical AR for minimally invasive procedures. *Chapter 1* outlines concepts, data, registration, challenges, display technologies, and benefits. *Chapter 2* summarises the aim of the thesis and the contributions.

State-of-the-art then focuses on existing work. *Chapter 3* reviews current surgical AR registration methods and related approaches, highlighting open gaps and motivating the research directions pursued in the rest of the thesis.

Part I (Traditional Rigid Registration Methods) examines rigid baselines and their usability and feasibility. In particular, *Chapter 4* reports a comparative usability study of human-computer interfaces for manual registration. The best-performing interface is then adopted in *Chapter 5*, presenting the EVA navigation system, which combines point-based and manual registration, evaluated for ureteral identification in an *in vivo* porcine model. Finally, in *Chapter 6* we extend the EVA system to implement AR navigation in robot-assisted surgery using a teleoperated robotic endoscope.

In **Part II** (Foundations for Learning-Based Registration), the focus shifts to deep-learning methods for rigid registration and the resources needed to support them. *Chapter 7* presents a benchmark of complete-to-partial point cloud registration techniques for laparoscopy, clarifying where rigid methods break down and establishing baselines for the subsequent non-rigid extension. *Chapter 8* introduces a dataset of kidney 3D models generated from CT segmentations provided by the KiTS challenge (MICCAI). *Chapter 9* describes an intraoperative surface reconstruction method based on Neural Radiance Field (NeRF) from one image.

Part III (Deep Learning Non-Rigid Point Cloud Registration) brings these elements together. *Chapter 10* introduces a patient-specific deformable registration approach for laparoscopy, leveraging deep learning and a physics-based deformation model to handle soft-tissue motion.

Finally, the **Conclusion** synthesises the findings and provides directions for future work.

State-of-the-art

Chapter 3

Surgical augmented reality registration methods: a review from traditional to deep learning approaches

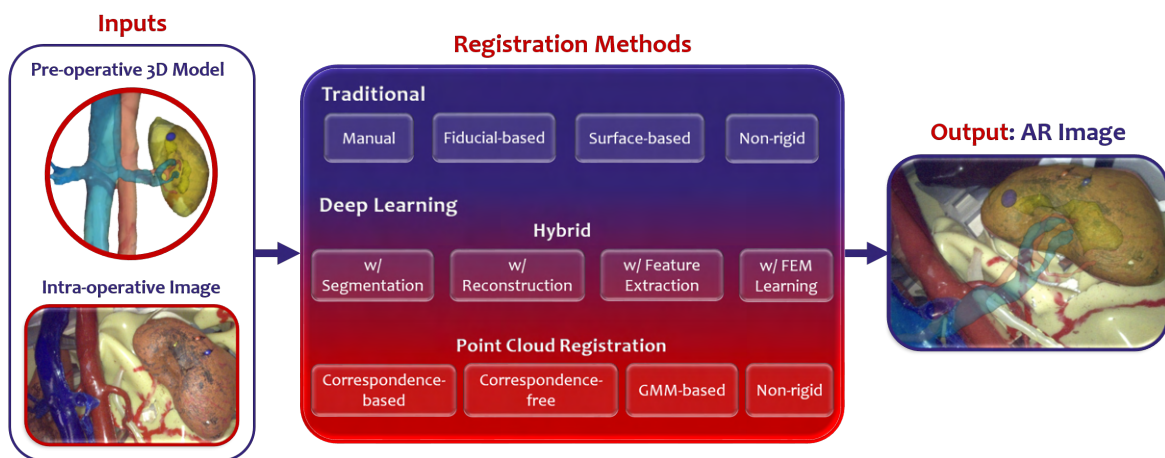


Figure 3.1 Overview of surgical AR registration methods. Inputs: a preoperative 3D organ model and intraoperative 2D or stereoscopic (3D) video. Registration estimates a rigid/non-rigid transform that aligns the model with the scene; the transformed model is then overlaid to produce the AR view.

This chapter explores the state-of-the-art in surgical AR registration methods, moving from traditional pipelines to modern learning-based systems. We begin with **traditional**

This work has been published as: Neri, A., Penza, V., Baldini, C., & Mattos, L. S. (2025). Surgical augmented reality registration methods: A review from traditional to deep learning approaches. *Computerized Medical Imaging and Graphics*. Publicly available at <https://doi.org/10.1016/j.compmedimag.2025.102616>.

methods, including *manual, fiducial-based, surface-based, and deformable approaches*. We then examine **hybrid, deep learning enhanced methods** that keep a classical structure but introduce learned components at key stages, such as *segmentation, 3D reconstruction, feature extraction, and FEM-based deformation learning*, to improve robustness and accuracy. Finally, we review **end-to-end deep learning for point-cloud registration**, covering *correspondence-based, correspondence-free, GMM-based, and non-rigid models*. Each category includes a brief discussion outlining its advantages, disadvantages, and suitability for use in laparoscopic procedures. Figure 3.1 provides a visual summary of the methods presented.

3.1 Traditional registration methods

3.1.1 Rigid registration methods

Table 3.1 Overview of traditional rigid registration methods

Reference	ROI	Approach	Peculiarity
Penza et al. (2023)	Kidney	Manual + Update	3D Mouse
Krug et al. (2023)	Liver	Manual	Ultraleap
Penza et al. (2020b)	Laparoscopy	Fiducial-based	External fiducials, tracking camera
Kuhlemann et al. (2017)	Endovascular	Fiducial-based	Hololens, magnetic tracking
De Paolis and De Luca (2019)	Laparoscopy	Fiducial-based	Pre/intraop fiducials, tracking camera
Puerto-Souza et al. (2014)	Kidney	Surface-based	ICP, automatic matching and tracking
Piccinelli et al. (2020)	Bladder	Surface-based	Mono endoscope, ICP
Wang et al. (2017)	Kidney	Surface-based	Multi-view images, ICP
Thompson et al. (2018)	Laparoscopy	Surface-based	Manual reg, ICP
Robu et al. (2018)	Liver	Surface-based	Shape matching, ICP
Beckmann et al. (2023)	Throat	Surface-based	TEASER, ICP

This section introduces registration methods that *do not include deep learning*. Most of these methods have already been addressed by previous reviews, so only the most recent ones will be presented in the following subsections. Rigid registration methods are summarized in Table 3.1. Rigid registration computes the optimal rigid transformation, precisely, the rotation and translation parameters, that best aligns two point clouds while preserving their inherent shape. In the following subsections, rigid registration methods are divided into manual, fiducial-based and surface-based approaches.

Manual rigid registration

Manual registration involves aligning two or more images by manually manipulating them using a Human-Computer Interface (HCI). In manual registration, an operator, often a trained medical professional, visually identifies corresponding anatomical features in the images and adjusts the position, orientation, and scale of one image to align it with the other.

A recent work using this type of registration is Penza et al. (2023); the authors developed a system where the user could perform an initial manual registration with a 3D mouse (Fig. 3.2b). The registration is conducted between the 2D projection of the 3D model and the 2D intraoperative image, then it is updated following the camera movements, tracked through a tracking camera and markers.

The 3D mouse is a HCI that allows the control of 6 DoF intuitively, but there are other applicable solutions. For instance, Schneider et al. (2020) use interaction with a touch screen monitor or a mouse, while Nguyen et al. (2020) use a Bluetooth keyboard (Fig. 3.2a). Another approach suggested by Krug et al. (2023) involves using Ultraleap (Fig. 3.2c), a device capable of recognizing mid-air hand gestures, which are exploited to register the 3D model manually.

Discussion on manual registration: Manual registration's primary drawback lies in its time-consuming nature. The task can be challenging since it requires the operator's expertise both in the anatomy recognition and in the manipulation of a 3D model, so it may take minutes, potentially lengthening the surgical procedure. Moreover, the registration should be adjusted in case of camera movements or organ movements. This is why tracking systems and/or automatic registration updates are necessary. Finally, this type of registration is inherently rigid.

On the other hand, the advantage of this approach lies in the operator having direct control of the registration. This can potentially make the technology easier to certify for use in the operating room compared to a completely automatic method.

Fiducial-based rigid registration

This approach relies on the identification of preoperative and intraoperative fiducials. In medical imaging, fiducials refer to markers or reference points which are visible in the imaging data and serve as landmarks for registration or navigation purposes. Fiducials can be physical objects, such as markers or instruments, placed on or within the patient's body, or they can be anatomical landmarks naturally present in the anatomy, such as prominent

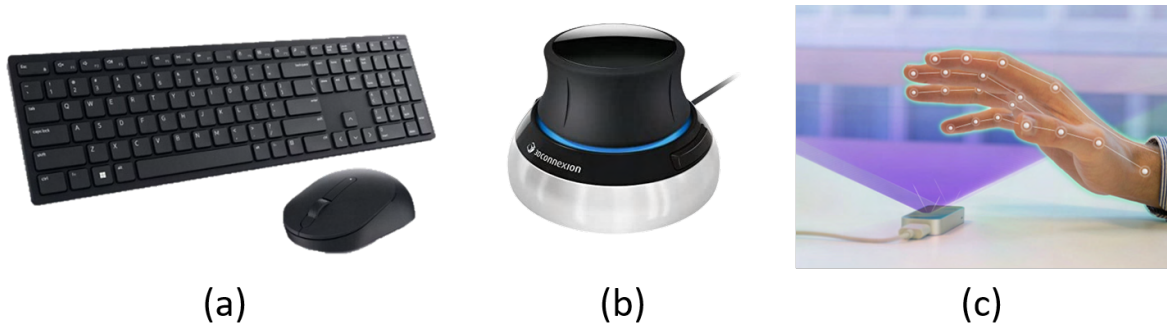


Figure 3.2 Some of the most commonly used HCI for manual registration in recent years. (a) Mouse and keyboard. (b) 3D mouse. (c) UltraLeap.

bony structures or vessel bifurcations. Fiducial-based registration methods focus on aligning the preoperative fiducial points with their corresponding intraoperative points.

A practicable fiducial-based registration method is described in Penza et al. (2020b). It exploits four fiducial points both on the virtual model and the patient. The setup includes a tracking camera, a tracked laparoscope and a pointer. Intraoperative fiducial points are chosen based on landmarks of the ribcage using a pointer tracked by the optical tracking system. Knowing the position of the relative points in the 3D model, they can compute the transformation that correctly aligns the fiducials and, consequently, the anatomy. However, the reported registration error is relatively large (5mm) because it is affected by various sources of errors. These include system calibrations and accuracy of the optical tracking system, in addition to the classic problems described in Sect. 1.2.5.

A similar work using HoloLens for endovascular interventions is described in Kuhlemann et al. (2017). In this case, they extract preoperative CT data and register it to the phantom patient using a magnetic tracking system and a magnetic pointer to touch different landmarks on the phantom's surface.

Ultimately, an alternative approach that omits the laparoscope, thus excluding real-time AR overlays on the images of the targeted organs, is De Paolis and De Luca (2019). Their proposal involves the use of 3 markers both in the preoperative and intraoperative phases. Since the correspondence between the 2 phases can be lost due to the patient's movements, a reference tool with reflective spheres is attached to the mannequin and detected by an optical tracker to stabilize registration during surgery. By exploiting the knowledge of the relative positions between these markers, they can perform rigid registration between the preoperative model and the mannequin.

Discussion on fiducial-based registration: In summary, few recent methods exploit the fiducial-based registration approach. Just like other manual methods, their popularity is

declining in favour of increasingly automated and rapid methods, which do not imply changes in surgical practices. The main disadvantage concerns the identification of fiducials on the 3D model and the area of interest, which is easily subject to errors, resulting in a deterioration of the registration. Furthermore, this approach is usually employed for rigid alignment. When the scenario permits finding effective fiducial points, the advantage is that fiducial-based methods offer an easy-to-implement solution that leads to good results. Finally, regarding time consumption, approaches like Penza et al. (2020b) complete the initial (semi-automatic) registration in less than two minutes, while the pose updates are performed in real-time based on the optical tracking.

Surface-based rigid registration

Surface-based registration involves aligning the preoperative model, often a surface mesh or point cloud (*src*), with the intraoperative data (mesh or point cloud - *tgt*) using geometric shape information. In particular, the *src* will be a *complete* point cloud, as it describes the entire organ of interest, while the *tgt* will be *partial*, representing only the visible surface. The first step consists of obtaining the intraoperative surface; various strategies are reported in the literature, such as stereovision, external 3D scanner, intraoperative imaging (CT, MR, US) and surface palpation combined with tracking. In this chapter, we will focus mainly on the case where the surface is extracted with stereovision, which is the most generic scenario and does not require adding other devices in the operating room. This technique leverages the disparity between the perspectives captured by the two eyes of the 3D camera to triangulate the position of each point on the surface of the organ within the surgical field (Stoyanov et al., 2010). Subsequently, the obtained point cloud is registered with the preoperative one.

One of the first approaches to surface-based 3D registration using stereovision was introduced in Su et al. (2009). It starts by performing an initial manual registration to fit the preoperative model with the intraoperative image visually. Then, they select several points on the kidney surface as fixed reference points and find the relative points on the 3D model. Given the two sets of relative points, they exploit the Iterative Closest Point (ICP) algorithm to perform the automatic surface registration.

Following this, Puerto-Souza et al. (2014) expand upon the concept of manual registration complemented by tracking and ICP. They add a matching algorithm to guarantee continuous tracking even in the case of occlusions, deformations or others. These two methods bridge the gap between fiducial-based and surface-based techniques by utilizing a series of anchor points extracted from the kidney's surface through triangulation.

Other studies have shifted their focus towards alternative approaches of reconstructing point clouds. An example that exploits a monocular endoscope is Piccinelli et al. (2020), which proposes a sparse 3D reconstruction obtained by ORB-SLAM (Mur-Artal et al., 2015). The obtained surface is registered to the preoperative model with ICP. Another scenario is Wang et al. (2017), in which the authors exploit a 3D endoscope capable of capturing five images from different viewpoints at the same time. By exploiting the DTAM algorithm (Newcombe et al., 2011), they reconstruct the intraoperative cloud, which will be registered with ICP to the 3D model.

Some other algorithms conduct two sequential registrations; one coarse, and the other fine. Recently, an approach of this type was tested in vivo on the liver (Thompson et al., 2018), obtaining a Re-Projection Error (RPE) of approximately 12mm. First, a dense surface reconstruction of the intraoperative surface is conducted. Next, manual coarse registration is executed and refined using surgical instrument tracking and the ICP algorithm.

Likewise, Robu et al. (2018) start with a dense surface reconstruction, followed by an initial global registration and subsequent application of ICP. The initial registration is conceptualized as a quadratic assignment problem, wherein the energy function aims to identify a set of correspondences that minimize dissimilarity between descriptors while preserving geodesic distances between pairs of correspondences.

Similarly, Beckmann et al. (2023) propose an initial coarse registration with TEASER (Truncated least squares Estimation And SEmidefinite Relaxation) method, followed by a fine registration with ICP. Belonging to the same category, but using an RGB-D camera to acquire the intraoperative surface is Huang et al. (2021b). Their solution employs the Sample Consensus Initial Alignment (SAC-IA) algorithm (Rusu et al., 2008) as a coarse registration and then ICP for fine registration.

Discussion on surface-based registration: To conclude, surface-based registration involves utilizing computer vision algorithms that can reconstruct the intraoperative 3D surface, forming the basis for the registration process. Subsequently, the registration can be executed using ICP in conjunction with other techniques like tracking and matching algorithms. Simultaneously, it is crucial to highlight the drawbacks and obstacles associated with this approach. The first drawback of the surface-based approach derives from the inaccurate or incomplete reconstruction of surfaces, primarily caused by the absence of distinctive features. This challenge is worsened by complex, texture-less, and deformable scenes, often interrupted by unpredictable instrument occlusions. While literature offers solutions to these challenges, the second drawback is that only a portion of the organ of interest is typically visible. The smaller this portion, the fewer features are available for matching, increasing ambiguity in

the registration process. This challenge varies depending on the organ under consideration. For instance, in a study (Benincasa et al., 2008) focusing on the kidney, the author indicated that a minimum of 28% of the visible surface is necessary for ensuring accurate registration. This requirement may only sometimes be met during real surgeries. Regarding the time consumption, determining the exact registration time is challenging since it depends on the number of points and the number of ICP iterations required; for example, Robu et al. (2018) achieve initial global alignment in approximately 20 seconds. Despite all this, these approaches emerge as the most convenient technique for surgical guidance, as it can be set up with minimal surgeon intervention and eliminates the need for selecting artificial fiducial markers. Its advantages are that it is automatic and allows the registration to be updated during the procedure without using markers. This is why surface-based registration is still a very active research topic.

3.1.2 Non-rigid registration methods

So far, all the methods discussed tackle the challenge of rigid registration, which still requires a comprehensive solution. As previously noted, rigid registration demonstrates accuracy in scenarios with minimal deformations, making it particularly effective in orthopaedic and neurosurgical applications. On the contrary, non-rigid registration is necessary to ensure greater accuracy in scenarios characterized by numerous soft tissues and resultant deformations, such as laparoscopic procedures. During the literature review, it was observed that the methods can be categorized into two categories: surface-based techniques and other deformation models. All the described methods are summarized in Table 3.2.

Surface-based non-rigid registration

Similar to the methods described in Sect. 3.1.1, surface-based non-rigid registration methods are a class of algorithms that perform registration exclusively using the spatial information contained in point sets. These methods assume that their discrete points can capture the essential geometry of the objects and that aligning these points accurately reflects the alignment of the underlying shapes. This category can be further divided into two classes: CPD-based and FEM-based methods.

Coherent Point Drift (CPD)-based CPD (Myronenko and Song, 2010) is a probabilistic approach for both rigid and non-rigid point set registration. This method treats the alignment of two point cloud as a probability density estimation problem. It fits the centroids of a

Table 3.2 Overview of traditional non-rigid registration methods

Reference	ROI	Approach	Peculiarity
Zhang et al. (2020)	Kidney	Surface-based, CPD	ICP, Gaussian RBF
Dhoju et al. (2021)	Liver	Surface-based, CPD	ICP
Reichard et al. (2017)	Liver	Surface-based, FEM	Mosaicing, SAC-IA
Teatini et al. (2020)	Liver	Surface-based, FEM	ICP
Yang et al. (2024)	Liver	Surface-based, FEM	Boundary-condition-free FEM
Espinel et al. (2022)	Liver	Surface-based, FEM	Monocular camera, ICP
Golse et al. (2021)	Liver	Surface-based, FEM	RGB-D camera, manual reg
Paulus et al. (2017)	Kidney	Surface-based, FEM	Robust to cuts
Adagolodjo et al. (2018)	Liver	Fiducial-based, FEM	Invasive marker, ICP
Collins et al. (2020)	Uterus	Others	Optimization, Semi-auto, Tracking
Collins et al. (2017)	Liver	Others	ICP, Elastic Model
MORELLI (2016)	Kidney	Others	Fiducial-based, B-Spline

Gaussian Mixture Model (GMM), representing the first point set, to the data of the second point set by maximizing the likelihood.

Zhang et al. (2020) propose a method based on CPD and optimized volume deformation. After reconstructing the surface intraoperatively, they perform a rigid registration with ICP followed by CPD. The latter, however, suffers in the case of deformations, which negatively impact its prediction. To solve this problem, the authors present an optimized volume deformation method using Gaussian RBF interpolation.

Dhoju et al. (2021) suggest a similar method. The reconstructed surface undergoes registration with the preoperative model using ICP. Subsequently, an improved CPD algorithm is employed. The method uses a hybrid optimization scheme to determine the weight parameters automatically. Specifically, modifying the parameter computation enhances the performance of non-rigid registration.

Finite Element Method (FEM)-based Reichard et al. (2017) suggest a FEM-based approach applicable to stereo-laparoscopy. The authors start by segmenting the organ of interest with random forests and superpixels, then reconstructing the intraoperative surface using disparity. The reconstruction from a single frame does not contain enough information to perform the registration, so they utilize mosaicing to gather a more descriptive point cloud. At this point, to establish the correspondences between the biomechanical model and the reconstructed surface, they perform an initial registration using SAC-IA and Physics-Based

Shape Matching (PBSM) (Suwelack et al., 2014). Finally, a dynamic registration is executed at every frame to update the biomechanical model, using the displacement between the model and the surface.

Another approach for stereo-laparoscopy, involving the reconstruction of the intraoperative surface through disparity, is presented in Teatini et al. (2020). In this case, the authors adopt Prevost et al. (2020) to conduct the initial registration, exploiting the combination of a fiducial-based method and ICP. Once the correspondences are established, the displacement between the reconstructed surface and the 3D model is applied to the FEM to calculate deformations.

Similarly, and more recently, Yang et al. (2024) use an initial rigid registration followed by surface matching with FEM. They formulate the FEM registration as the optimization of surface forces applied to a volumetric liver model. By stabilizing the stiffness matrix (“soft springs”), their approach removes the need for explicit zero-boundary conditions and iteratively refines closest-point correspondences during deformation.

A monocular camera method is presented in Espinel et al. (2022). Initially, the authors integrate several frames captured by the laparoscope to obtain a more descriptive intraoperative point cloud. Then, they combine visual cues, FEM, and ICP to register the preoperative 3D model with a laparoscopic view (Koo et al., 2017). More precisely, the method employs an elastic biomechanical model as the deformation law for the liver’s preoperative volume, while utilizing shading and contour as visual cues from the single laparoscopy image to identify deformations.

In contrast, Golse et al. (2021) explores liver resection within the context of open surgery, using an RGB-D camera. An initial manual registration is performed to find the correspondences between the preoperative 3D model and the reconstructed point cloud. Then, they exploit surface displacements to compute the deformations of the 3D model through FEM.

A particular example that exploits the use of markers is Adagolodjo et al. (2018). At the beginning of surgery, the surgeon places the optical markers around the area of interest. After this, two phases follow: an initial rigid registration and a tracking phase. The initial registration is performed with ICP, allowing for the alignment of the FE preoperative model with respect to the intraoperative configuration. In this way, it is possible to identify the position of the markers on the 3D model. Subsequently, the FE preoperative model is deformed based on the 3D displacements of the markers, tracked by an infrared camera. The authors also showed its applicability in the operating room during liver resection surgery performed on a human patient, obtaining promising results.

Finally, Paulus et al. (2017) stands out as a notable contribution, being among the few methods that tackle the challenge of non-rigid registration post intraoperative cuts. The authors use a monocular camera to capture organ images, extracting significant features using a Sped-up Feature Detector (Bay et al., 2008) and tracking them with Lucas-Kanade optical flow (Bruhn et al., 2005). Following an initial registration, these features are projected onto the 3D elastic model. After the organ's incision, the cut line is transferred to the model, enabling the deformation estimation through a node snapping technique and remeshing approach (Paulus et al., 2015).

Other deformation models for non-rigid registration

These include different deformation models or optimization methods that are also applicable to the laparoscopic scenario.

Collins et al. (2020) propose a semi-automatic AR guidance method for uterine laparoscopy that involves two stages. Initially, the operator roughly aligns the preoperative 3D model with a keyframe by selecting corresponding points. Next, an exploratory video is used to generate a dense 3D reconstruction of the organ's surface using SfM. Finally, numerical optimization performs non-rigid registration by minimizing an energy function that integrates the reconstructed surface and the operator-indicated organ contour. Once the initial registration is complete, real-time tracking is performed by matching SIFT keypoints between the laparoscopic image and the preoperative model, with the pose further refined using the Levenberg-Marquardt algorithm. The approach assumes minimal organ deformation during the tracking, allowing the transformation to be modelled as rigid.

Another example is the method presented by Collins et al. (2017). The authors perform a first rigid registration exploiting salient feature-weighted ICP. This allows them to establish correspondences between the 3D model and the intraoperative surface, refining the non-rigid registration using a linear elastic biomechanical model (Rucker et al., 2013). In particular, the boundary conditions are derived from signed closest point distances between the data and the model surface.

Finally, MORELLI (2016) exploits an initial rigid registration employing a fiducial-based method with markers, followed by a deformable registration using non-rigid FFD (Free Form Deformation) with B-spline (Khodadad et al., 2011).

Discussion on non-rigid registration In summary, most efforts in recent years have been directed towards FEM approaches. The general idea is to conduct an initial registration between the preoperative model and the intraoperative surface. Following this step, the

most common approach is to extract and track interesting features from the intraoperative surface, which will then be projected onto the biomechanical model to serve as input for the FEM. Non-rigid registration poses even greater challenges than rigid registration, as it not only inherits the existing limitations but also introduces new ones. Indeed, similar to rigid registration, its accuracy heavily relies on the quality of intraoperative reconstruction, which, it is worth noting, only represents a fraction of the involved organ, compounded by the challenge of instrument occlusions.

Concerning the newly introduced drawbacks, the first involves ensuring accurate deformable registration to enable the correct projection of features onto the 3D model. FEM can be time-consuming for complex and high-resolution simulations, but modern hardware, optimized solvers, and parallel processing can help mitigate these costs. For instance, Reichard et al. (2017) performed the initial registration in 1-2s and the dynamic one in 100ms, allowing 10 FPS updates. Simultaneously, it is crucial to identify biomechanical parameters that accurately characterize the organ and are mutually compatible. Otherwise, the FEM may fall into numerical instabilities. Another problem concerns validation, as surgical environments often lack ground truth data for evaluating registration performance, making it challenging to assess the reliability of FEM-based techniques in practice. Lastly, non-rigid registration should address topological changes like resection, cuts, and anastomosis, which significantly alter the organ's structure. These are subjects that only a few recent studies have explored (Paulus et al., 2017).

Although non-rigid methods are the most difficult to implement, they have the advantage of being the only ones to consider soft tissue deformations. Therefore, they provide the most comprehensive solution to the problem of laparoscopic AR registration, as they retain the benefits of surface-based methods, including automaticity and the absence of markers.

Summary of traditional registration methods

Fig. 3.3 provides an overview of traditional registration methods. Both surface-based and non-rigid approaches exploit two input point clouds: the preoperative surface extracted from a 3D model and the intraoperative surface reconstructed via stereovision. In non-rigid pipelines, any of the rigid techniques described previously can serve as the initial rigid registration step before applying deformation-aware logic.

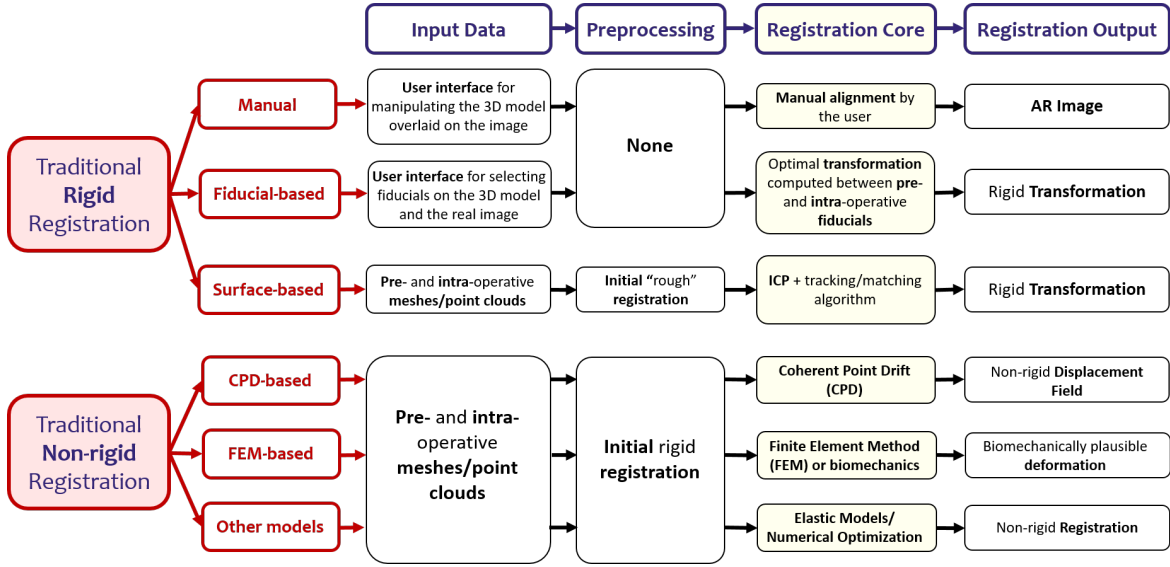


Figure 3.3 Schematic overview of traditional registration methods. Each horizontal lane corresponds to a method category, while the vertical columns represent the common pipeline stages: input data, preprocessing, registration core, and registration output.

3.2 Deep Learning methods

In recent years, the rise of Deep Learning (DL) has been overwhelming, revolutionizing the landscape of computer vision and enabling significant performance improvement in many tasks, including feature extraction, segmentation, image classification and image reconstruction. One of the main reasons for this success lies in representation learning, which consists of autonomously discovering data features and patterns, surpassing the performance of handcrafted features used in traditional approaches. Additionally, the availability of large-scale annotated datasets, coupled with advancements in high-performance computing hardware and algorithmic innovations, has allowed researchers to train increasingly complex models on massive datasets, unlocking unprecedented levels of performance and scalability. This popularity has also affected the field of registration, as highlighted in Fig. 3.4. In this context, DL methods were initially used to extract features to guide traditional algorithms (Sect. 3.1). Subsequently, the transformation was predicted directly, initially with supervised methods and later with unsupervised and weakly-supervised approaches.

As discussed in Section 1.2.2, this thesis specifically explores surgical AR registration methods rather than medical image registration, which has been thoroughly investigated in other reviews, such as Fu et al. (2020); Haskins et al. (2020). In particular, we will explore how DL can be incorporated into specific stages of the surface-based approaches, such as the segmentation or reconstruction of the intraoperative cloud. Then, we will delve deeper into

methods that allow direct transformation estimation. These methods focus on point cloud registration, where the network takes two point clouds as inputs: one extracted from the preoperative 3D model, denoted as *src*, and the other from the intraoperative surface, referred to as *tgt*.

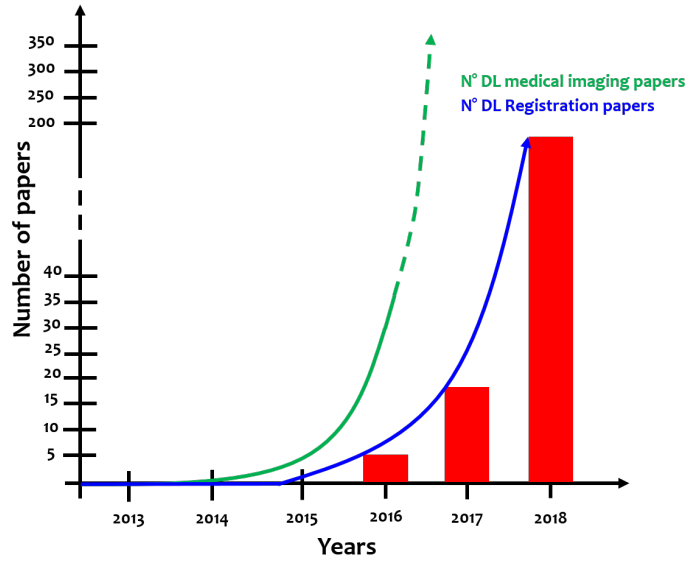


Figure 3.4 Popularity of deep learning in medical imaging papers and registration papers. As observed in Haskins et al. (2020), registration papers emerged slightly later than medical imaging papers and are now following a similar trend.

3.2.1 Hybrid DL-enhanced methods

Several studies have investigated the application of DL across different stages of the traditional surface-based methods to enhance their efficacy and the resulting registration quality. These methods are described in Table 3.3.

For instance, DL can be employed in the initial segmentation phase to identify the organ of interest on 2D endoscopic images. Various models have been proposed to address this challenge, as outlined in Den Boer et al. (2022); Rivas-Blanco et al. (2021).

An example of a framework integrating one of these methods can be found in Zhang et al. (2022, 2019). Initially, they segment the kidney with Mask R-CNN. The network is supervised and trained on manually segmented renal regions in clinical laparoscopic images. Then, to obtain the intraoperative surface, stereovision combined with point-cloud stitching is used; in this way, different point-cloud patches captured by different views of the endoscope are "stitched" together. Ultimately, two registrations are conducted. An initial registration,

Table 3.3 Overview of Hybrid DL-Enhanced Methods

Reference	ROI	Approach	Peculiarity
Zhang et al. (2019)	Kidney	Surface-based	Mask-RCNN, ICP, CPD
Luo et al. (2020)	Liver	Surface-based	DL-based reconstruction, unsupervised ICP
Labrunie et al. (2022)	Liver	Surface-based	U-Net features, RANSAC
Tanzi et al. (2021)	Prostate	Surface-based	U-Net catheter segmentation (extended in Padovan et al. (2022))
Chandelon et al. (2025)	Kidney	Surface + Image-based	Image encoder, iterative error feedback pose regressor
Pfeiffer et al. (2019)	Liver	Surface + FEM	Nonrigid CNN displacement field (extended in Pfeiffer et al. (2020))
Brunet et al. (2019)	Liver	Surface + FEM	Nonrigid U-Net volumetric deformation
Wang et al. (2024)	Liver	Surface + FEM	Initial FEM + learned residual correction
Gadoux and Bartoli (2025)	Liver	Landmark + FEM	Landmark segm, learned latent deformation model
Mhiri et al. (2025)	Liver	Landmark	Linear deformation subspace, MLP

coarse in nature, aligns the stitched point cloud with the preoperative model using ICP. Subsequently, a fine registration is executed to achieve deformation alignment between the current model and the preoperative one, utilizing the CPD algorithm.

Another application of DL is surface reconstruction. Compared to traditional disparity calculation methods, which frequently produce sparse or semi-dense surface reconstructions, DL approaches offer promising results for dense reconstruction (Han et al., 2019). An interesting example is Luo et al. (2020). To extract the intraoperative point cloud of the liver, the authors rely on an unsupervised CNN capable of estimating the depth starting from the stereoscopic images. The movements of the laparoscope are tracked through a tracker and a tracking camera. Subsequently, the Go-ICP (Yang et al., 2015) is used to carry out the registration between the pre- and intraoperative surfaces.

Another paper that exclusively focuses on reconstruction, rather than registration, is Wei et al. (2022). Their model accepts stereo images as input and applies a Hierarchical Deep Stereo Matching (HSM) network (Yang et al., 2019) to estimate the depth. The architecture is supervised, initially trained on KITTI and then fine-tuned on SERV-CT and SCARED datasets. It identifies features with large receptive fields and multi-scale properties to facilitate the establishment of precise pixel-level correspondence between left and right binocular

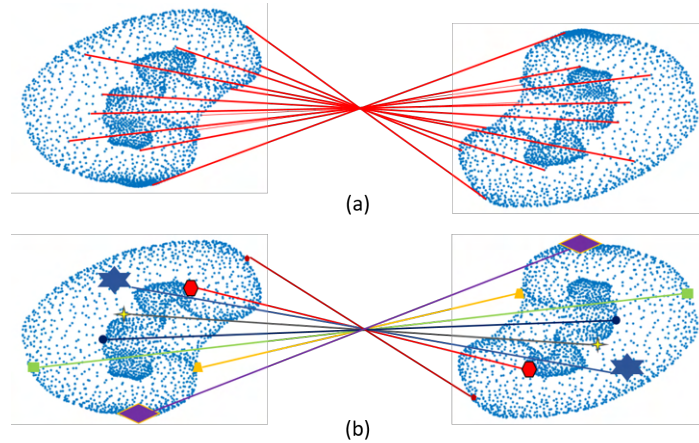


Figure 3.5 Types of matching used in registration approaches. (a) Point matching focuses only on the point coordinates (ICP). (b) Feature matching, in addition to the coordinates, uses other information typically extracted using deep learning methods. Features could be local or global, depending on the scale of the information they capture.

images. The obtained depth map is provided to a dense visual reconstruction algorithm that recovers the entire 3D structure of the tissue surface. The entire framework can achieve real-time performance.

Additionally, other techniques use deep learning for feature extraction to aid registration, as demonstrated in Fig. 3.5. Traditional methods establish correspondences by using only the coordinates of each point (Fig. 3.5a). In contrast, deep learning methods perform feature-based matching (Fig. 3.5b), enabling more accurate matches. For instance, Labrunie et al. (2022) employ a U-Net network to detect critical landmarks on the liver's surface. In this case, the selected landmarks are the lower anterior ridge of the liver and the junction between the falciform ligament and the liver. To train the network, annotated endoscope images provided by surgeons were used to create the dataset. During surgery, once the U-net identifies the landmarks, point-to-point correspondences (3D-2D) with the 3D model are established, and registration is achieved by iteratively applying RANSAC to find the optimal estimate.

Another particular example of a DL-enhanced method is Tanzi et al. (2021). The method is tailored for robot-assisted radical prostatectomies, specifically focusing on the phase where a catheter is inserted into the pelvic cavity. The authors exploit the catheter segmentation information as an artificial, easily identifiable feature to guide the virtual-over-real overlay. They propose employing a U-net for catheter segmentation in endoscopic images, followed by shape detection to extract the points needed to correctly align the 3D model (containing the catheter) with the real-time video. This work was then extended and generalized by Padovan et al. (2022), in which the authors exploit the segmentation outputs as input to a

CNN that estimates the rotations necessary to perform the registration. Furthermore, they introduce an organ tracking system based on optical flow. The new architecture is more generic, as it can conduct registration without relying on the catheter.

Chadelon et al. (2025) propose a landmark-free kidney AR registration pipeline that regresses the 3D-2D registration directly from a monocular RGB endoscopic image, avoiding explicit anatomical landmark matching. The method introduces a generic kidney “shape prototype”: preoperatively, the patient-specific digital twin is aligned to this prototype via a coarse-to-fine procedure (canonical-frame initialisation followed by ICP and NR-ICP), and intraoperatively, a neural network predicts the prototype pose in the camera frame, which is then transferred back to the patient twin. The network leverages a pretrained image encoder (DINOv2) and an iterative error feedback regressor to refine rotation/translation parameters. Reported results show real-time inference (9 ms) and sub-centimetric tumour-contour alignment error on clinical cases.

Two interesting proposals that combines DL with FEM are Brunet et al. (2019); Pfeiffer et al. (2019). They propose the use of a CNN to predict a displacement field for all points within an organ based solely on the displacement of a portion of its surface. The authors create synthetic datasets by simulating FEM models of random, organ-like meshes. This approach provides pairs of original and deformed models along with ground truth deformation information. With this data, they train their network, which is a U-Net, allowing for three-dimensional input. Subsequently, their framework was further developed in Pfeiffer et al. (2020).

A different approach is LIBR+ (Wang et al., 2024), a hybrid intraoperative liver registration method that learns to correct a fast linear-elastic FE registration (LIBR Heiselman et al. (2020)). Given a preoperative liver mesh and a set of intraoperative sparse measurements (e.g., sparse surface points), the pipeline first runs the biomechanical solver LIBR to obtain an initial deformed mesh. Then, a neural network is trained to take as input the preoperative mesh, the LIBR output, and the sparse measurements, and to predict a dense residual displacement field over the mesh vertices. The final deformation is obtained by refining LIBR’s result with the learned correction (i.e., applying the predicted residual on top of LIBR output).

Gadoux and Bartoli (2025) propose ADeLiR, an automatic deformable 3D–2D liver registration pipeline for laparoscopic AR based on anatomical landmarks. The method proceeds in three stages: (1) it segments the relevant 2D landmarks (i.e. liver ridge and silhouette) in the intraoperative image using a neural model; (2) it performs a coarse global alignment by regressing a 3D affine transformation with a network trained preoperatively on

synthetic landmark projections, and reduces the synthetic-to-real gap with domain adaptation; (3) it refines the result with a non-rigid step. In particular, during training, they learn a latent deformation model with an autoencoder trained on FEM-simulated liver deformations. Then, during testing, they optimise the latent code to minimise the Hausdorff distance between projected deformed-model landmarks and the segmented 2D landmarks.

Mhiri et al. (2025) propose a patient-specific learning-based 3D–2D deformable registration for laparoscopic liver AR. Instead of regressing all mesh vertices, they build a linear deformation subspace and represent a deformed mesh by coefficients. Then, a lightweight MLP takes a fixed-length set of 2D anatomical landmark points (ridge/silhouette) and regresses low-dimensional deformation coefficients. The MLP is trained preoperatively on per-patient synthetic deformations generated with an ARAP-based (Igarashi et al., 2005) simulator and camera projection.

Discussion on hybrid DL-enhanced methods: Fig. 3.6 provides an overview of hybrid DL-enhanced methods. DL segmentation and reconstruction improve the quality of the intraoperative point cloud extracted from the video, which is then aligned using traditional surface-based registration. More recently, several works have also proposed segmenting liver anatomical landmarks (e.g., ridges and silhouettes) and using these cues to guide non-rigid registration. Overall, the hybrid DL methods offer a versatile solution to the registration task across various scenarios, typically enhancing accuracy and reducing execution times. Segmentation has long reached high accuracy and processing speed, and now reconstruction methods are trending in the same direction. Similarly, integrating deep learning with FEM aligns with these trends, notably cutting estimation times. Their distinguishing factor lies in their capacity to learn complex and discriminative data representations, allowing for more accurate registration even with noise in the source data. However, one drawback is the necessity to construct appropriate datasets to train these networks. Sometimes the required data may be unavailable or lacking ground truth. For this reason, it is often necessary to resort to synthetic data, with the risk that they may not be adequately representative and generalizable to address real-world scenarios. The advantages are that these methods are mostly surface-based, thus automatic and marker-free, offering superior performance compared to traditional methods. For instance, methods like Luo et al. (2020) leverage DL to reconstruct higher-quality intraoperative surfaces while still achieving registration update rates of 10–12 FPS.

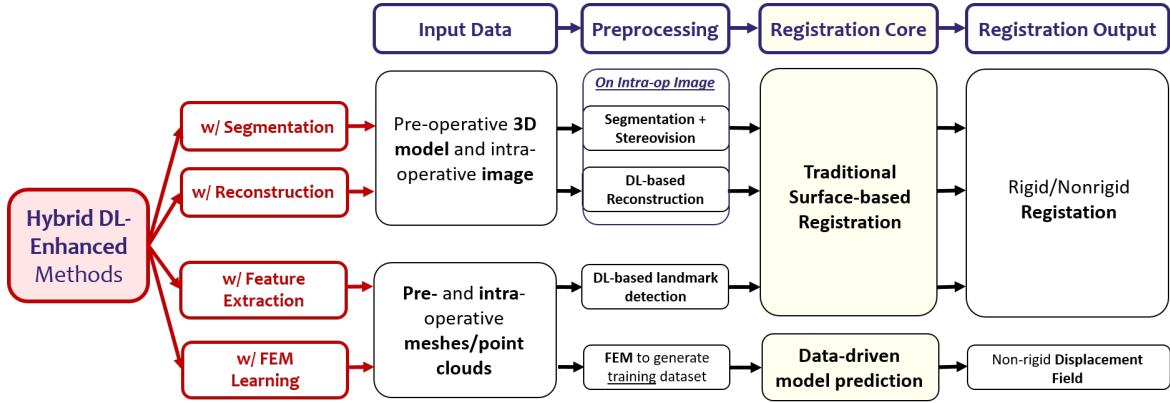


Figure 3.6 Schematic overview of Hybrid DL-Enhanced methods. Each horizontal lane corresponds to a method category, while the vertical columns represent the common pipeline stages: input data, preprocessing, registration core, and registration output.

3.2.2 DL-based point cloud registration methods

Similar to surface-based methods, the inputs for point cloud registration methods are always *src*: a *complete* point cloud derived from the preoperative 3D model, and *tgt*: a *partial* point cloud obtained from the intraoperative image. For this reason, the problem to be solved is a complete-to-partial registration. Given the scarcity of algorithms targeting this issue, we will also explore approaches designed for partial-to-partial registration, which typically exhibit a generalizability level extending to the former scenario. Furthermore, this section broadens the scope of the review from the specific field of laparoscopy to a more general perspective, due to the lack of targeted studies on surgery.

In the DL context, end-to-end networks take these two point clouds as input and predict the transformation that aligns them as the output. As reported by Huang et al. (2021c), the idea is to convert the registration problem into a regression problem. It is crucial to emphasize that these architectures differ from those employed for segmentation and detection primarily because of the fundamental difference in the input data: RGB images in one scenario and point clouds in the other. Point clouds are defined as a set of 3D points without a specific order, making them less compatible with conventional CNNs that excel with images featuring regular grids, as noted by Monji-Azad et al. (2023). Hence, exploring alternative methods to process point cloud data and achieve effective feature extraction became necessary. One approach involves converting point clouds into volumetric 3D grids and then applying CNNs, as demonstrated in Pfeiffer et al. (2020).

Differently, PointNet (Qi et al., 2017) is a pioneering method that processes raw point clouds directly by addressing key challenges such as unordered input, local feature extraction from neighbouring points, and transformation invariance.

Another option for point cloud processing is to use DGCNN (Dynamic Graph Convolutional Neural Network) (Wang et al., 2019), which operates on graph-structured data, allowing each node to have variable connections with other nodes, unlike traditional CNNs. They excel at learning geometric features derived from point clouds with neighbouring relations, typically defined on KNN (K-Nearest Neighbors) graphs.

Finally, methods incorporating transformers (Guo et al., 2021) are increasingly popular, leveraging the inherent permutation invariance of the transformer and the attention mechanism to facilitate feature learning on point cloud data.

We divided the examined registration-related methods into four main categories and summarized them in Table 3.4.

Correspondence-based methods

These methods focus on establishing one-to-one correspondences between the *src* and the *tgt* feature embeddings and subsequently estimate the transformation with an optimization method. Their core aspect involves computing a correspondence matrix using the similarity of the extracted features. Most of the methods listed in Table 3.4 are trained in a supervised manner using ModelNet40 (Wu et al., 2015). This dataset contains normalized 3D CAD models across 40 categories: chairs, cars, aeroplanes, and other common objects.

Deep Closest Point (DCP) (Wang and Solomon, 2019a) is one of the first complete-to-complete registration (Fig. 3.7a) methods that extract features from each point cloud using DGCNN and refine them through self- and cross-attention layers to establish soft correspondences via a softmax function. The rigid transformation is then computed from these correspondences using Singular Value Decomposition (SVD).

Regarding partial-to-partial registration (Fig. 3.7b), PRNet (Wang and Solomon, 2019b) is one of the pioneering approaches. It employs a feature extraction process similar to DCP and then identifies keypoints in both clouds. Keypoints-to-keypoints correspondences are used to estimate the transformation, which is subsequently applied to *src*.

The concept of leveraging local features, concentrating on neighbouring points, has frequently been employed to address the registration challenge. In recent years, significant efforts have been dedicated to refining this approach to minimize the number of wrong matches. In the literature, two refining classes can be identified.

Table 3.4 Overview of DL-based Point Cloud Registration. Methods in **bold** are applied to the medical field.

Reference	Approach	Peculiarity
DCP (Wang and Solomon, 2019a)	Corr-based	Complete-to-complete, self/cross-att
PRNet (Wang and Solomon, 2019b)	Corr-based	Keypoints, self/cross-att
PREDATOR (Huang et al., 2021a)	Corr-based, Overlap refined	Overlap score
LiverMatch (Yang et al., 2023)	Corr-based, Overlap refined	Liver , Visibility score, transformer, matching
RegTR (Yew and Lee, 2022)	Corr-based, Overlap refined	Overlap score, transformer
UTOPIC (Chen et al., 2022)	Corr-based, Overlap refined	Overlap uncertainty
INENet (Wu et al., 2022)	Corr-based, Overlap refined	Complete-to-partial, threshold prediction
STORM (Wang et al., 2022b)	Corr-based, Overlap refined	Gumbel-Softmax
MRA (Chen et al., 2023a)	Corr-based, Overlap refined	Masking, reconstruction
RGM (Fu et al., 2021a)	Corr-based Geom refined	Deep Graph Matching
RPMNet (Yew and Lee, 2020)	Corr-based, Geom refined	Robust point matching
RIENet (Shen et al., 2022)	Corr-based, Geom refined	Inlier evaluation
DIT (Chen et al., 2023b)	Corr-based, Geom refined	Full transformer, confidence score
GeoTransformer (Qin et al., 2022)	Corr-based, Geom refined	Geom transformer
KCR-Net (Zhang et al., 2024)	Corr-based, Geom refined	Liver , KNN Feature Fusion
(Dai et al., 2025)	Corr-based, Hybrid	Liver , GeoTransformer, overlap
(Huang et al., 2025)	Corr-based, Hybrid	Liver , GeoTransformer, overlap
RORNet (Wu et al., 2023b)	Hybrid	Overlap, corr-matrix, dual-branch
PointNetLK (Aoki et al., 2019)	Corr-free	Not end-to-end, LK alg
OMNet (Xu et al., 2021)	Corr-free	Overlap masks
(Guan et al., 2023)	Corr-free	Liver , OMNet improved with RPMNet
FINet (Xu et al., 2022b)	Corr-free	Dual-branch
PANet (Wu et al., 2023a)	Corr-free	Multi-branch, different features scales
DeepGMR (Yuan et al., 2020)	GMM-based	Complete-to-complete, 2 GMMs
UGMM (Huang et al., 2022)	GMM-based	Complete-to-complete, 1 GMM
OGMM (Mei et al., 2023a)	GMM-based	2 GMMs, overlap score
UDPReg (Mei et al., 2023b)	GMM-based	2 GMMs, overlap score
ProRegNet (Fu et al., 2021b)	Non-Rigid, Corr-based	Dataset with FEM, medical MR-TRUS reg
CorrNet3D (Zeng et al., 2021)	Non-Rigid, Corr-based	Complete-to-complete, MLP, extended in (Hu et al., 2022)
Lepard (Li and Harada, 2022a)	Non-Rigid, Corr-based	Matching confidence, N-ICP
NDP (Li and Harada, 2022b)	Non-Rigid	Hierarchical structure, MLP
GraphSCNet (Qin et al., 2023b)	Non-Rigid	Local spatial consist., N-ICP
Self-P2IR (Zhou et al., 2025)	Non-Rigid	Liver , rigid + non-rigid, self-supervised

Refined with overlap points PREDATOR (Huang et al., 2021a) is a pioneering model for partial-to-partial registration that identifies overlapping areas between two point clouds by exchanging information through cross-attention and assigning each point an overlap score. This overlap information, combined with soft correspondences, significantly enhances the quality of the matched points.

LiverMatch (Yang et al., 2023) is inspired by PREDATOR and represents one of the first approaches for complete-to-partial liver registration under deformation conditions. The architecture utilizes an encoder-decoder structure with self- and cross-attention mechanisms to extract point features, which are used to compute a similarity matrix. This matrix, combined with a visibility score, is later employed to predict correspondences between the two point clouds. Once these correspondences are established, registration is performed using ICP. To train the network for matching under deformation, the authors created a synthetic dataset by applying deformations and crops to 16 livers from the 3D-IRCADb-01 dataset (Soler et al., 2010).

RegTR (Yew and Lee, 2022) also estimates an overlap score of each keypoint, which is used to predict correspondences. In particular, unlike the more common approach of computing correspondences via nearest neighbour feature matching, this network is trained to estimate matching points through transformer layers and directly predict the final transformation.

MRA (Chen et al., 2023a) rethinks the point cloud registration problem as a masking and reconstruction task. In particular, they use the invisible parts of each point cloud as inherent masks; consequently, reconstructing a complete cloud using the mask of the other is equivalent to registration.

Other popular approaches in this category are UTOPIC (Chen et al., 2022), INENet (Wu et al., 2022) and STORM (Wang et al., 2022b).

Refined with geometric constraints RPMNet (Yew and Lee, 2020) estimates the parameters for the Robust Point Matching algorithm (Gold et al., 1998) to identify outliers and improve the correspondence matrix.

GeoTransformer (Qin et al., 2022) proposes a new geometric transformer module that learns a geometric invariant representation of point clouds for robust superpoint matching. Identifying geometric structures within individual point clouds and ensuring geometric consistency across point clouds enables accurate estimation of correspondences.

DIT (Chen et al., 2023b) employs a predominantly transformer-based architecture to extract salient features and establish correspondences. Subsequently, each correspondence is

assigned a confidence score based on geometric constraints to selectively utilize the most reliable matches for estimating the transformation. Other popular approaches in this category are RGM (Fu et al., 2021a) and RIENet (Shen et al., 2022).

Recently, KCR-Net (Zhang et al., 2024) introduced a keypoint matching network extendable to liver registration. Their network leverages a Neighborhood Feature Fusion Module to extract and merge each point's features with those of its neighbours, creating representations that capture local geometric structures. This integration is achieved by combining KNN-max pooling with self- and cross-attention mechanisms. Then, the extracted features are used to compute correspondence weights, from which a set of correspondences is predicted. The final registration is performed using SVD. To adapt this model for liver registration, following an approach similar to LiverMatch, they generated a synthetic training dataset by applying deformations and cropping techniques to the livers from the 3D-IRCADb-01 dataset.

An hybrid approach that leverages both overlap and geometric features for liver registration is introduced in Dai et al. (2025). Initially, a geometric transformer module, similar to GeoTransformer, refines the features. Next, a deep graph matching module computes graph edges via an Edge Generator, which are subsequently used to derive a similarity matrix. Lastly, the most representative matches are filtered using an overlap mask, and the final transformation matrix is computed with RANSAC and ICP. Also in this case, the training dataset is generated following the strategy proposed by LiverMatch.

Huang et al. (2025) extend this work by proposing an entire pipeline for landmark-free rigid liver 3D–2D registration. From a single laparoscopic RGB frame, they reconstruct an intraoperative liver surface point cloud by combining monocular depth estimation with a SAM2-based liver mask. A two-stage matching pipeline then aligns this partial cloud to the preoperative liver model. KPConv-FPN features, and a geometric transformer establish coarse superpoint correspondences, which are propagated to dense matches via sparse-to-dense transport. The final transformation is recovered by weighted SVD over matched pairs and selected based on inlier support.

Discussion on correspondence-based methods: Overall, correspondence -based methods stand out as the most popular within the domain of DL-based point cloud registration, underscored by the extensive literature available. This technique leverages the estimation of point correspondences by calculating the similarity matrix between point features. Recent advancements in these methods have surpassed the performance of earlier approaches, thanks to refined techniques that ensure the extraction of more informative features and, thus, higher levels of accuracy. Collectively, this progress has paved the way for some of the first point cloud registration methods in laparoscopy, exemplified by LiverMatch and KCR-Net.

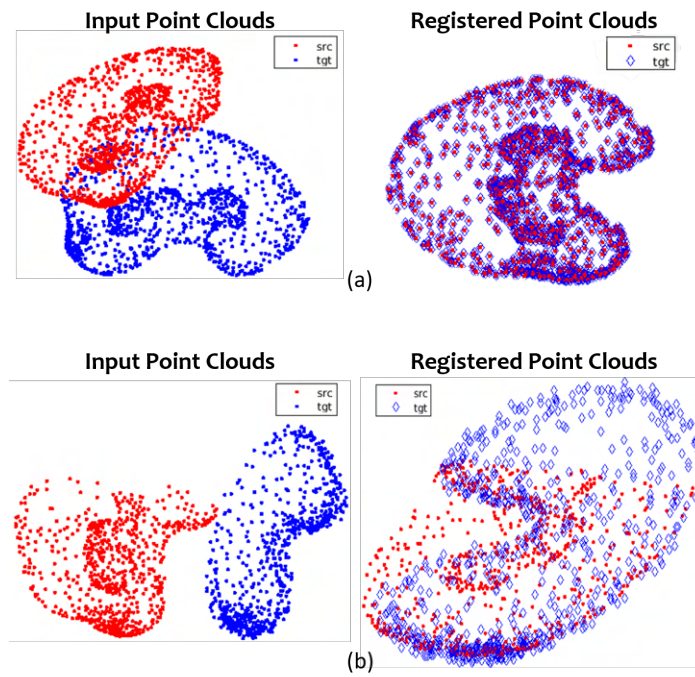


Figure 3.7 Two examples of kidney registration in different conditions. (a) Complete-to-complete registration with exact one-to-one correspondences. In the registered plot, each point of the *src* (red) falls within the diamond of the *tgt* (blue). (b) Partial-to-partial registration with noise and density variation. The input point clouds partially describe the kidney, sharing an overlap area. In this case, there are no exact one-to-one correspondences not even in the overlap area, as shown in the registered plot.

However, a limitation of correspondence-based methods is the requirement of strict one-to-one correspondence, as shown in Fig. 3.7a. This means that each point in one point cloud must be uniquely associated with exactly one corresponding point in the other point cloud. In the partial-to-partial case, this condition refers only to the points in the overlap area, and it occurs if they are distributed equally and have the same density. For instance, Fig. 3.7b shows how noise and density variation affect one-to-one correspondences in the overlap region. Gaussian noise is applied to each point of the *tgt* (blue), which is more noticeable along the edges of the kidney, causing them to appear irregular. The density variation is also visible as the two point clouds represent different kidney regions but contain the same number of points, resulting in a different distribution. This scenario closely resembles the surgical one, where the input data originates from two distinct sources (preoperative model and intraoperative reconstruction). Therefore, it is highly improbable that the point clouds are distributed equally, resulting in the absence of exact one-to-one correspondences and potentially leading to suboptimal registration.

In general, the refinement techniques presented aim to make the algorithms more robust in the presence of noise and to accept non-exact correspondences. However, not all algorithms address the issue of different densities between *src* and *tgt*.

Another issue with correspondence-based methods, which also applies to all other DL methods, is the generalization problem. These methods achieve high performance when the testing conditions match the training ones. However, the further the test conditions deviate from the training conditions, the more likely the algorithm is to fail. Such variations might include the presence of noise, differences in overlap percentages or between synthetic and real data. For instance, liver registration methods are typically trained on large synthetic datasets that simulate various deformations, suggesting that inter-patient variability may critically influence the models' generalizability.

Regarding time consumption, these methods depend significantly on hardware performance and the number of input points. For example, KCR-Net can perform a single registration in 100ms with 4096 input points.

Despite all these problems, the advantages of correspondence-based methods are significant and lie in improved inference times and accuracy, surpassing both traditional and hybrid methods. Furthermore, these methods offer the notable advantage of easy extension to non-rigid registration. While current surgical approaches remain focused on rigid registration by identifying a limited number of high-accuracy correspondences in deformed tissues and applying techniques like ICP, future advancements are expected to shift towards non-rigid registration. Developing more sophisticated networks capable of predicting a more exten-

sive set of correspondences could enable the effective use of non-rigid algorithms such as non-rigid ICP (N-ICP) or physics-based methods, expanding the possibilities for surgical applications. This will pave the way for developing real-time solutions for laparoscopic AR registration.

Correspondence-free methods

Instead of directly seeking correspondences, these methods focus on other aspects of point clouds, such as their overall geometric properties, local structures, or global features, to estimate the transformation that aligns them.

The pioneering method in this category is PointNetLK (Aoki et al., 2019), which is not an end-to-end method because it utilizes conventional alignment techniques based on Lucas-Kanade (LK) algorithm (Baker and Matthews, 2004) for aligning the PointNet features extracted from the two clouds.

End-to-end methods were subsequently developed; an interesting one is OMNet (Xu et al., 2021). The method is based on overlapping masks, which filter out non-overlapping points, preserving only the overlapping ones that are more valuable for estimating the transformation. The registration regression is based on global features of the overlap region.

Inspired by this idea, Guan et al. (2023) proposed the first DL-based approach to 3D-3D laparoscopic liver registration. The method uses the learned overlapping mask derived from the preoperative 3D model and intraoperative reconstruction to filter out non-overlapping regions, converting the two point clouds into a uniform shape. Subsequently, it aligns the extracted overlapping regions based on local and global features. Compared to OMNet, this method improves the extraction of local features following the RPMNet model.

A fascinating hybrid approach is RORNet (Wu et al., 2023b), which combines overlap masks and the correspondence matrix to obtain excellent registration results. The model involves estimating the overlap masks through a dual-branch network, where one branch calculates the overlap scores by exploiting the global features, and the other determines the similarity scores using pointwise features. Combining these two scores enables them to estimate a more refined mask, which effectively filters out non-overlapping points, thus improving correspondences.

Finally, FINet (Xu et al., 2022b) and PANet (Wu et al., 2023a) are two noteworthy methods that leverage distinct multi-branch strategies.

Discussion on correspondence-free methods: Overall, it is seen that correspondence-free methods estimate the transformation using global features, making them more suitable for solving problems where one-to-one correspondence cannot be assured, contrasting with

correspondence-based methods. However, depending solely on global features may yield less accurate predictions. To address this concern, some approaches have also begun incorporating point-wise or multi-scale features. Even in this case, the problem of generalization remains. Despite this, their performance remains comparable to correspondence-based methods and thus superior to traditional and hybrid methods. Moreover, as demonstrated by Guan et al. (2023), the approaches in this category effectively apply to surgical AR registration.

Gaussian Mixture Model (GMM)-based methods

The rationale here is to exploit point-to-distribution correspondences, which consists of mapping the points into probabilistic distributions and estimating the transformation by distribution alignment. In particular, GMM is perfect for this task as it describes the point cloud as a weighted sum of Gaussian distributions (Fig. 3.8).

First proposed by DeepGMR (Yuan et al., 2020) with a complete-to-complete registration, their method learns the GMM parameters for both *src* and *tgt*, then predicts the optimal transformation by minimizing the KL-divergence between them.

Unlike DeepGMR, which needs to build two GMM distributions for two inputs, UGMM (Huang et al., 2022) models both point clouds as samples from a single unified GMM. By leveraging self- and cross-attention, it predicts posterior mappings between points and GMM components to determine their parameters, ultimately estimating the transformation between the unified GMM's virtual centre and the input point clouds.

Recently, OGMM (Mei et al., 2023a) proposed a supervised method for partial-to-partial registration with an overlap-guided GMM. As in DeepGMR, two different GMMs are estimated for *src* and *tgt*, but in this case, the overlap scores are estimated to filter out

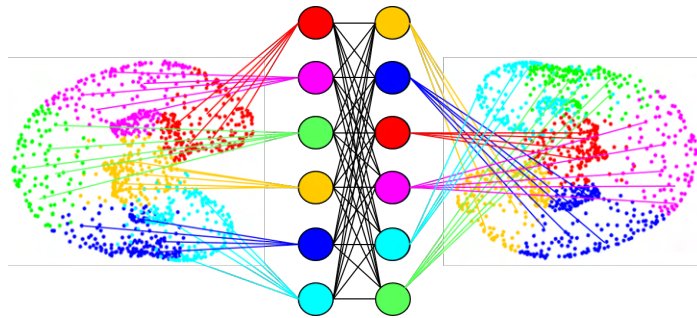


Figure 3.8 Distribution-to-distribution matching. GMMs represent point clouds as a combination of Gaussians, depicted in the figure as coloured circles. Each point is assigned to one of these Gaussians. The algorithm searches for matches between the Gaussians of each GMM.

the points that are not useful for partial registration. The same author also developed an unsupervised network version, UDPReg (Mei et al., 2023b).

Discussion on GMM-based methods: Altogether, only a few algorithms exist within this particular category. However, their accuracy frequently matches those of the previously discussed algorithms. The strength of these methods lies in the ability of GMM to describe point clouds as probability distributions, transforming the problem into a distribution-to-distribution registration without the need to establish explicit point correspondences. Hence, these methods can effectively address the absence of strict one-to-one correspondence. A key limitation is the challenge of extending these methods to non-rigid registration, as, to the best of our knowledge, no GMM-based DL approach has yet addressed this issue. Additionally, these methods often exhibit high computational complexity, requiring extensive training on large datasets. Lastly, generalization remains a significant concern. On the other hand, GMM-based algorithms demonstrate superior performance in scenarios characterized by noise, varying point densities, or repetitive patterns (Mei et al., 2023a,b). Finally, these methods seem highly efficient in terms of processing time; for instance, DeepGMR can complete a single registration in just 2 ms with 1,437 input points. Therefore, they may represent a good solution for surgical AR registration.

Non-rigid registration methods

All the algorithms described so far addressed the problem of rigid registration. However, discussing non-rigid registration in the surgical scenario would be more appropriate. Only a few DL-based methods address non-rigid point cloud registration, so we summarize them in this category.

ProRegNet (Fu et al., 2021b) stands out as one of the few non-rigid medical image registration methods incorporating point clouds within its architecture. The approach involves processing two prostate images, an MR and a TRUS (Transrectal ultrasound), and registering them. It comprises three main steps. In the first step, tetrahedron meshing generates volumetric point clouds from automatically segmented MR and TRUS prostate masks. In the second step, the training dataset is prepared, wherein the deformation field between the two images is estimated using FEM, along with establishing ground-truth correspondences. In the third step, these correspondences are used to train the network to estimate the deformation field from the two volumetric point clouds.

The concept of leveraging FEM to construct a training dataset is intriguing and potentially applicable even in scenarios involving 3D models. For a comprehensive overview of networks designed explicitly for non-rigid point cloud registration, refer to Monji-Azad et al. (2023).

An interesting algorithm for complete-to-complete registration is CorrNet3D (Zeng et al., 2021). Initially, it estimates the correspondence matrix between the two input clouds, which is then used to reorder the elements of the two input vectors for alignment purposes. Subsequently, these vectors are fed into an MLP block, which predicts the final registered clouds. This method is further developed by NrtNet (Hu et al., 2022), which integrates transformers and makes it unsupervised.

Moving on to partial-to-partial registration, Leopard (Li and Harada, 2022a) utilizes both self- and cross-attention mechanisms of transformers alongside a matching confidence score to infer correspondences. These are then used to estimate the registration using a N-ICP (Amberg et al., 2007) approach.

NDP (Li and Harada, 2022b) is a different approach based on MLP. Their intuition is that natural non-rigid motion typically exhibits a hierarchical structure, with higher hierarchies representing the global movements (i.e. global position, orientation) and lower hierarchies representing the local deformation. So, they present a pyramidal architecture, decomposing the global transformation into sub-motions.

Lastly, GraphSCNet (Qin et al., 2023b) proposes an outlier rejection network to find reliable correspondences, followed by a N-ICP to perform the registration. Their network leverages graph-based local spatial consistency to measure the local geometric compatibility among correspondences within a specific region.

Self-P2IR (Zhou et al., 2025) proposes a complete-to-partial point cloud registration pipeline. First, an intraoperative visible-surface point cloud is reconstructed by reprojecting a monocular depth map inside the liver mask (with a PCA-based scale normalization). Then, a KPConv + transformer module estimates robust correspondences and solves the initial SE(3) alignment via SVD, while a second stage predicts a dense non-rigid warp field using a structure-regularized deformation pyramid modulated by learned similarity weights. Notably, the approach is trained in a self-supervised manner on in-vivo data through differentiable rendering constraints, avoiding the need for ground-truth 3D–2D registration labels.

Discussion on non-rigid point cloud registration methods: In general, only a few non-rigid methods have been proposed for partial-to-partial registration. These methods predominantly rely on correspondence and use N-ICP for registration, with NDP being an exception. Similar to other DL methods for point cloud registration, challenges with exact correspondences, noise and generalization persist. Additionally, issues related to the deformation level arise because obtaining datasets with ground truth for non-rigid deformations is challenging. Many studies had to create their own customized datasets, which complicates algorithm performance comparison. Regarding time consumption, state-of-the-art methods like Leopard

predict correspondences in 100ms, with the overall registration time varying based on the number of N-ICP iterations required.

As discussed for correspondence-based methods, we anticipate that such non-rigid approaches, and even correspondence-free techniques, will emerge, facilitated by using masks to delineate deformation areas. Only one of the papers in this category has been developed specifically for medical imaging (Fu et al., 2021b), despite the necessity of having an end-to-end system in surgical AR to accelerate processes and enable registration to adapt to the dynamic intraoperative environment.

Summary of DL-based point cloud registration methods

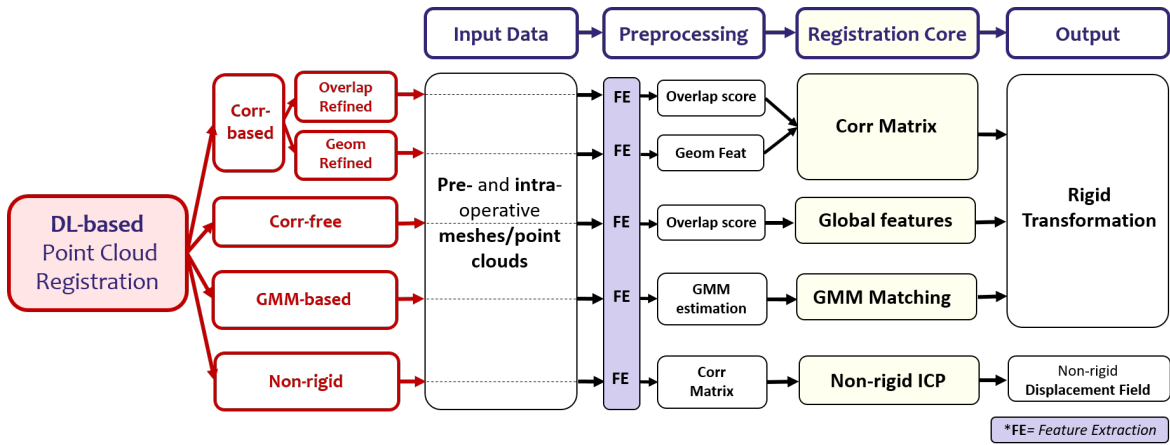


Figure 3.9 Schematic overview of DL-based point cloud registration methods.

Fig. 3.9 illustrates an overview of DL-based point cloud registration methods. Each approach takes two point clouds as input: the intraoperative cloud, obtained via segmentation, stereovision, or DL-based surface reconstruction (see Fig. 3.6), and the preoperative cloud, extracted directly from the 3D model. All methods include a feature-extraction (FE) stage, most commonly implemented with DGCNN, KPConv, or PointNet architectures. Feature extraction may be applied either to the full dense cloud or to a subset of keypoints after downsampling. Most approaches further refine point features by incorporating both transformer-based self-attention and cross-attention mechanisms. Beyond feature extraction, each method category employs a distinct registration engine tailored to its underlying logic. Importantly, any rigid correspondence-based model can, in principle, serve as the foundation for non-rigid registration: once accurate correspondences are predicted between deformed clouds, a downstream deformation-refinement step can produce the final alignment.

Part I

Traditional rigid registration methods

Introduction

This part examines conventional **rigid registration** approaches for surgical AR. We begin with these methods because they are widely used in the literature and provide a clear baseline: the achieved results, the associated time and operator workload, and the conditions under which rigidity assumptions fail. This foundation enables the identification of strengths and weaknesses before transitioning to learning-based methods in later stages.

This part is organised into three chapters:

- *Chapter 4* investigates **manual rigid registration** through a comparative usability study of human-computer interfaces for 3D model manipulation, identifying an interface that balances accuracy, speed, and operator workload. Manual registration remains the most accessible approach in surgical AR, so understanding its limits is essential before proposing automation. Therefore, we evaluated manual alignment on two fronts: **objective performance** (accuracy and time) and **user experience** (workload and usability) to establish a fair baseline and expose potential limitations. Using two representative interaction methods (3D mouse vs. Leap Motion), we show that manual alignment can achieve accurate overlays under controlled conditions.
- *Chapter 5* presents a pre-clinical study of **EVA Navi**, the surgical AR navigation system developed in our lab in earlier works. The system combines **point-based** patient initialization, **manual** registration refinement, and optical **tracking** to maintain real-time alignment of the 3D model with the laparoscopic view as the endoscope moves. We evaluate it **in vivo** during conventional laparoscopic surgery on a porcine model, focusing on **ureter identification**. This setup allows us to assess, in a realistic surgical scenario, how such a solution performs in terms of guidance and usability. In this study involving surgeons and residents, EVA reduced ureter identification time compared with the conventional view and was judged to be functional and efficient, supporting its role as a practical guidance tool.
- *Chapter 6* extends the **EVA** navigation system to **robot-assisted** laparoscopy. We remove the external tracking camera used in Chapter 5 and instead **update the AR overlay based on the kinematics of the robotic arm** holding the camera, after an initial manual registration. The solution couples a teleoperated robotic endoscope (software-defined RCM, hand-eye calibration) with our AR pipeline to maintain alignment as the camera moves. Overall, the system maintained a precise overlay after manual registration, showing the positive influence of an accurate camera tracking.

Chapter 4

Comparative usability study of human-computer interfaces for 3D model manipulation in surgical augmented reality application

4.1 Introduction

As discussed in Sect. 1.1.1, AR surgical navigation overlays patient-specific preoperative information (e.g., 3D models from CT/MRI) onto real-time intraoperative images to support anatomical understanding and guidance. A typical first step in an AR system is an initial rigid registration of the 3D models to the surgical image (Bernhardt et al., 2017b; Collins et al., 2021). Most of the approaches are manual or semi-automatic. While the latter requires physical landmarks or 3D surface features, which are usually hard to find in real organs; the manual approach requires Human-Computer Interfaces (HCI) to manipulate the models. The choice of the HCI is fundamental to develop an easy-to-use application with a short learning curve. In fact, the way the manual registration is performed can impact the users' physical and mental stress, influencing their final performances.

This chapter presents a human-centred usability study that aims to evaluate two modern HCI devices commonly proposed for 3D model manipulation in augmented reality applications (Mendes et al., 2019), i.e., the SpaceMouse[®] (3Dconnexion, Germany) and UltraLeap[®]

This work has been published as an **abstract**: A. Neri, V. Penza, and L. S. Mattos, “Comparative usability study of human-computer interfaces for 3d model manipulation in surgical augmented reality applications,” Hamlyn Symposium on Medical Robotics 2023, 2023.

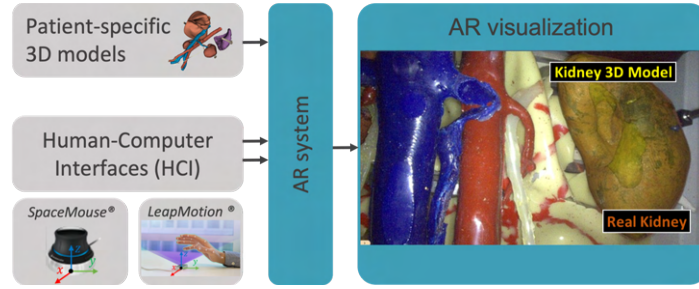


Figure 4.1 Overview of the AR system architecture and the HCI used to perform the manual registration.

(LeapMotion, USA). Accordingly, a software interface was developed, to allow the manipulation of 3D models in a minimally invasive partial nephrectomy scenario. Subjects were asked to perform a manual registration of a kidney's 3D model on the image of the real kidney. Quantitative and qualitative evaluation was performed to assess the cognitive and physical load imposed on the operators, which can support an informed selection of HCI to maximize system usability. These trials were performed under study protocol 229/2019 - ID 4621, approved by the Regional Ethics Committee of Liguria (Italy).

4.2 Materials and Methods

An AR system was developed to overlay, on the surgical image, the 2D projection of patient-specific 3D organ models, as shown in Fig. 4.1. The system takes as input (i) a set of patient-specific 3D models and (ii) the commands from two different HCI, the SpaceMouse® and the UltraLeap®. These HCI are used to perform the rigid manual registration between a 3D model and the corresponding phantom organ on the surgical image.

The AR visualization consists of a QT virtual environment with a virtual camera configured to behave like an endoscope. In fact, its projection matrix is set equal to the matrix of the undistorted camera. The *position*, *focal point* and *view up* axis of the virtual camera are set to visualize the 3D organ models located in the origin of the virtual environment. The endoscope image is represented behind the 3D models by mapping the undistorted surgical image on a plane parallel to the camera. The transparency of the 3D models can be changed to visualize the surgical image behind them.

Two different HCI, SpaceMouse® and UltraLeap®, were selected to control 6 DOF of a virtual 3D model to be overlaid on the surgical images. The UltraLeap® control is activated by the finger pinch. While the pinch is detected, the rotation and translation of the hand are used to transform the models with respect to the camera frame. The SpaceMouse® control

is implemented by applying the translation and twist offset as the model's velocity with respect to the camera frame. A scale factor is added in both HCI to facilitate a precise model manipulation.

4.2.1 Experimental protocol

The experiments included 13 subjects without experience in 3D object manipulation. Each experimental session consisted of two trials, where the task was respectively and randomly performed with the SpaceMouse® or the UltraLeap®. The task consists in accurately aligning the 3D model of the left kidney with the real video image of the corresponding kidney phantom, with the 3D virtual model always located at the scene centre as suggested by Bergström et al. (Bergström et al., 2021). For each trial, the user was asked to perform the same task on 3 different images of the kidney phantom in a different position and orientation. The sequence of images was the same for both trials.

At the beginning of each trial, the users had 5 to 10 minutes to familiarize themselves with the manipulation of HCI. After each trial, the user was asked to compile the SURG-TLX survey (Wilson et al., 2011). The task was considered completed when the user was satisfied with the achieved alignment.

Quantitative measures consisted of: (i) intersection over union (IOU), computed between the real kidney mask (GT) and the 3D model projection mask, to evaluate the alignment accuracy for each task, (ii) task execution time. Qualitative information is extracted from the survey answers describing each interface's perceived workload. Due to the low sample size, a non-parametric statistical test (Wilcoxon rank sum test) was used to analyze the data for significant differences, as it could not be evaluated under a normal distribution.

4.3 Results

Fig. 4.2 shows the results related to IOU and execution time for the two HCIs. The SpaceMouse® median IOU (0,84) is slightly better than the UltraLeap® one (0,82), while the latter holds a lower median execution time (55s) against the 81s of the first. For both IOU ($p = 0,77$) and execution time ($p = 0,19$), no significant differences have been found. It is noteworthy that for each task executed with the second interface, whatever it is, the execution time decreases, and the IOU obtained increases. This could mean that the learning curve of manipulating a 3D object has a very steep initial slope. Figure 2b presents the result of the surveys. All the dimensions, except for Physical Demands, have equivalent median values,

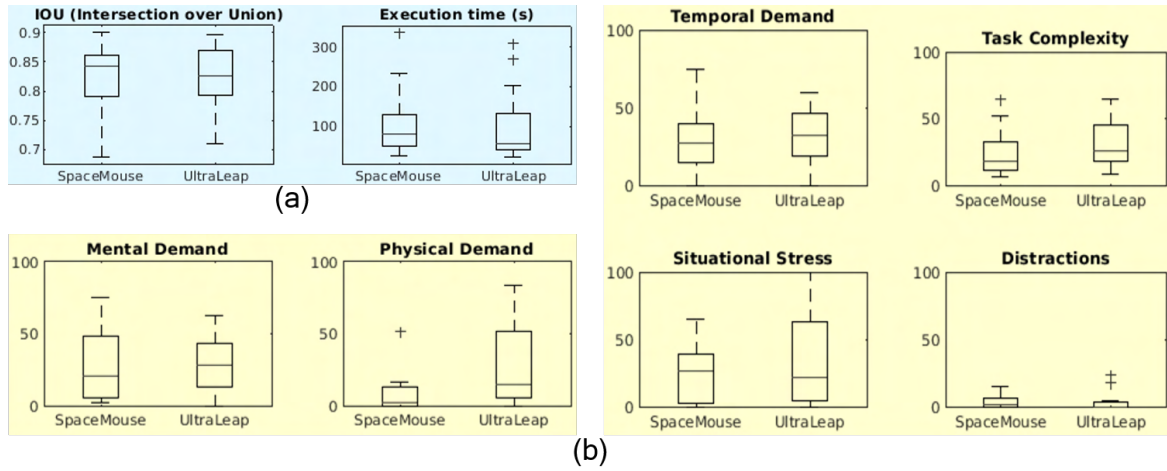


Figure 4.2 (a) IOU and execution time box plots (blue background). (b) SURG-TLX dimensions (yellow background).

and their Wilcoxon rank sum test does not show any difference. This indicates that the users found these dimensions comparable with both interfaces. The case of Physical Demands is particular; with a p-value of 0,068, it is barely over the threshold; collecting more data could lead to a significant difference between the two HCIs, where the UltraLeap[®] results in more physically demanding.

4.4 Discussion

In general, inexperienced users managed to complete the task with a good IOU ($>0,8$) and in less than 1min and 30s in both cases; this suggests that the two interfaces are natively intuitive and suitable to perform this task. Some users were very critical of UltraLeap[®], finding it challenging to complete the task with it. This was because the position with the arm raised was uncomfortable for them, and they had problems with the control algorithm. On the other side, the majority stated that they felt more in control with the SpaceMouse[®], which could explain why they decided to invest more time in it, feeling they could get better alignment. Considering that the choice of the interface turns out to be user-dependent, in this study the SpaceMouse[®] comes out as a slightly better option. The following steps will be to improve the UltraLeap[®] control algorithm to find less stressful hand gestures to manipulate the objects.

4.5 Conclusion

Overall, manual registration is a valuable component of the workflow, but it is best used in combination with other techniques (e.g. anatomy tracking and automatic rigid/non-rigid registration) rather than as a stand-alone solution in the operating room. Some limitations that emerged in this study are: (i) a stable view is required during alignment; keeping the laparoscope fixed is feasible but disrupts workflow and can prolong the procedure; (ii) without an automatic registration update, any subsequent camera motion invalidates the overlay, making repeated manual re-registration impractical and (iii) purely rigid alignment cannot compensate for soft-tissue motion and deformation (respiration, insufflation, instrument interaction). Manual methods are therefore best positioned as *refinement* tools within a surgeon-in-the-loop workflow, complementing automated tracking and non-rigid updates that maintain accuracy as the scene evolves.

Chapter 5

Evaluating the EVA surgical navigation system for ureteral identification in an in vivo porcine model

5.1 Introduction

According to the World Health Organization (WHO), unsafe surgical care causes substantial problems, such as major complications, occurring at an alarming rate of 25% during major surgeries and leading to a mortality rate of 0.5% – 5% (Safety et al., 2009). Minimally Invasive Surgery (MIS) and the introduction of robotic systems in the Operating Room (OR) have brought various advancements in surgical techniques in terms of manoeuvrability, precision and enhanced surgical field visualization. Despite that, a series of complications remain closely associated with the surgeon's experience level.

Specifically, Ureteral Injury (UI) is one of the most common forms of iatrogenic urological injuries, arising in approximately 0.2 – 1% of pelvic surgeries. Among these cases, 50% are observed during obstetric and gynecological surgeries, with 30% specifically occurring during radical hysterectomies (Jha et al., 2004; Kumari Rani, 2020). UI also occurs in colorectal surgery with an incidence ranging between 0.2 – 1.5% (Croghan et al., 2019). Other less common causes of UI include ovarian tumor removal, transabdominal urethropepy, and abdominal vascular surgery. More than 50% of UI that occurs intraoperatively are not

This work has been published as: Penza, V., Soriero, D., Sperotto, B., Neri, A., Ortiz, J., Pertile, D., ..., Mattos, L. S. (2025). Evaluating the EVA surgical navigation system for ureteral identification in an in vivo porcine model. *Scientific Reports*, 15(1), 16976. Publicly available at: <https://doi.org/10.1038/s41598-025-00138-8>.

immediately recognised and treated, especially during emergency surgery (Burks and Santucci, 2014). Consequently, these injuries are associated with increased morbidity, mortality, hospital stay, and healthcare costs that can exceed 30,000 US\$ per case (Halabi et al., 2014). UI also represents a frequent reason for legal disputes with serious psychological, economic and professional burdens for both patients and doctors (Jha, 2015); in fact, this complication represents 17% of non-obstetric lawsuits filed against gynecologists in the USA (Jha et al., 2004).

Currently, there is not a single non-invasive prevention strategy that effectively reduces the risk of UI (Brollo et al., 2023). Surgical exposure of the ureter and prophylactic ureteral stenting can help in its identification, but are both invasive approaches, not free from risks and therefore not included in any management protocol (Brollo et al., 2023). Nonetheless, in case of high-risk patients, prophylactic placement of ureteral stents can be considered as an important strategy for the prevention of ureteral injury. Pokala et al. conducted a comparison between simultaneous intraoperative vs sequential ureteral stent placement. Their findings revealed that simultaneous stent placement significantly reduced the overall operative time by 19 minutes, without leading to an increase in morbidity (Pokala et al., 2007). Studies investigating the identification of ureters using ureter-specific vital dyes are still in the preclinical phase, showing promise for future applications (Lau et al., 2020). Notably, indocyanine green, the most extensively studied vital dye, does not undergo renal metabolism. Consequently, direct instillation into the ureter is required, either via cystoscopy using a rigid cystoscope or by pre-placing a ureteral stent. However, these procedures are invasive and significantly elevate the risk of complications, along with prolonging operative times (White et al., 2021).

The use of computer-assisted technologies during intraoperative surgical procedures offers a promising solution for enhancing safety in clinical settings. These technologies aim to improve the surgeon's awareness and capabilities by providing additional information from preoperative examinations, monitoring the surgical site, and alerting the surgeon to potential adverse events. Virtual Reality (VR) and Augmented Reality (AR) tools, particularly in laparoscopic assistive technologies, have shown significant potential in achieving these objectives over the past decade (Amparore et al., 2022; Bernhardt et al., 2017b; Penza et al., 2017, 2020a). Atallah *et al.* demonstrated that stereotaxic surgical navigation in rectal cancer surgery is a valid technique for guiding dissection during surgery (Atallah et al., 2016, 2019). This technique enables the identification of the optimal dissection point using radiological imaging, thereby facilitating the preoperative localization of anatomical structures before their visualization in the surgical field. However, it is important to note that



Figure 5.1 An example of AR visualization of EVA Surgical Navigation System during the in-vivo study on animal model.

while radiological anatomy remains static, surgical anatomy inherently exhibits deformable characteristics, which may lead to disparities between the two. On the other hand, Kiyomatsu *et al.* demonstrated that during minimally invasive surgery in a porcine model, the difference between radiological anatomy and surgical anatomy, especially concerning pelvic vessels, is minimal, being less than 2.1 mm (Kiyomatsu et al., 2021). These results support the idea that an intraoperative navigation system capable of integrating radiological and endoscopic images could serve as a significant support for surgeons.

This chapter presents an in-vivo study focused on evaluating the feasibility of using the surgical navigation system EVA for intraoperative ureteral identification on an animal model. EVA is a system previously developed by the authors to enhance surgical navigation, aimed at improving the quality and safety of MIS (Penza et al., 2020b). In fact, acting like an intraoperative GPS for surgery, it allows the visualization of a dynamic VR and/or AR environment that shows 3D models of patient-specific organs in real-time, aligned and synchronized with the endoscopic view. The EVA-VR system was evaluated in a laboratory setting by 14 surgeons, as reported in previous publication (Penza et al., 2020b). Results showed a navigation precision lower than 5mm, a reduction in task execution time by 15% and a reduction in the size of the peritoneal incision (to reach the lesion) by 46% compared to the control group. Surgeon's feedback highlighted how the use of EVA system improves the sense of orientation during surgery, therefore also helping in the identification of structures of interest.

Based on the preliminary findings, this study aims to further explore the potential of AR technology as a step forward from the VR GPS technology in surgical navigation. To this end, this chapter reports the outcomes of a feasibility study focused on evaluating the EVA-AR system under conditions closely resembling those of real surgery, using a live porcine model

during MIS (Fig. 5.1). The primary objective of this study is to evaluate the efficacy and usability of the system in identifying the ureter during MIS. The secondary endpoint involves assessing the system's initialization and calibration in an OR setting, measuring both the time required for these processes and the calibration error. Here, the description of the in-vivo study follows the ARRIVE 2.0 guidelines (Kilkenny et al., 2010a,b; Percie du Sert et al., 2020).

This chapter is structured as follows: Section Materials and Methods provides a brief overview of the EVA surgical navigation system and describes the study design and protocol. Technical and clinical results are presented in Section Results, followed by discussion and conclusions in Section Discussion.

5.2 Materials and Methods

5.2.1 EVA-AR surgical navigation system

EVA-AR is a surgical navigation system designed to offer an AR view of the surgical field. It provides a dynamic AR environment that shows in real-time 3D models of patient-specific organs overlaid and synchronized with the endoscopic view, showing Structures Of Interest (SOI) identified during the preoperative planning (i.e., lesions, critical vessels, etc.). This overlay is obtained by manually registering the 3D organ models with the actual ones, using fiducial landmarks selected on rigid structures, such as bones. This process will be referred to as *patient calibration*. The alignment is then continuously updated during the surgical intervention based on the movements of the endoscope, monitored by an external Optical Tracking System. The idea behind EVA is to make the preoperative planning easily accessible in the surgical room, relieving the surgeons' cognitive load required to mentally remember and fuse the plan with the intraoperative endoscopic images. To this end, EVA allows intuitive navigation and guidance, supporting the surgeon's decision-making process. The system capabilities are enhanced by advanced functionalities, such as changing the transparency of organs to visualize internal structures (e.g., tissue abnormalities, vessel structures, etc.). As shown in Fig. 5.2, the system is configured as a medical cart that integrates (i) an Optical Tracking Camera to track the endoscope motions, (ii) a full-HD monitor for the visualization of the AR functionalities, and (iii) a computer for the generation of the AR view. The tracking of the endoscope is achieved through a maker tool positioned on the body of the surgical endoscope.

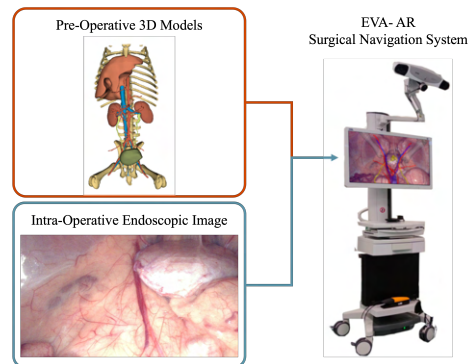


Figure 5.2 The EVA surgical navigation system: on the left, the 3D models of the pig's selected organs generated with HA3D™ (Medics, Italy) and the intraoperative endoscopic image; on the right, the system medical cart, composed of a screen, an optical tracking system and a computer.

Patient-specific 3D models reconstruction

The EVA system requires patient-specific 3D models as input, which can be obtained from diagnostic images such as CT scans or MRI scans. For this study, the HA3D™ service (Medics Srl, Italy) was used to generate the 3D models using CT scans, as illustrated in Fig. 1. HA3D™ technology integrates engineering modeling knowledge with clinical know-how in order to obtain 3D patient-specific medical-grade models based on 2D images (Medics 3D, 2024). HA3D™ reconstructions can be manipulated by the surgeon, generating 3D model views that can be fitted, rotated, zoomed and virtually dissected, allowing a clearer study of the lesion itself. The reconstruction can also reveal possible anomalies near the neoplasm, such as inflammation, lymph nodes or smaller adjacent lesions. This service has already been verified in a pilot study aimed at evaluating the efficacy of high-resolution preoperative 3D reconstructions for lesion localization in oncological colorectal surgery (Soriero et al., 2022).

5.2.2 Study protocol

The assessment study of EVA-AR was conducted at the Mini Invasive Advanced Surgery (MIAS) Academy with approval from the Italian Ministry of Health n°654/2023-PR (Risp. a prot. AC827.8). All experiments involving a live animal were performed in accordance with the ARRIVE guidelines and relevant regulations (Kilkenny et al., 2010a,b). Experiments involving human participants were conducted in compliance with the Declaration of Helsinki and all applicable ethical standards and regulations. Informed consent was obtained from all

participants for study participation and publication of any identifying information/images in an open-access format.

A single male Yorkshire pig weighing between 20 to 30 kg, without genetic modification and in good health status, was used for the study. The animal was housed at the MIAS site and allowed a 12-hour acclimatization period before undergoing an abdominal CT with contrast under general anesthesia. The CT scan included an arterial phase, a late venous phase, and a urographic phase with contrast medium bolus to simultaneously enhance all anatomical structures, including arterial, venous, and urological components.

Before image acquisition, the animal was positioned supine in the same orientation it would assume during surgery, with the goal of minimizing position-induced anatomical deformation. The radiological images were acquired with a slice thickness of 1.5mm. Subsequently, the animal was returned to the MIAS animal facility and remained there until the day of the experiment. The radiological images were transmitted via the Medics platform and used to generate 3D graphical reconstructions within three working days. The reconstructed structures were the retroperitoneal and pelvic organs, which are minimally affected by position-induced changes and pneumoperitoneum.

The experiment took place in the MIAS operating room, where the animal received general anesthesia administered by a trained veterinarian. The animal was positioned supine and in Trendelenburg position. Three trocars were used at the level of the transverse supra-umbilical line, and the bladder was emptied and anchored to the upper abdominal wall with a fixed trans point. At the conclusion of the study, euthanasia was performed in accordance with approved veterinary protocols. Standard minimally invasive surgery instrumentation, including a laparoscopic column equipped with TIPCAM 1 S 3D LAP, 0° optics (KARL STORZ - ENDOSKOPE, Tuttlingen, Germany), and dissecting and grasping instruments were used.

All anesthetic and euthanasia procedures followed current veterinary best practices. The animal received an intramuscular administration of tiletamine at 4 mg/kg, zolazepam 4 mg/kg (Zoletil), along with dexmedetomidine at 10 mcg/kg (Dexdomitor). Intravenous access was established in both ear veins using 20-gauge catheters. Subsequently, after IV administration of propofol at 4 mg/kg (Propovet) the animal was intubated and mechanically ventilated with a tidal volume of 10-13mL/kg. General anesthesia was maintained with 2~3% sevoflurane (Sevflo) throughout the procedure. Analgesia was provided by an IV constant rate infusion of fentanyl at 30mcg/kg/h (Fentadon). At the end of the procedure, the animal was euthanized using the pharmacological procedure of intravenous administration of pentobarbital sodium (150 mg/kg). This method ensures a painless procedure, as it rapidly depresses the central

nervous system, leading to loss of consciousness and cessation of respiratory and cardiac functions.

The experimental evaluation was divided in two phases. The first phase was focused on assessing the system's initialization and calibration in an OR setting, later referred to as *Technical Evaluation*. The second phase was focused on evaluating the efficacy and usability of the system in identifying the ureter during MIS, later referred to as *Clinical Evaluation*.

Technical evaluation

A trained surgeon, with 10 years of experience in MIS and RMIS, and 5 years of experience with animal surgical experimentation, was asked to conduct a technical evaluation of the EVA system. The evaluation consisted in:

- **OR Feasibility** Qualitative assessment of the feasibility of integrating EVA's cart in an operating room;
- **Configuration Assessment** Qualitative assessment was performed for the system start-up configuration, to evaluate its proper initialization and ease of use; while a quantitative assessment of the *patient calibration* accuracy ($patientCalib_{acc}$) was performed by computing the re-projection error between the corresponding virtual and real points selected on the models and on the pig, respectively; Specifically, the reprojection error was measured as the Euclidean distance between corresponding virtual points (selected on the preoperative 3D model) and their real counterparts (manually identified on the pig's anatomy), which were re-projected onto the virtual 3D model using the rigid transformation identified during the calibration process. These points were selected based on anatomical landmarks visible in both the virtual and real environments. The real points were acquired by pointing the tip of a pointer tracked by the tracking system on the skin of the animal model.

Clinical evaluation

Four surgeons participated to the study, including 2 residents and 2 expert surgeons, all male with an age range of 31 to 43 years and an average of 36 years. All of them were from the field of general surgery. Each experimental session included two trials, during which participants performed the same task with and without the support of the EVA-AR system.

To optimize the use of the animal, non-destructive tasks were selected to prevent alterations in its anatomy that could affect the task's validity for subsequent participants. The

chosen task was the identification of the ureter. The anatomy of the pig's ureter, like all retroperitoneal organs, closely resembles that of humans. However, a notable difference lies in the representation of retroperitoneal adipose tissue, which is minimal in pigs. Consequently, retroperitoneal structures are easily visible through the peritoneum. For this reason, the task was designed to be performed by residents rather than experienced surgeons, who would have found it too easy. This setup allowed for a better assessment of how the EVA system supports less experienced surgeons in identifying key anatomical structures while ensuring validation from an experienced supervisor.

The participants were divided into two groups (Group A and Group B), each comprising an experienced surgeon and a resident. Group A was assigned the task of identifying the left ureter using the EVA-AR system and the right ureter without its assistance, while Group B performed the opposite. The task was considered complete when the participant confidently identified the ureter. For each group, the resident conducted the identification task, while the expert surgeon manipulated the endoscope and confirmed the position of the ureter. The task required the participant to move the intestinal loops and localize the ureter. For each task, the completion time t_{task} was measured, and the correspondence between the ureter 3D model and the real structure was qualitatively evaluated. In addition, the subjects were asked to fill the following questionnaires:

- **Face Validity.** A custom face validity questionnaire was developed to assess the relevance of the setup, the system and the tasks executed. The specific questions are reported in Fig. 5.6.
- **System Usability Scale (SUS).** The SUS is a widely used questionnaire-based method for assessing the perceived usability of a system (Brooke et al., 1996). It consists of ten statements that participants can evaluate on a Likert scale ranging from "strongly disagree" to "strongly agree". The statements are designed to cover aspects of usability, efficiency, and learnability. After collecting individual responses, scores are calculated and transformed to provide a global usability score (SUS_{score}) on a scale from 0 to 100.
- **User Experience Questionnaire (UEQ).** The UEQ is a fast and reliable questionnaire to measure the User Experience of interactive products (Laugwitz et al., 2008). The questionnaire investigates different aspects of the user experience: classical usability aspects, such as efficiency, perspicuity, dependability, and user experience aspects, such as originality and stimulation. The questionnaire uses pairs of contrasting attributes, and participants rate their experience on a scale from 0 to 7.

- **The Surgical Task Load Index (Surg-TLX).** The Surg-TLX questionnaire is designed to assess the perceived workload experienced by individuals performing surgical tasks. It takes into account various factors contributing to the overall workload, including mental, physical, and temporal demands, as well as factors like task complexity, situational stress and distractions, on a scale from 0 to 100.

While Face Validity, SUS, and UEQ were completed by both residents and surgeons, SurgTLX was filled out only by the residents who directly performed the tasks. Moreover, the surgeons completed the questionnaire only once, after completing both tasks. The results from these questionnaires were analyzed by computing the median and interquartile values, which are represented using boxplots, except for the UEQ results that were analyzed using the standard data analysis tool provided for the questionnaire.

5.3 Results and Discussion

5.3.1 Technical evaluation results

OR Feasibility

The dimensions of the EVA system cart are similar to those of a conventional laparoscopic column. Fig. 5.3 demonstrates its compatibility with the placement of other instruments, indicating its potential for seamless integration into conventional ORs. The cart must be positioned such as to place the Optical Tracking Camera at a distance between $1m$ and $2.4m$ from the operating field, enabling installation outside the sterile area. During our experiment, we ensured that the tracking camera maintained a clear line of sight to the surgical field by instructing the surgeon to orient it toward the center of the abdomen while keeping the camera as high as possible. This positioning was chosen to optimize visibility and tracking accuracy while minimizing occlusions. No significant tracking occlusions were observed, suggesting that the selected placement was effective in maintaining a continuous and stable tracking signal. From an ergonomic point of view, the monitor for visualizing the AR environment is positioned at eye level when standing. The endoscope marker tool, attached on the laparoscopic endoscope body, is small in size and did not interfere or impede the surgical procedure, as indicated also by the results of the face validity questionnaire. Due to its position, this tool should be made of a sterilizable material.



Figure 5.3 Positioning of EVA Surgical Navigation System in a OR-like setting.

Configuration assessment

The Graphical User Interface (GUI) of the system was considered simple and intuitive. In order to perform the *patient calibration* it was necessary to choose at least 3 anatomical landmarks that are fixed (i.e. located on the ribcage), easily palpable and recognisable from outside by the surgeon. We chose the pubic symphysis and the medial margin of the last ribs, bilaterally. Regarding the choice of ribcage landmarks, we acknowledge that respiratory motion affects the ribcage. However, we selected the ribcage because it is a relatively rigid structure that maintains a stable spatial relationship with the preoperative 3D model and remains consistent throughout the intervention. While the ribcage moves due to breathing, this motion was minimal compared to other anatomical regions in the pig model. In human procedures, iliac crests would be the preferred landmarks for patient calibration due to their stability and independence from respiratory motion. However, in pigs, this was not feasible due to anatomical differences, making the ribcage the most viable alternative for calibration. Fig. 5.4 shows the process of landmarks selection. The calibration process took on average 1.04min.

A $patientCalib_{acc}$ of $1.7cm \pm 0.002$ was measured. The sources for this calibration error can be attributed partially to the pneumoperitoneum-induced changes and to the landmarks detection error. In fact, the virtual landmarks chosen on the 3D models did not take into account the skin-fat-muscle layers covering the bones.



Figure 5.4 An instance of identifying anatomical landmarks during the calibration process between the preoperative 3D models and the animal model.

5.3.2 Clinical evaluation results

The average task execution time t_{task} for all the tasks conducted with and without the support of the EVA-AR system were respectively $41.41s \pm 11.11s$ and $81.41s \pm 62.10s$; Tab. 5.1 shows raw values for task execution time for all the trials.

Trial	Execution Time (s)
Trial 1 - NO EVA	125.33
Trial 1 - EVA	33.55
Trial 2 - EVA	49.27
Trial 2 - NO EVA	37.50

Table 5.1 Execution times for all trials with and without EVA-AR.

The surgeon had access to both the traditional laparoscopic column screen and EVA's screen. The EVA screen displayed AR images, where the 3D models were continuously updated in real time to align with the endoscope view. The AR views were presented without noticeable delays compared to the endoscopic images and without any inconsistencies. Following the initial *patient calibration*, the 3D models demonstrated a high level of congruence with the anatomical structures of the pig. In both tasks conducted with the support of EVA, manual adjustment using the mouse was necessary at the beginning of the trial to refine the registration between the 3D models and the target structures. The reference points used for alignment included the kidneys and some features of the visible pelvic anatomy. In both cases, surgeons confirmed that the projected 3D model of the ureter accurately corresponded to the real anatomical structure, as shown in Fig. 5.5.

Fig. 5.6 reports the boxplot of the results from the **Face Validity questionnaire**, indicating the median values and interquartile range for each question, with responses separated

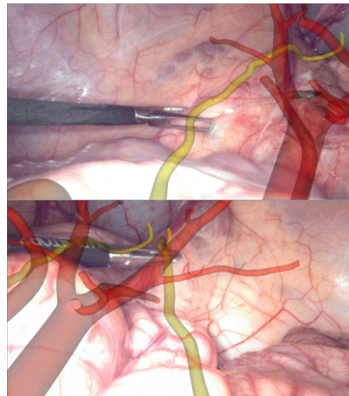


Figure 5.5 Sample frames from ureter identification task with the support of EVA-AR. In these pictures, the surgeon was indicating with the instrument the location of the ureter.

between surgeons (in blue) and residents (in orange). An analysis of the results is provided below:

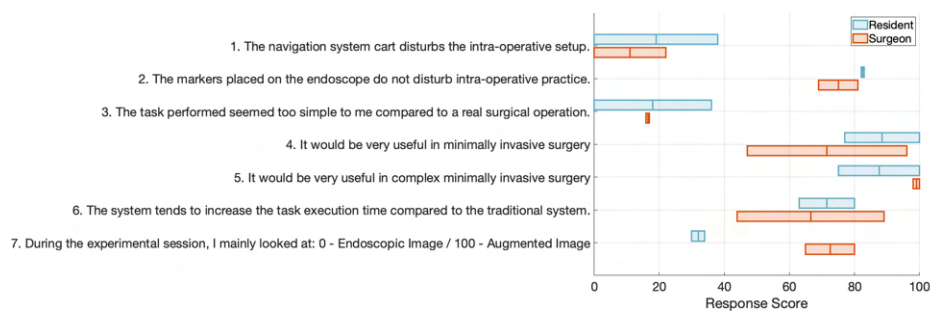


Figure 5.6 Face Validity Questionnaire Results.

1. The navigation system cart's disturbance to the intraoperative setup received a relatively low score, indicating that it may not significantly disrupt the surgical environment;
2. Both residents and surgeons reported that the markers placed on the endoscope did not interfere with intraoperative practice;
3. Participants felt that the task performed was not too simple compared to a real surgical operation, as indicated by the relatively high score.
4. Participants considered the system to be useful in MIS, in particular for the surgeons, while performing complex surgeries.
5. The system's impact on task execution time compared to traditional methods received a moderately high score, suggesting that it may increase the time required for certain

tasks; This opinion however is in contrast with the measure execution time, which appears to be lower when using the EVA system. We hypothesize this refers to the time to configure the system together with its usage during the surgery.

6. During the experimental session, participants' attention was divided between the traditional laparoscopy screen and EVA's augmented reality screen. A difference in preference was observed between residents and surgeons, the latter focusing more on the AR display than on the endoscopic video.

While the Face Validity questionnaire provided valuable qualitative feedback on the execution task, system usability and intraoperative integration of the EVA system, we acknowledge the potential for social desirability bias in custom-made questionnaire. This limitation is particularly relevant given the small number of participants and the exploratory nature of the study. To mitigate this, the questionnaire was designed to include both positive and critical evaluation aspects. Additionally, future studies will incorporate standardized usability assessments with a larger participant pool to further validate these findings.

Results of the **SUS questionnaire** are presented in Fig. 5.7, where boxplots are reported with responses separated between surgeons and residents. The overall score obtained with the SUS questionnaire is $SUS_{score} = 73.75$, with very similar results from residents (72.5) and surgeons (75). This indicates a general positive perception of the system's usability, as a SUS score above 68 is considered above the average. Focusing on the single questions, the subjects expressed a high likelihood of frequent system usage and found the system easy to use. While there were mixed feelings regarding the system's complexity, subjects generally perceived good integration of system functions and low levels of inconsistency. The majority of respondents felt confident in using the system and expressed confidence in its learnability, expecting most people to learn to use it quickly (more for surgeons). While there was some indication of a moderate need for learning before becoming proficient with the system, overall, the findings suggest a favorable perception of the system's usability with room for improvement in addressing perceived complexity and learning needs.

The **UEQ questionnaire** results, show in Fig. 5.8, indicate that the EVA-AR system was rated highly across several user experience attributes. The highest mean score was for Attractiveness (2.250 ± 0.68 for both residents and surgeons), suggesting a positive overall impression of the system. Perspicuity and Efficiency were evaluated quite differently from residents and surgeons, with generally positive scores from the residents (2.25 ± 0.5 and 2.135 ± 0.28 respectively) and lower scores from the surgeons (0.875 ± 0.28 and 0.625 ± 0.28 respectively). This suggests that, in general, participants found the system relatively easy

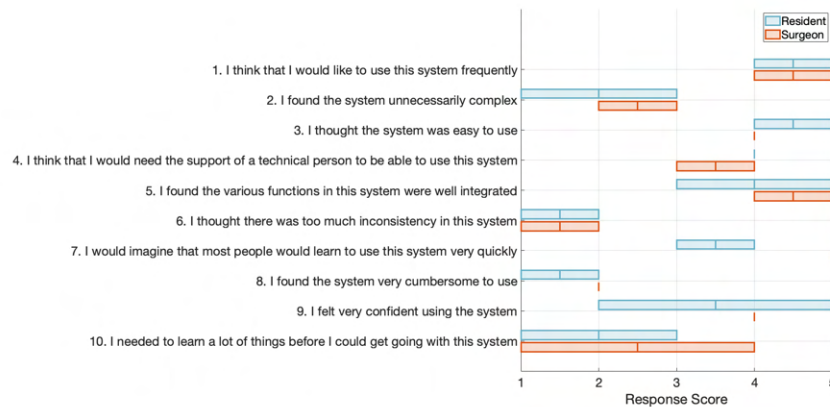


Figure 5.7 System Usability Scale (SUS) Results.

to learn and use, although there is room for improvement in these areas. Dependability and Stimulation also received favorable evaluations both groups, with residents providing slightly higher scores (2.0 ± 2.0 and 2.215 ± 1.53 respectively) than surgeons (1.38 ± 1.53 and 1.38 ± 0.03 respectively), highlighting the system's ability to engage and interest users. Finally, Novelty was the only metric evaluated more positively by surgeons (2.25 ± 0.0) than residents (1.5 ± 0.0), but it still remained positive overall.

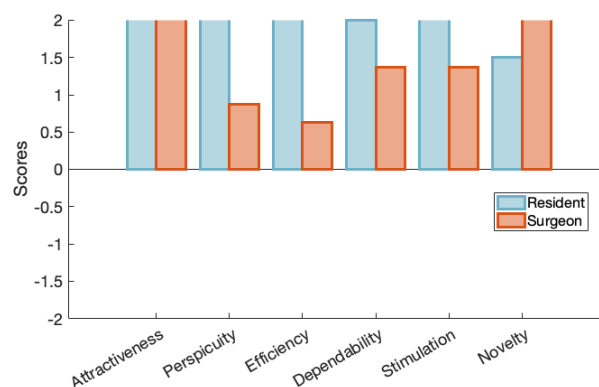


Figure 5.8 UEQ questionnaire results

The results obtained from the **Surg-TLX questionnaire** are presented in Fig. 5.9. These results reflect feedback exclusively from the residents who directly performed the task. Due to the small sample size, no statistical analysis was performed. However, the box plots shows an increased median and interquartile range in particular for the Mental Demand, Task Complexity and Distractions for the task executed without the support of EVA. These

findings suggest that EVA may alleviate the challenges associated with surgical navigation and identification of critical structures.

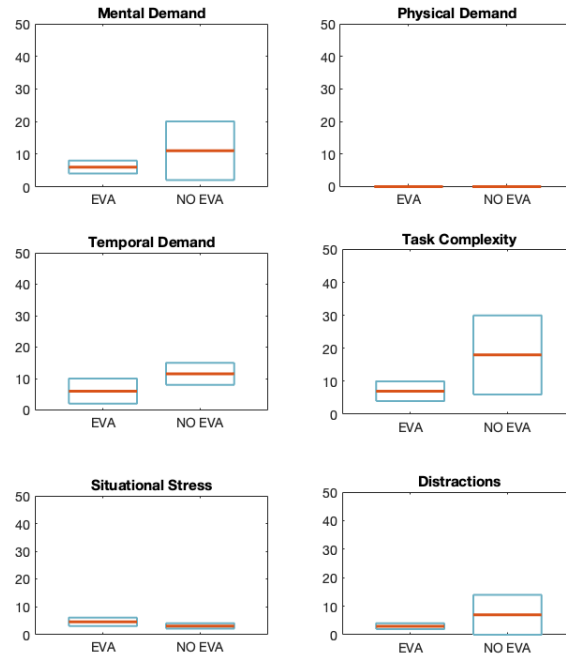


Figure 5.9 Surg-TLX questionnaire results. The y-axis was truncated at 50 for better visibility of the results.

5.4 Conclusion

This chapter presented an in-vivo study on an animal model performed to explore the feasibility of the EVA-AR surgical navigation system for uretral identification. Based on an optical tracking system, EVA aligns 3D reconstructions of patient-specific organs, generated from preoperative radiological images, with intraoperative images captured by the endoscope. EVA-AR continuously updates the view of the 3D model in real-time, enabling a congruent alignment between virtual anatomical structures and those observed by the surgeon.

Given the challenges of applying AR navigation in general abdominal surgery, particularly the deformation of soft tissues and their positional changes relative to preoperative planning, we focused this study on retroperitoneal and pelvic organs. Despite being soft, these organs exhibit relative stability in position, with less susceptibility to modification by factors such as

patient positioning and pneumoperitoneum-induced stimuli. This decision was influenced by findings from studies conducted on a porcine model by Kiyomatsu (Kiyomatsu et al., 2021).

This study demonstrated how EVA-AR can enable intraoperative surgical navigation. The EVA cart integrated well into the operating room and did not hinder the use of other equipment. The system was fully calibrated and operational in less than 15 minutes, requiring the involvement of two individuals, one at the cart and one at the operating table. According to questionnaire results, the participants who used EVA-AR believe that the system is functional and efficient, easy to learn, and user-friendly. The participants observed both the endoscopic video on the traditional laparoscopic display and on EVA's AR display during task execution, defining EVA as a supportive tool.

However, further research is still necessary to validate this initial finding. Limitations of this study include: i) The limited number of subjects who tested EVA; ii) The fact that experiments were performed on a single animal; iii) The exclusion of more invasive actions during the ureter identification trials; iv) The absence of retroperitoneal fat in pigs, which facilitates ureter identification through the parietal peritoneum. Further studies are already planned to address the shortcomings identified in this feasibility study before experimenting with EVA in the operating room on humans.

Overall, this research focused on the problem of ureter identification, which is a crucial aspect in numerous surgical procedures. In this context, the support of EVA-AR navigation system appears to have positively influenced the task, facilitating the localization of the actual ureter. Our registration pipeline combined an initial point-based registration, optional manual refinement, and real-time updates via tracking. In this setting, it was not possible to quantitatively assess registration accuracy, since the ureter could not be fully exposed without risking damage; consequently, no IoU or TRE could be computed on its surface. Nevertheless, the execution-time analysis showed that surgeons were faster in identifying the ureter when using EVA. This suggests that the registration quality was sufficient to provide useful guidance and to simplify the task in practice.

Chapter 6

Augmented Reality navigation in robot-assisted surgery with a teleoperated robotic endoscope

6.1 Introduction

Sect. 1.1.1 highlighted the benefits of AR in minimally invasive surgery, notably improving situational awareness and reducing cognitive load. These advantages can naturally extend to robotic surgery, where coupling AR with robotic platforms can further support intra-operative guidance and decision-making. In fact, since surgical robotics is reaching an unprecedented technological level in most facets, one of the next challenges is the introduction of awareness into the surgical room, which an AR system can represent. Context-aware surgical guidance will not only enhance the surgeon's capabilities and support surgeons with different degrees of experience, but it will also enable autonomous systems in the future.

In recent years, several applications of AR applied to robotic surgery have been demonstrated, as reported in (Makhataeva and Varol, 2020; Qian et al., 2019). Different approaches were used to register the model with the endoscopic image, via manual alignment (Neri et al., 2023), fiducial-based registration (Penza et al., 2020b), optical-flow-based semi-autonomous registration (Sengiku et al., 2017) or other advanced algorithms (Amparore et al., 2022; Penza et al., 2017, 2018b). So far, only a few studies used the robot kinematics to update the

This work has been published as: Penza, V., Neri, A., Koskinopoulou, M., Turco, E., Soriero, D., Scabini, S., ..., Mattos, L. S. (2023, October). Augmented reality navigation in robot-assisted surgery with a teleoperated robotic endoscope. In 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 4621-4626). IEEE. Publicly available at: [10.1109/IROS55552.2023.10342282](https://doi.org/10.1109/IROS55552.2023.10342282).

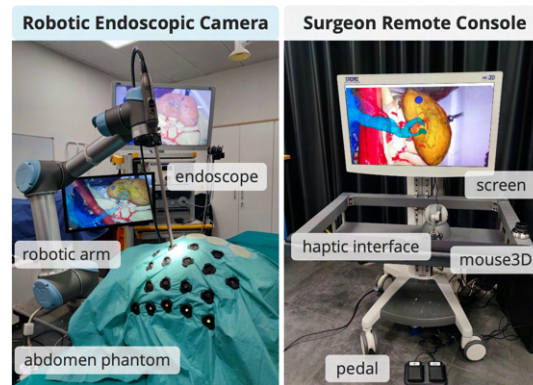


Figure 6.1 Overview of the AR navigation system integrated with the Robotic Endoscopic Camera teleoperated through the Surgeon Remote Console.

alignment obtained with the registration, and to the best of our knowledge all of them were using the da Vinci surgical system (Coste-Manière et al., 2004; Singla et al., 2017).

This chapter presents an AR navigation system integrated with a 6 DoF (Degree of Freedom) robotic endoscopic camera, that can be remotely teleoperated, as shown in Fig. 6.1. The AR navigation system dynamically visualises the patient-specific 3D models overlaid on the endoscopic image, inspired by (Penza et al., 2020b). This is achieved by manually registering the 3D organ models with the actual ones depicted in the endoscopic images. The resulting overlap is then updated based on the endoscope's movement, which is attached to a UR5 robotic arm controlled by a teleoperation system with a haptic interface and two pedals. The robot end-effector movements will be constrained to a virtual Remote Center of Motion (RCM). This is necessary since the context is minimally invasive surgery and the UR5 does not have a mechanical RCM. Thus, the main contribution of this work lies in providing AR guidance capabilities to an accurate robotic arm holding the endoscopic camera, improving the operator's (and the robotic system's) awareness of the surgical field.

The chapter is organized as follows: the main technical features of the system are described in Section 6.2, comprising the remote console (6.2.1), the teleoperated robotic camera development (6.2.2, 6.2.3), the system calibration (6.2.4), and the AR navigation system (6.2.5). System assessment experiments are presented in Section 6.3 and the results are presented in Section 6.4. Finally, discussion and conclusions are reported in Section 6.5.

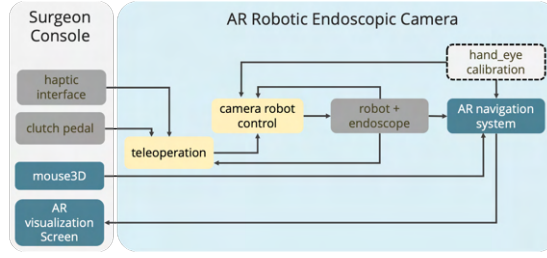


Figure 6.2 Robot-integrated Augmented Reality Navigation System Architecture.

6.2 Material and Methods

The proposed system allows AR guidance while performing a robotic surgical intervention. Whenever the support of an AR navigation system is required, the surgeon can manually perform the registration between a patient-specific virtual model, obtained during the pre-operative planning, and the real organs on the endoscopic image. Once the registration is done, the overlay pose is updated based on the kinematic of the teleoperated robotic arm holding the endoscopic camera. The teleoperation is performed keeping the Remote Center of Motion (RCM) fixed at the endoscope entry point on the abdomen phantom ($\hat{p}_{\{M\}}$). The proposed framework architecture is shown in Fig. 6.2, while further details are reported in the following sub-sections. Fig. 6.3 shows all the reference frames involved in the presented system.

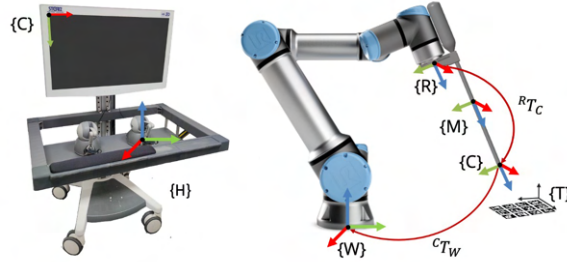


Figure 6.3 Reference frames of all the individual components of the teleoperation console and the robotic endoscope respectively namely: $\{C\}$: camera frame, $\{H\}$: haptic interface frame, $\{R\}$: robot's end-effector, $\{M\}$: RCM, $\{W\}$: robot's base, ${}^R T_C$: transformation between the robot end-effector and the camera link, ${}^C T_W$: transformation between the robot's basis and the camera link and $\{T\}$: the calibration target.

6.2.1 Surgeon remote console

The *Surgeon Remote Console* consists of:

- 6 DoF serial haptic interface to teleoperate the robotic arm holding the endoscope;
- Two pedals: the first used to activate the teleoperation as a clutch, and the other one as a trigger to switch from variable to fixed RCM;
- 32" 4K screen for the visualization of the endoscopic images, with AR if required;
- Mouse3D for the 3D models manipulation in the AR navigation system.

6.2.2 Robotic endoscopic camera

On the patient side, a 6 DoF robotic arm is equipped with a medical-grade endoscopic camera. An adapter to attach the endoscope to the robot end-effector $\{R\}$ was designed and 3D printed, as shown in Fig. 6.4. The pose of the endoscope with respect to $\{R\}$ was chosen to take into account the specific endoscope movements required by a surgical robotic system. In fact, the endoscope motion needs to have 4 DoF, i.e. three translations and the roll rotation around its main axis. Thus, to particularly allow the robot to execute the latter DoF, the endoscope was installed along the outgoing axis of $\{R\}$. A hand-eye calibration method, explained in detail in Sec.6.2.4, is used to compute the correct roto-translation ${}^R T_C$ between R and the camera frame C . This transformation is added to the kinematic chain of the robot, in order to directly control C while teleoperating.

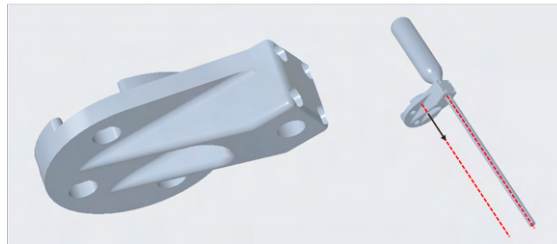


Figure 6.4 On the left, 3D design of the adapter to attach the endoscopic camera to the robot end-effector. On the right, an assembly of the adapter and the endoscope, with their reference systems.

6.2.3 Teleoperation with remote center of motion

To replicate an RMIS scenario, the robotic arm is teleoperated using the 6 DoF haptic interface from the *Surgeon Remote Console*. The increment of position $\mathbf{p}$ of the haptic

interface is used to control $4DoF$ of the endoscope robotic arm accordingly, keeping $\hat{p}_{\{M\}}$ fixed at the insertion point. The teleoperation is performed by mapping the haptic interface reference frame $\{H\}$ to the camera reference frame $\{C\}$ (Fig. 6.3).

To constrain the robotic arm movements in $\hat{p}_{\{M\}}$, this chapter proposes a software-based RCM generation method.

The RCM can be assumed as a fixed point with respect to the world reference frame, i.e. the robot base $\{W\}$; in fact, it coincides with the insertion point in the patient's abdomen. At the beginning of the teleoperation, it is essential that the end-effector frame, i.e. $\{C\}$, is placed at the insertion point $\hat{p}_{\{M\}}$. By pressing a pedal, it will be possible to fix the coordinates of $\hat{p}_{\{M\}}$ with respect to $\{W\}$ for the entire procedure. However, during the insertion of the endoscope into the abdomen, the RCM position changes with respect to $\{C\}$, as $\hat{p}_{\{M\}}$ slides along the endoscope during insertion. Every time an insertion is performed, the distance (l) between $\{M\}$ and $\{C\}$ changes: knowing l is fundamental to compute the correct rotations to apply around the $\{M\}$ in order to have the desired translation in $\{C\}$.

As shown in Fig. 6.5, the rotations around the x and y axes of $\{M\}$ are computed as follows:

$$\theta = \text{atan2}\left(\frac{\Delta x}{l}\right) \quad (6.1)$$

$$\phi = \text{atan2}\left(\frac{\Delta y}{l}\right) \quad (6.2)$$

The rotation around the z -axis of $\{M\}$ (ψ) is computed from the roll variation angle of the haptic interface.

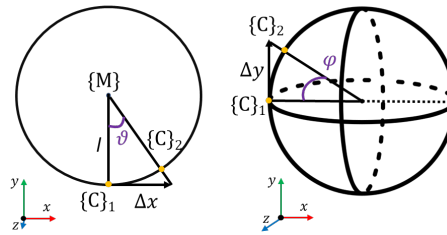


Figure 6.5 2D and 3D illustration of the displacement Δx and Δy applied by the haptic interface. l is the distance between the RCM and end-effector $\{C\}$. θ and ϕ are the resulting angles.

These rotation angles are then converted into a quaternion ΔQ to be applied to $\{C\}$. In order to calculate the new position of $\{C\}$ and to ensure that $\hat{p}_{\{M\}}$ remains fixed, we follow the steps shown in Fig. 6.6: (1) the $\hat{p}_{\{M\}}$ is calculated applying a rigid translation (T_1) of $-l$

along the z-axis of $\{C\}$; (2) the ΔQ is applied to $\{M\}$ to find the new orientation; and (3) another rigid translation (T_2) of l along the new direction of the z-axis is applied.

The 4th DoF is the insertion movement; in fact, a ΔZ motion on the haptic interface frame $\{H\}$ is mapped as a translation on the z-axis of $\{C\}$. While performing this movement, the l is updated to track the correct position of the $\{M\}$ with respect to $\{C\}$.

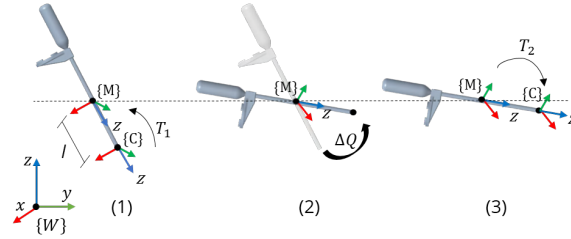


Figure 6.6 Steps to reproduce rotation around RCM. (1) Compute the RCM pose starting from EE pose and applying T_1 (rigid translation). (2) Apply ΔQ to find the new RCM orientation. (3) Find final EE pose applying T_2 (rigid translation) to RCM.

6.2.4 Hand-eye calibration

In order to find the transformation between the robot end-effector reference frame $\{R\}$ and the endoscopic camera frame $\{C\}$, a hand-eye calibration process is performed. Hence, the eye-in-hand calibration process is adopted as provided by Moveit-calibration package⁰ in ROS Noetic. The process involves the acquisition of a number of samples recording: (i) the transformation between $\{C\}$ and a fixed calibration target $\{T\}$ in the camera view and (ii) $\{R\}$ pose. In this work, 32 samples were recorded, varying significantly the robot's pose configuration in the space with respect to an *aruko* marker of 4x3 squares of 10.6mm each and an intermediate distance of 1.2mm.

6.2.5 Augmented Reality navigation system

An AR navigation system was developed to overlay, on the surgical image, the 2D projection of patient-specific 3D virtual organ models. To this end, the main steps required consist in:

- Extraction of patient-specific 3D virtual models during the preoperative planning of the surgical intervention;

⁰https://github.com/ros-planning/moveit2_calibration

- Development of the AR navigation and visualization software to dynamically overlay the 3D virtual models on the endoscopic image, based on the kinematics of the robotic arm;
- Development of the method for the manual registration of the virtual models with the real patient.

Patient-specific 3D virtual models

In the surgical routine, diagnostic medical imaging is performed on the patient before any surgical intervention that requires the knowledge of the shape and position of a target structure. This includes, for example, computed tomography (CT) scan or magnetic resonance imaging (MRI). During the preoperative planning of the intervention, these data are commonly examined by the surgeon to plan the surgical strategy. In the recent years, it is a common practice to exploit these data to extract the 3D surface rendering of organs, vessel structures, and malformations to have a better 3D visualization of the operative field. The proposed system uses these 3D models to create a virtual model of the patient which is used for real-time AR navigation during the surgery. Different software are available to extract the surface rendering from medical imaging (Pieper et al., 2004; Rosset et al., 2004; Udupa et al., 1994). In this work, we used directly 3D models of a patient provided by 3D-IRCADb-02¹. As described in (Penza et al., 2018a), these models were also used to create a corresponding phantom of abdomen, which is used as the *real patient* in the experimental validation.

AR navigation and visualization

The AR navigation system outputs an image representing the 2D virtual models projection overlaid on the endoscopic image. This is achieved by developing a virtual environment where the 3D virtual models are loaded. A virtual camera is placed in the virtual environment to mimic the movement of the robotic endoscopic camera, thus pointing at the models with the same pose of the endoscopic camera. In order to have the same field of view of the endoscopic camera and a correct match between the projected 3D models and the real image, the virtual camera projection matrix is set equal to the matrix of the undistorted endoscopic camera. In addition, the virtual camera *focal point* and the *view up* axis are set equal to the real camera, while the *position* is updated based on the robotic arm movements (taking into account the hand-eye calibration described in section 6.2.4). The endoscopic image is represented behind the 3D models by mapping the undistorted surgical image on a plane parallel to the camera image plane. The transparency of the 3D models can be changed to

¹<http://www.ircad.fr/research/3dircadb/>

visualize the surgical image behind them.

Manual real-virtual patient registration

In order to have a match between the real and virtual patient, an initial rigid registration is needed. Most of the approaches in literature use manual or semi-automatic methods. While the latter exploits physical landmarks on the tissue or 3D surface features, the manual approach requires Human-Computer Interfaces (HCI) to manipulate the models. The choice of the HCI is fundamental to develop an easy-to-use application with a short learning curve. In fact, the way the manual registration is performed can impact the users' physical and mental stress, influencing their final performances. Based on a human-centred study that the authors conducted (Neri et al., 2023), we decided to exploit the manual registration using a 3D mouse. During this registration phase, the endoscopic image is static and the required 3D virtual model appears in the middle of the image. The 3D mouse is used to manipulate the 3D models and adjust its position and orientation in order to match with the real organ in the image. The 3D mouse control is implemented by applying the translation and twist offset as the model's velocity with respect to the camera frame. A scale factor is added to facilitate a precise model manipulation.

6.3 Experimental validation

6.3.1 System architecture

The *Surgeon Remote Console* consisted of a Phantom Omni Haptic Interface (3D Systems, USA), a fullHD 32" 3D Monitor (KARL STORZ, Germany), two USB pedal and the SpaceMouse® (3Dconnexion, Germany). The robotic arm used is a UR5 (Universal Robot, Denmark), while the endoscope is a TIPCAM 1 S 3D LAP, 0deg (KARL STORZ, Germany). The images were captured with an acquisition card (Magewell Pro Capture AIO 4k Lp PCIe X4, 1-channel). The program ran on multiple machine in the same network with GNU/Linux Ubuntu 20.04 operating system. All the code was written in C++, using VTK Library for the development of the virtual environment (Geveci et al., 2012) and with the Robot Operating System (ROS1) as framework (Quigley et al., 2009).

The robotic arm is controlled using velocity commands. These velocities are calculated using the Levenberg-Marquardt inverse kinematics solver available within the Kinematics and Dynamics Library (KDL) (Smits et al., 2011).

6.3.2 Experimental protocol

Experiments were conducted to demonstrate (i) the accuracy of the hand-eye calibration, (ii) the feasibility and the accuracy of the teleoperated robotic camera and (iii) the accuracy of the AR navigation system.

The accuracy of the hand-eye calibration process was computed as the reprojection error $err_{HandEye}$, provided by Moveit-calibration.

Regarding the feasibility and the accuracy of the teleoperated robotic camera, we first assessed the variation of $\hat{p}_{\{M\}}$ in the 3D space with respect to the $\hat{p}_{\{M\}1}$ set at the beginning of the procedure, as explained in Sec 6.2.3. Accordingly, the poses of all the $\{R\}$, the $\{M\}$ and the $\{C\}$ have been recorded during a cross-shape movement along all the directions of the 3D space. Three different insertion depths (i.e. 3cm, 6cm and 9cm) were examined for each experiment. For each one, the user at the *Surgeon Remote Console* was asked to perform a movement of 42cm on the Phantom Omni's x-axis and y-axis, both in the positive and negative direction, resulting in the cross-shape trajectories depicted in Fig. 6.7. The experiment was executed 3 times, starting from 3 different points in the space. The RCM position accuracy (err_{RCM}) was computed as the mean of the differences between each $\hat{p}_{\{M\}i}$ and the initial $\hat{p}_{\{M\}1}$, where i is the frame index ($i = 1..n$), as shown in Eq. 6.3.

$$err_{RCM} = \frac{\sum_{i=1}^n |\hat{p}_{\{M\}i} - \hat{p}_{\{M\}1}|}{n} \quad (6.3)$$

The evaluation of the robot's accuracy in reaching the desired position was done measuring the error in the 3D space between the desired goal pose and actual pose at $\{C\}$. As in the previous experiment, the user at the *Surgeon Remote Console* performed movements towards all the directions of the 3D space for approximately 2 minutes. The mean values and standard deviations of the Cartesian distance ($errPos_{\{C\}}$) and the difference in the Euler angles ($errRoll_{\{C\}}$, $errPitch_{\{C\}}$, $errYaw_{\{C\}}$) of the respective poses were computed.

The accuracy of the AR Navigation system was evaluated by measuring the Intersection over Union (IoU), i.e. the intersection between the mask of the projected 3D virtual model and mask obtained from manually labelled endoscopic image, over their union. A video of 2 minutes was recorded during the teleoperation of the robotic arm and the measurement was performed on 96 images. The evaluation was performed during the execution of different movements (position $errPos_{AR}$ on 67 images, rotation $errRot_{AR}$ on 23 images, zoom $errZoom_{AR}$ on 23 images), as well as the overall accuracy $errAll_{AR}$. The initial manual registration error was separately measured in order to understand the impact of the starting error in the overall accuracy.

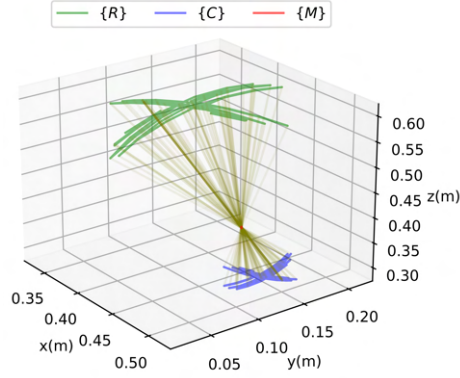


Figure 6.7 Plot of the $\{R\}$ (light green), the $\{M\}$ (red) and the $\{C\}$ (blue) during a cross-shape movement of the Phantom Omni of 42cm around x and y axis and 6cm insertion depth. The dark green lights represent the endoscope length.

6.4 Results

The reprojection error $err_{HandEye}$ of the hand-eye calibration process was calculated as 0.79 pixels.

Fig. 6.7 shows the resulting trajectories of the $\{R\}$ (light green), the $\{M\}$ (red) and the $\{C\}$ (blue) from one experiment of 6cm depth for simplicity (two more depths of 4cm and 8cm were also captured with similar results). The dark green lights can be considered as the endoscope length. As evident, RCM remains fixed during the execution with a positional error (err_{RCM}) derived from Eq. 6.3 of $(8.96 \cdot 10^{-5}, 8.36 \cdot 10^{-5}, 1.06 \cdot 10^{-5})$. Which means that RCM displacement is close to zero while $\{R\}$ and $\{C\}$ poses vary in space.

The errors referred to robot's accuracy in reaching the desired position are reported as: $errPos_{\{C\}} = 0,0011 \pm 0.0006$, $errRoll_{\{C\}} = (0.0013 \pm 0.0241$, $errPitch_{\{C\}} = (0.0009 \pm 0.0139$, and $errYaw_{\{C\}} = 0.0010 \pm 0.0128)$. Therefore, it can be summarized that the calculated position error is close to 1mm while the rotational one gives an error about 0.001degrees with a higher variance.

The computed errors of the AR navigation system are reported in Table 6.1. Note that the initial error represents an important part of the overall error, and that there is a low dependency of the type of movement. An example of the AR visualization during the initial manual registration is shown in Fig.6.8, while Fig. 6.9 shows sample images of the AR guidance visualization during the motion of the robotic endoscopic camera along different axes.

Table 6.1 AR navigation evaluation

	AR Accuracy (IoU)	
	Initial registration	All
$errPos_{AR}$	0.9610	0.9359 ± 0.0262
$errRot_{AR}$	0.9624	0.9232 ± 0.0295
$errZoom_{AR}$	0.9719	0.9459 ± 0.0237
$errAll_{AR}$	0.9651 ± 0.0059	0.9356 ± 0.0271



Figure 6.8 Example of initial manual registration between the 3D virtual organ model and the image of the corresponding kidney phantom. As an example, a representation of a lesion is visible thanks to the transparency of the virtual model.



Figure 6.9 Sample images of the AR guidance visualization in four different positions of the teleoperated robotic endoscopic camera during the experiments. The kidney virtual 3D model remains overlaid with high accuracy on the image of the real organ independently of the position.

6.5 Discussion and Conclusions

This chapter presents a novel integrated system for AR navigation in robotic surgery with a teleoperated robotic endoscope. The system includes a surgeon console composed by a haptic interface, a 3D mouse, a visualization screen and two pedals. The console allows the control of the endoscope attached to the robotic arm, which maintains the RCM constraint. Moreover,

the endoscopic image is augmented by the AR navigation system, which contributes to an enhanced awareness of the surgical scene.

The system was tested to evaluate the accuracy of the calibration, spatial accuracy of the RCM and camera movement, and the real-time tracking of the 3D virtual models during camera motion. Experiments demonstrated promising accuracy results for the system as a whole and for the individual components. The software-defined RCM was shown to remain fixed in space while the endoscopic camera moved to a wide range of different positions.

Regarding the registration method, we implemented an initial manual alignment using a 3D mouse, followed by real-time updates driven by the kinematic data of the robotic arm holding the endoscopic camera. The AR navigation system was able to maintain high overlay accuracy after careful manual registration, indicating that precise camera tracking directly benefits the proposed AR pipeline. This work focuses on rigid registration in a laboratory setup with an abdominal phantom, without modelling organ deformation. Nevertheless, it represents a fundamental first step towards a more advanced system, first by integrating automatic rigid registration and subsequently by incorporating non-rigid organ motion. The accuracy achieved in both registration and camera-motion tracking provides a strong foundation for these future developments.

Part II

Foundations for learning-based registration

Introduction

Motivated by the literature trend toward **deep learning**, showing higher accuracy and robustness than many traditional pipelines, we transition here to learning-based **rigid** registration for surgical point clouds and the resources needed to support them. The goals are: (i) to measure how well state-of-the-art rigid deep-learning methods perform in laparoscopic settings with limited deformation; (ii) to identify when rigidity becomes the limiting factor, guiding the design of non-rigid methods; (iii) to explore and propose training data, and reconstruction strategies, needed to develop and fairly evaluate non-rigid registration methods in the following part.

- *Chapter 7* benchmarks state-of-the-art **complete-to-partial point-cloud** registration methods for laparoscopic AR, using a **preoperative 3D model** and a **partial intra-operative surface** as inputs. It evaluates **traditional** (ICP, CPD, GMMReg) and **deep-learning** approaches (OMNet, OGMM, Leopard, LiverMatch) alongside our Transformer-based correspondence method with a refined overlap module. All the methods were tested across two datasets: a **deformed IRCAD** set (synthetic cases, small deformations i.e. $< 3cm$, varying visibility) and **DePoll** (real cases, large deformations i.e. $> 5cm$). The study establishes performance baselines, evaluates the operating range of rigid deep learning approaches, and highlights conditions under which accuracy degrades. Using these findings, we formulate practical recommendations for building non-rigid methods.
- Building on the benchmarking in Chapter 7, *Chapter 8* addresses a practical bottleneck: the lack of organ-specific point-cloud data to train and evaluate learning-based registration. We construct a kidney dataset of **688** 3D point clouds, covering healthy and tumorous cases, derived from the KiTS (MICCAI, (Heller et al., 2023)) segmentations via a semi-automatic mesh-to-point cloud pipeline. While the dataset provides **complete** kidney surfaces, it pairs naturally with the data processing introduced in Sect. 7.2.3 to generate partial views and deformations, enabling effective training and testing for rigid and non-rigid registration in surgical AR.
- *Chapter 9* investigates intraoperative surface reconstruction using Neural Radiance Fields (NeRF) under realistic surgical constraints. We present a NeRF-based method that, unlike conventional approaches requiring many calibrated intraoperative views, is trained from a **single surgical image** combined with **preoperative MRI**. Preoperative scans are used to predefine camera poses and generate a synthetic multi-view training

set, while an unpaired **neural style transfer** pipeline (WCT2 + STROTSS) transfers the appearance of the intraoperative image onto these preoperative renders. The resulting styled dataset is then used to train an InstantNGP-based NeRF specific to each patient, enabling fast 3D reconstruction and novel-view synthesis from minimal intraoperative data. The method is evaluated on four **neurosurgical** cases, showing good agreement with a multi-view NeRF baseline and with ground-truth microscope images, and thus providing a practical route toward NeRF-based intraoperative 3D/2D registration and AR in the operating room.

Chapter 7

Benchmarking complete-to-Partial point cloud registration techniques for laparoscopic surgery

7.1 Introduction

As discussed in Part I, augmented reality overlays patient-specific 3D anatomical models onto the surgical view to provide intuitive access to preoperative information for tasks such as tumour localisation and vessel avoidance, and has seen growing adoption across neurosurgery, orthopaedics, and laparoscopy with potential efficiency and safety benefits (Bernhardt et al., 2017a; Prasad et al., 2024). However, Sect. 1.2.5 highlighted key challenges arising from soft-tissue motion and deformation, which cause the intraoperative anatomy to differ from the preoperative images acquired before surgery. Aligning preoperative models with intraoperative images during laparoscopy remains a key focus of research, with many challenges still unresolved (Neri et al., 2025b).

Conventional surface-based methods align the preoperative model (surface mesh or point cloud) with intraoperative data using geometric shape information. This process relies on computer vision algorithms to reconstruct the intraoperative 3D surface and typically employs techniques like Iterative Closest Point (ICP) (Besl and McKay, 1992), along with tracking and matching algorithms (Puerto-Souza et al., 2014), to perform the registration. However, it faces challenges such as incomplete reconstructions due to occlusions and lack of distinctive

This work has been published as: Neri, A., Penza, V., Haouchine, N., Mattos, L. S. Benchmarking Complete-to-Partial Point Cloud Registration Techniques for Laparoscopic Surgery. *Frontiers in Robotics and AI*, 12, 1702360. Publicly available at: <https://doi.org/10.3389/frobt.2025.1702360>.

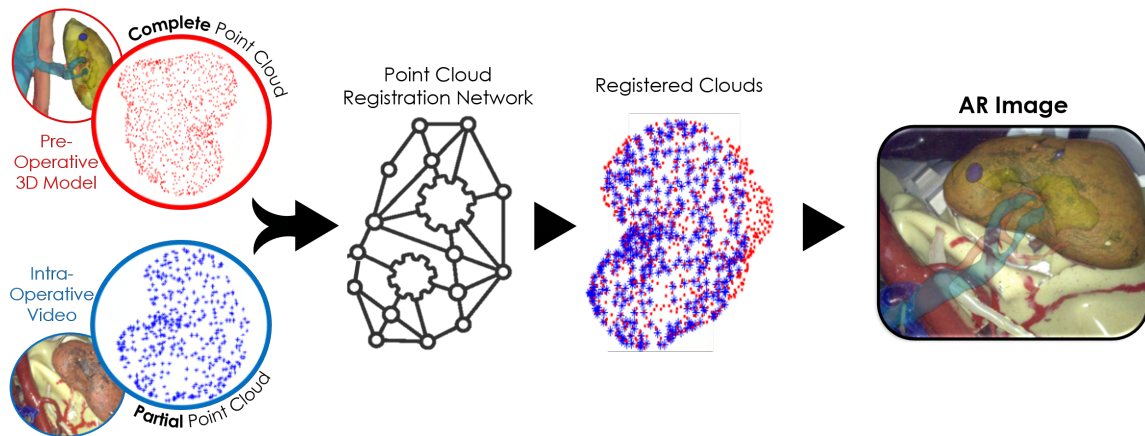


Figure 7.1 DL-based point cloud registration methods use organ point clouds extracted from the preoperative model and intraoperative stereo-video as inputs and estimate the transformation that aligns the two point clouds, allowing the creation of the AR image for surgical navigation.

features, further worsened by complex, texture-less, and deformable scenes (Marques et al., 2015). To address the limitations of traditional surface-based methods, various algorithms are being developed incorporating Deep Learning (DL). One approach involves hybrid DL methods, which enhance conventional surface-based techniques by integrating DL at specific stages, thereby improving registration effectiveness and outcomes. For instance, DL can be applied to tasks such as image segmentation (Zhang et al., 2022), intraoperative surface reconstruction (Luo et al., 2020), or feature extraction (Labrunie et al., 2022). Alternatively, fully DL-based algorithms, such as end-to-end networks for point cloud registration (Huang et al., 2021c), have been employed. These networks take two point clouds as inputs and generate the transformation required to align them (Fig. 7.1).

Recently, a limited number of studies have been published in this last category. They can mainly be divided into two classes: correspondence-free, which do not require the prediction of one-to-one point correspondences, and correspondence-based, which explicitly predict such correspondences.

A correspondence-free approach is (Guan et al., 2023), which introduced the first deep learning-based approach for 3D-3D laparoscopic liver registration. The method builds on OMNet (Xu et al., 2021) and does not utilise Transformers, instead leveraging local and global feature extraction to learn overlapping masks from the preoperative 3D model and the intraoperative reconstruction. These masks are used to filter out non-overlapping regions and standardise the point clouds before aligning the overlapping areas. Thus, this method

directly predicts the final transformation without estimating the point correspondences.

LiverMatch (Yang et al., 2023) is a correspondence-based method to register complete-to-partial synthetic point clouds of liver anatomy. The method consists of a transformer encoder-decoder network that learns feature descriptors, which are then fed to a matching module that predicts point correspondences. The promising results suggest that estimating correspondences between sets of point cloud descriptors leads to good registration results, even in the presence of small deformations.

Dai et al. (Dai et al., 2025) introduce a correspondence-based method that uses a Transformer encoder-decoder architecture. Unlike LiverMatch, which applies the Transformer to encoder features, they employ a geometric Transformer (Qin et al., 2023a) on decoded features and complement it with deep graph matching guided by overlap masks to refine correspondence quality.

Finally, Zhang et al. (Zhang et al., 2024) propose KCR-Net, a correspondence-based method built on an encoder-only Transformer. KCR-Net first extracts keypoint descriptors using a Neighbourhood Feature Fusion Module (NFFM) that employs both self- and cross-attention, and then estimates keypoint correspondences. Unlike (Yang et al., 2023) and (Dai et al., 2025), which recover the final transform from dense correspondences, KCR-Net computes the transformation just from the sparse keypoint matches.

Among these, LiverMatch is the only open-source algorithm.

These approaches are applicable to surgical guidance, since point clouds can be extracted from both preoperative images (e.g., CT scans) and intraoperative images (e.g., stereo cameras), as shown in Fig. 7.1. However, although these point clouds represent the same organ geometry, they differ not only due to deformations, but also because of varying levels of partiality and noise. For instance, the registration to be solved is complete-to-partial; in fact, point clouds derived from CTs (obtained from well-established segmentation techniques (Isensee et al., 2021)) are *complete*, noise-free and dense. In contrast, stereo-camera point clouds are *partial*, capturing only the surface regions visible to the camera ($\approx 30\%$ of the organ (Koo et al., 2022)), and typically noisy, even with the latest 3D reconstruction methods (Zha et al., 2023).

As presented in (Neri et al., 2025b), various end-to-end rigid registration approaches exist, but current evaluations remain restricted mainly to partial-to-partial experiments on classical vision datasets such as ModelNet40 (Wu et al., 2015), which contains rigid, noise-free objects; consequently, little is known about how these methods behave under the conditions that characterise surgical point clouds. To address this gap, we provide the first systematic benchmarking of state-of-the-art registration networks in a surgical scenario, assessing

their robustness and limitations in synthetic intraoperative settings, including complete-to-partial matches, noise and soft-tissue deformation. Among the methods compared, we introduce a refined correspondence-based registration method with an improved overlap-estimation module that yields more accurate correspondences and competitive performance against current baselines. Although all tested algorithms perform rigid registration while the underlying problem is non-rigid, rigid alignment is a useful intermediate step: (i) it brings the two clouds into closer correspondence for a subsequent non-rigid refinement, and (ii) rigid architectures provide a convenient backbone that can be extended to predict dense deformations. Because the solution quality depends strongly on the magnitude of deformation, our evaluation progresses from small deformations across varying levels of partiality to the extreme cases represented in the DePoll dataset (Modrzejewski et al., 2019) (large deformations). Overall, the benchmark isolates the essential building blocks for reliable registration and provides a practical guideline for developing new methods in surgical scenarios.

7.2 Benchmarking protocol

7.2.1 Problem formulation

Let $\mathbf{X} \in \mathbb{R}^{M \times 3}$ be the complete point cloud of the organ of interest (from preoperative planning), and let $\mathbf{Y} \in \mathbb{R}^{N \times 3}$ be a partial point cloud of the same organ (e.g., captured using an endoscopic camera), where $N \ll M$. We define $\mathbf{X}_{\text{visible}} \subset \mathbf{X}$ as the subset of points in $\mathbf{X}$ that correspond to the partial cloud $\mathbf{Y}$, so that $\mathbf{X}_{\text{visible}} \approx \mathbf{Y}$. The goal of point cloud registration is to determine the unknown rigid transformation, composed of a rotation $\mathbf{R} \in \text{SO}(3)$ and a translation $\mathbf{t} \in \mathbb{R}^3$, that aligns $\mathbf{X}$ with $\mathbf{Y}$, i.e., we seek a transformation $\mathcal{T}$ such that $\mathcal{T}(\mathbf{X}_{\text{visible}}) = (\mathbf{R}\mathbf{X}_{\text{visible}} + \mathbf{t}) \approx \mathbf{Y}$.

7.2.2 Competing methods

The following sections present the state-of-the-art open-source methods we evaluated to identify the baseline that best generalizes to real surgical scenarios. To ensure fair comparisons we selected methods according to three criteria: (a) they operate on 3D point clouds (rather than multi-modal pipelines that require image-based tracking or fiducials); (b) their implementations are open-source and can be adapted to the complete-to-partial evaluation setting; and (c) they produce the same output object (a global rigid transformation) so that all methods can be assessed with the same metrics. Accordingly, the following sections first

review traditional (non-deep learning) registration techniques and then cover deep learning approaches. Because publicly available, deep learning complete-to-partial registration methods are scarce (e.g., LiverMatch is an exception), we primarily selected partial-to-partial algorithms. This category is the closest available match to our complete-to-partial scenario and can be adapted to our benchmark under the constraints above. Table 7.1 summarizes all the selected methods.

Table 7.1 Summary of the methods compared

Name	Registration Type	Approach	Main Components
<i>ICP</i>	Rigid	Iterative	Point-to-Point
<i>DCP</i>	Rigid	Iterative, Probabilistic	1 GMM
<i>GMMReg</i>	Rigid	Iterative, Probabilistic	2 GMM
<i>OGMM</i>	Rigid	GMM-Based	Transformers, Overlap Score, 2 GMM
<i>OMNet</i>	Rigid	Correspondence-Free	Overlap Masks, Global Features
<i>Lepard</i>	Matching	Correspondence-Based	Transformers, Repositioning
<i>LiverMatch</i>	Matching	Correspondence-Based	Transformers, Visibility Score
<i>Ours</i>	Rigid	Correspondence-Based	Transformers, Overlap Score

Traditional methods

We classify "Traditional methods" as those that do not rely on deep learning. Among these, we have selected several popular approaches, including Iterative Closest Point (ICP)(Besl and McKay, 1992), Coherent Point Drift (CPD)(Myronenko and Song, 2010), and Gaussian Mixture Models Point Set Registration (GMMReg)(Jian and Vemuri, 2011).

ICP ICP is a widely used rigid registration algorithm that aligns two point clouds by iteratively minimizing the distance between corresponding points. In each iteration, the algorithm identifies the nearest neighbours between the datasets and computes the optimal transformation that reduces the alignment error.

CPD CPD is a probabilistic point cloud registration algorithm that treats one point set as centroids of a Gaussian mixture model while aligning it to the other point set. It enforces smooth motion by assuming nearby points to move coherently, which helps maintain local geometric structure during the transformation.

GMMReg GMMReg is a probabilistic framework representing both input point sets as Gaussian mixture models. In this formulation, the point set registration task is transformed into aligning the two mixtures to minimize a statistical discrepancy measure between them.

Deep learning methods

According to the classification proposed in (Neri et al., 2025b), we selected deep learning point cloud registration methods belonging to different categories such as: correspondence-free, GMM-based, correspondence-based and liver-specific.

Correspondence-free, OMNet We selected OMNet (Xu et al., 2021) to represent correspondence-free deep learning methods. Its core concept involves using overlapping masks to discard non-overlapping points, thereby retaining only the overlapping regions that are most useful for estimating the transformation through global feature regression. Notably, OMNet inspired the work of Guan et al. (Guan et al., 2023), which improved OMNet local feature extraction following the RPMNet (Yew and Lee, 2020) model. Despite that, Guan et al. (Guan et al., 2023) closed-source nature led us to opt for testing OMNet instead.

GMM-based, OGMM OGMM (Mei et al., 2023a) introduces an overlap-guided probabilistic registration approach that estimates the optimal transformation by matching Gaussian Mixture Model parameters. Similarly to GMMReg, the method reformulates registration by aligning two Gaussian mixtures to minimize statistical discrepancies. Additionally, a Transformer-based detection module is employed to identify overlapping regions, using the resulting overlap scores to guide the GMM representation and alignment of the input point clouds.

Correspondence-based, Leopard Leopard (Li and Harada, 2022a) is a learning-based method for partial point cloud matching in rigid and deformable scenes, predicting correspondences that are later registered using ICP or N-ICP. Its architecture combines a fully convolutional feature extractor (KPFCN) with a Transformer employing self- and cross-attention to compute a differentiable similarity matrix. A repositioning module further refines the relative positions between point clouds, enhancing cross-attention and matching effectiveness, which makes Leopard one of the leading methods in non-rigid point cloud registration.

Liver-specific, LiverMatch LiverMatch (Yang et al., 2023) stands out as one of the few open-source deep-learning approaches designed explicitly for surgical laparoscopic registration. It employs an encoder-decoder architecture enriched with self- and cross-attention mechanisms to extract point features that are then used to compute a similarity matrix. This matrix, in combination with a visibility score, is utilized to predict correspondences between

the two point clouds. Following correspondence determination, registration is carried out using ICP. To train the network to handle deformations, the authors generated a synthetic dataset by applying deformations and cropping techniques to 16 livers from the 3D-IRCADb-01 dataset (Soler et al., 2010).

Refined RegTR (Ours) We extend RegTR (Yew and Lee, 2022) with a refined overlap-estimation module designed to enhance the accuracy in predicting the final transformation (Fig. 7.2). RegTR leverages keypoint features, which are fundamental since correspondences are determined among these keypoints rather than using all points. A KPConv backbone is employed to extract a reduced set of keypoints ($\mathbf{K}_X \in \mathbb{R}^{M' \times 3}$, $\mathbf{K}_Y \in \mathbb{R}^{N' \times 3}$) and their associated features ($\mathbf{F}_{K_X} \in \mathbb{R}^{M' \times D}$, $\mathbf{F}_{K_Y} \in \mathbb{R}^{N' \times D}$) from the input clouds, which are then projected to a lower dimension (256) and enriched with sinusoidal positional encoding (PE).

We form position-augmented tokens as $\tilde{\mathbf{F}}_{K_X} = \mathbf{F}_{K_X} + \text{PE}(\mathbf{K}_X)$ for the complete cloud and $\tilde{\mathbf{F}}_{K_Y} = \mathbf{F}_{K_Y} + \text{PE}(\mathbf{K}_Y)$ for the partial cloud. In the *self-attention* blocks, each cloud is processed independently and the same token set is used as query, key, and value (i.e., $Q = K = V = \tilde{\mathbf{F}}_{K_X}$ and $Q = K = V = \tilde{\mathbf{F}}_{K_Y}$), enabling intra-cloud contextual feature aggregation among keypoints. In the *cross-attention* blocks, we adopt a bidirectional conditioning scheme across clouds: complete-to-partial attention uses $Q = \tilde{\mathbf{F}}_{K_X}$ with $K = V = \tilde{\mathbf{F}}_{K_Y}$, while partial-to-complete attention uses $Q = \tilde{\mathbf{F}}_{K_Y}$ with $K = V = \tilde{\mathbf{F}}_{K_X}$. This setup allows each keypoint descriptor to be updated by retrieving a weighted combination of features from the opposite cloud, while remaining explicitly position-aware through the added sinusoidal encoding.

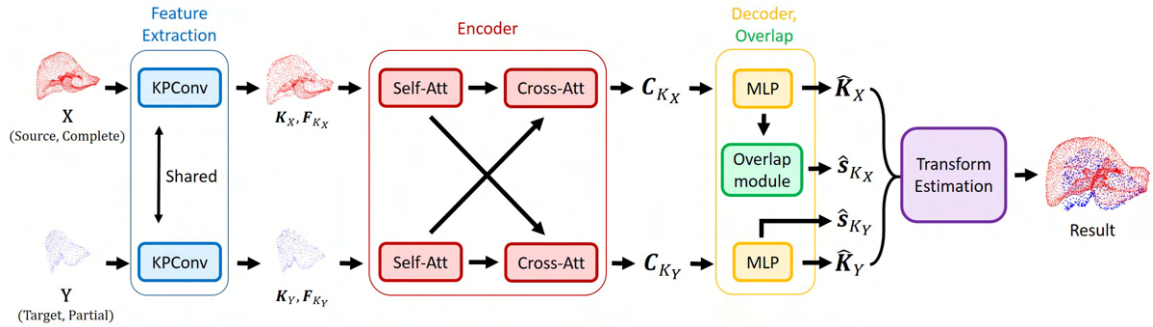


Figure 7.2 The network leverages KPConv to extract keypoints and their associated features. These features are further refined using self- and cross-attention mechanisms and then passed to the decoder, which predicts overlap scores and the corresponding keypoints. This information is ultimately used to estimate the final transformation.

The conditioned features are fed into a two-layer MLP to predict the transformed keypoint coordinates ($\hat{\mathbf{K}}_X \in \mathbb{R}^{M' \times 3}$ and $\hat{\mathbf{K}}_Y \in \mathbb{R}^{N' \times 3}$).

Self- and cross-attention produce conditioned features ($\mathbf{C}_{K_X} \in \mathbb{R}^{M' \times d}$ and $\mathbf{C}_{K_Y} \in \mathbb{R}^{N' \times d}$) that effectively identify accurate correspondences and filter outliers, serving a role similar to RANSAC in traditional approaches.

Overlap module. In parallel with the decoder, overlap scores for the two point clouds are predicted. The overlap score, denoted as $\hat{\mathbf{s}} = [\hat{\mathbf{s}}_{K_X}, \hat{\mathbf{s}}_{K_Y}]$, represents the probability that a keypoint belongs to the overlap region. Unlike (Yew and Lee, 2022), where overlap scores are predicted for both point clouds, our method predicts the overlap score only for the complete point cloud $\mathbf{X}$. Indeed, since the problem we are solving is a complete-to-partial registration, we know by definition that the partial point cloud is fully contained within the complete one ($\mathbf{X}_{\text{visible}} \subset \mathbf{X}$ and $\mathbf{Y} \approx \mathbf{X}_{\text{visible}}$). Consequently, every point in $\mathbf{Y}$ belongs to the overlap area, and its overlap score is 1. Therefore, predicting overlap scores for the partial cloud provides no additional information but does introduce an extra source of estimator error: imperfect predictions on $\mathbf{Y}$ can produce false negatives or false positives that harm the overall performance.

To predict which points of the complete model are observed in the partial scan, we apply a linear fully-connected layer with an elementwise sigmoid activation to produce per-point overlap scores as follows:

$$\hat{\mathbf{s}} = \begin{cases} \hat{\mathbf{s}}_{K_X} = 1 / (1 + e^{-(\mathbf{C}_{K_X} \mathbf{w}_1 + \mathbf{b}_1)}) \\ \hat{\mathbf{s}}_{K_Y} = \mathbf{1}_{N'} = (1, 1, \dots, 1)^\top \in \mathbb{R}^{N' \times 1} \end{cases} \quad (7.1)$$

where $\mathbf{w}_1$ and $\mathbf{b}_1$ are learnable weights and biases parameters and $\mathbf{C}_{K_X}$ are the conditioned features. $\hat{\mathbf{s}}_{K_Y} = \mathbf{1}_{N'}$ ensures that all points in the partial cloud are part of the overlap region (i.e. overlap score = 1).

Transformation prediction. Finally, the predicted transformed keypoint coordinates are concatenated to form correspondence pairs. Unlike methods such as Lepard and Liver-Match, which rely on similarity matrices and matching losses, RegTR directly predicts the transformed coordinates and treats them as correspondences for final transform estimation; therefore, it does not construct a similarity matrix or depend on correspondence supervision. The rigid transformation is estimated by leveraging the correspondences and overlap scores and minimizing the weighted sum of squared distances between the corresponding points. We solved it using a weighted variant (Gojcic et al., 2020) of the Kabsch-Umeyama algorithm (Umeyama, 1991).

Losses and Optimization. Our method employs a weighted sum of three losses similar to (Yew and Lee, 2022): (i) a registration loss that minimizes the error between the predicted transformed keypoint positions and their ground truth counterparts, weighted by the overlap

confidence; (ii) a conditioned feature loss that encourages the network to consider geometric properties and to distinguish correct correspondences from incorrect ones in the context of feature matching; and (iii) an overlap loss designed to optimize the overlap scores, which measure the confidence that a keypoint from $\mathbf{X}$ has a valid correspondence in the overlapping region of $\mathbf{Y}$. Early stopping is applied: training terminates if the validation loss does not improve for 12 consecutive epochs. On an NVIDIA Tesla V100 GPU, the full training run takes approximately 5 hours.

7.2.3 Dataset and pre-processing

We employed two datasets with different deformation magnitudes to assess how rigid registration algorithms generalize to non-rigid scenarios. First, we generated a customized complete-to-partial version of the 3D-IRCADb-01 dataset, incorporating small random deformations, noise, and varying levels of partiality. To further stress-test the algorithms, we also used the DePoll dataset, which features large deformations, irregular noise, and severe partiality.

Deformed IRCAD

The original IRCAD dataset (Soler et al., 2010) consists of 3D CT scans from 10 women and 10 men, with hepatic tumors present in 75% of the cases. The dataset also contains the VTK models of each liver. For each VTK, we extracted the point cloud that describes its surface and sampled 3500 points, which provided a balance between geometric detail and GPU memory constraints. The points were normalized in the range $[-1, 1]$ across all three axes, producing 20 different source point clouds. Subsequently, inspired by the approaches proposed by Livermatch (Yang et al., 2023) and (Dai et al., 2025), we generate several corresponding partial, deformed targets for each source point cloud. Differently from them, our pipeline employs the As-Rigid-As-Possible (ARAP) deformation algorithm from Open3D (Sorkine and Alexa, 2007), which lets us deform the liver mesh via user-defined control and anchor points (Neri et al., 2025a). Specifically, we randomly choose one of the two liver lobes and apply a random translation of up to $\pm 25\text{mm}$ along both the x and z axes. This range, corresponding to 2.5 cm in real scale, was selected empirically to simulate small yet realistic liver deformations (Fig. 7.3); considering that an adult liver measures roughly 20 cm in width (Jones J, 2009).

To simulate realistic endoscopic views, we place a virtual camera aimed at the organ’s visible surface. We sample random camera positions in spherical coordinates, constraining

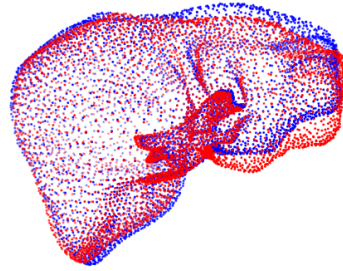


Figure 7.3 An example of synthetic deformation applied to the complete preoperative liver (red), where the right lobe is shifted upward (blue). The final intraoperative cloud is then cropped from the deformed model (blue).

polar and azimuthal angles to match typical intraoperative ranges. For each point on the surface, we compute the dot product between the camera direction and the point normal, retaining only 15% of points whose angle to the camera is less than 80° (empirically found). Finally, we apply a random rigid transformation to the target, with a rotation up to 45° and a translation in the $[-50\text{mm}, 50\text{mm}]$ range. We add element-wise Gaussian noise with zero mean and a standard deviation of 0.01 to each point in every dimension. Regarding seed management, a random seed is assigned to each training pair. This seed controls the generation of all stochastic parameters, including the selected lobe to deform, the magnitude of its displacement, the camera viewpoint, the applied rigid transformation, and the noise level.

For the training set, we used 17 livers (i.e. livers numbered 3 through 20) as source point clouds and generated 560 partial target clouds for each, yielding a total of 10,080 pairs. Using the same workflow, we created 4 testsets based on the remaining 2 livers (i.e. liver 1 and 2). Each set includes 50 partial targets per source (100 pairs total). The variation arises from the crop ratio: 5%, 10%, 15%, or 25%, chosen to mimic different levels of intraoperative organ exposure, and from the random deformations generated as in training. The selected levels of partiality reflect the surgical context, where the algorithm must handle varying exposure of the organ to provide AR guidance. In practice, surgeons typically expose only about 20–30% of the organ surface during a procedure (Benincasa et al., 2008; Koo et al., 2022); we therefore included lower partiality levels ($<20\%$) to stress-test the robustness of the evaluated algorithms under particularly challenging visibility conditions. Partiality below 20% poses challenges due to insufficient discriminative features in the intraoperative point cloud (Benincasa et al., 2008). Although surgeons generally expose as much of the organ as possible before resection, visibility rarely exceeds 50% of the surface area, as the opposite side remains occluded.

Because these datasets are synthetically generated, we retain complete ground-truth annotations, including the applied rigid transformations, known point correspondences, and overlap scores. Finally, to enable the computation of the Target Registration Error (TRE), we choose n landmark points on the original (source) mesh and identify their exact correspondences on the deformed (target) mesh. To ensure these points are not part of the input clouds provided to the registration algorithm, we select them from the cropped-out regions, applying farthest point sampling.

Code and data to reproduce our deformed IRCAD dataset are available at: https://github.com/Alberto-Neri/Laparoscopic_Organ_Deformation_wARAP.

DePoll

DePoll (Deformable Porcine Laparoscopic Liver) dataset (Modrzejewski et al., 2019) comprises preoperative and intraoperative pig liver surface point cloud data under different deformation states. Specifically, it includes a preoperative point cloud of the pig liver, which is complete and derived from a CT scan. Regarding the intraoperative data, there are 13 cases of the same liver under various deformation conditions. Each case contains a partial point cloud extracted from an intraoperative CT scan and a partial point cloud obtained from video reconstruction. Fig. 7.4(a), 7.4(b) and 7.4(c) show one representative case, displaying the complete preoperative liver point cloud alongside two corresponding intraoperative partial reconstructions—one from CT and one from video. To normalize the points in the range $[-1, 1]$, we applied min-max normalization using the maximum and minimum values of the complete preoperative point cloud. The dataset authors obtained the ground truth registration using pre- and intraoperative markers and the point clouds are provided pre-registered. Moreover, since the data come from two different sensors (i.e., CT scan and endoscope), they already exhibit noise and density variations. For this reason, the only pre-processing we applied was generating a random rigid transformation within the range proposed for IRCAD dataset.

7.2.4 Metrics

To evaluate the rigid registration quality, we compute the mean absolute error ($\text{MAE}(\mathbf{R})$, $\text{MAE}(\mathbf{t})$) between the predicted and the ground truth values for both the rotation angle and translation. All methods directly estimate a rigid transformation, except Lepard and LiverMatch, which output point-to-point correspondences. For these two, we recover the rigid pose by feeding their correspondences into Open3D’s RANSAC-ICP routine, as proposed

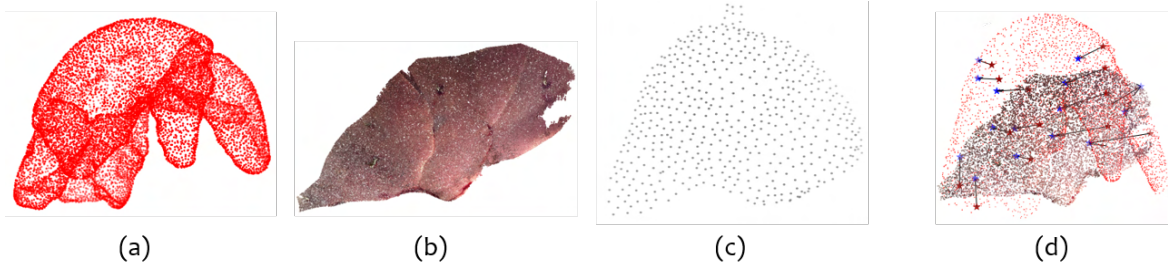


Figure 7.4 (a) The complete preoperative pig liver point cloud with segmented lobes. (b) The intraoperative video reconstructed point cloud (partial), relative to case 1. (c) The intraoperative CT scan point cloud (partial), relative to case 1. (d) Point cloud (a) and (b) registered with the ground truth rigid transformation. The dark red stars represent the surface marker on the preoperative anatomy; the blue stars represent the same surface markers on the intraoperative anatomy (after non-rigid deformations). The black arrows show the displacements between the corresponding markers, representing the effect of non-rigid deformation. The GT TRE is computed as the sum of the distances indicated by the black arrows. For simplicity, the figure shows only surface markers and the two clouds with a lower resolution.

by LiverMatch. We set the *max_correspondence_distance* parameter to 0.05, producing the best alignment results.

We also report the TRE (mm) for each experiment. For the IRCAD dataset, we use the landmark coordinates identified during preprocessing (see Sect. 7.2.3), while for DePoll we rely on the preoperative and intraoperative markers provided by the dataset authors.

7.2.5 Experimental setup

ICP was run with an identity initialization, a convergence threshold of 0.001, and a maximum of 30 iterations. CPD was configured with the same threshold and up to 50 iterations. For GMMReg, the number of Gaussian components was set equal to the number of points in the target cloud.

All deep learning models were trained on our deformed IRCAD dataset (15% crop ratio), with minor code adjustments and hyperparameter tuning to ensure optimal convergence. For OMNet, which expects two partial point cloud of equal size, we padded the smaller target clouds to preserve its input structure and replaced all BatchNorm layers with GroupNorm to stabilize training with our small batch sizes. We also substituted the authors' overlap score estimation with our ground-truth overlap scores, which are better suited to deformed data. Each of these adjustments led to significantly improved convergence. Similar modifications were applied to OGMM, yielding performance gains but still less optimal convergence; in this case,

we opted not to introduce further changes to respect the original design. Finally, we improved Lepard and LiverMatch’s convergence by setting the hyperparameter *first_subsampling_dl* to 0.03.

All algorithms were evaluated on our four IRCAD deformed test sets, each defined by a different crop ratio (5%, 10%, 15%, and 25%), to mimic varying levels of intraoperative organ exposure. We also tested each algorithm on the DePoll dataset for further generalization evaluation on large deformations.

7.3 Results

7.3.1 Evaluation on the IRCAD deformed dataset

Fig. 7.5 presents the algorithms’ qualitative results at varying partiality levels.

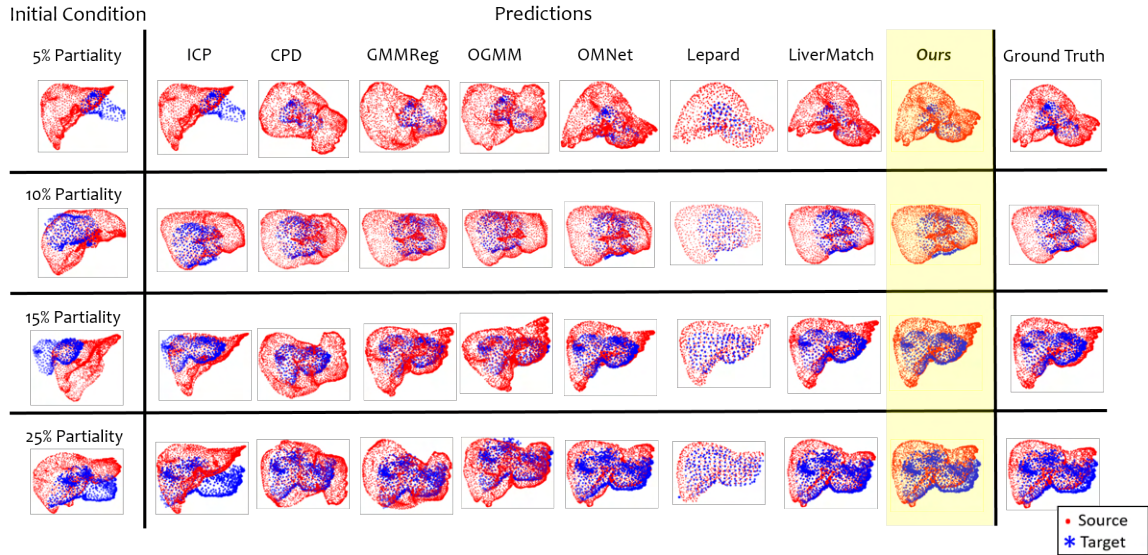


Figure 7.5 Each model’s qualitative results were evaluated under four distinct levels of partiality (i.e., 5%, 10%, 15% and 25%). The first column illustrates the initial conditions (i.e. after pre-processing), the following shows the predicted registration, and the last provides the ground truth registration.

Table 7.2 shows the performance (mean $\pm$ std) of the algorithms on the IRCAD deformed test set at the four partiality levels analyzed. This table does not include results from the original RegTR, as they are marginally worse than those of the proposed method (e.g., at 15% partiality: $TRE = 8.60mm \pm 2.51mm$, $MAE(R) = 2.87^\circ \pm 1.49^\circ$, $MAE(t) =$

$2.97mm \pm 1.26mm$; at 10% partiality: $TRE = 13.41mm \pm 6.70mm$, $MAE(R) = 3.98^\circ \pm 2.62^\circ$, $MAE(t) = 4.94mm \pm 2.10mm$).

Table 7.2 Performance Metrics at different levels of partiality. Legend: **best**, second best

	5% Partiality			10% Partiality			15% Partiality			25% Partiality		
Algorithm	TRE (mm)	MAE(R) (°)	MAE(t) (mm)	TRE (mm)	MAE(R) (°)	MAE(t) (mm)	TRE (mm)	MAE(R) (°)	MAE(t) (mm)	TRE (mm)	MAE(R) (°)	MAE(t) (mm)
ICP	62.69 ± 27.79	22.60 ± 8.21	21.34 ± 7.11	62.19 ± 28.44	24.03 ± 8.76	19.95 ± 7.37	66.15 ± 29.51	25.30 ± 8.30	22.56 ± 6.94	65.22 ± 28.61	25.61 ± 6.85	21.69 ± 6.42
CPD	80.05 ± 30.55	43.09 ± 21.93	19.43 ± 6.68	79.52 ± 32.70	46.62 ± 24.12	20.82 ± 5.64	78.44 ± 34.78	47.09 ± 24.52	22.56 ± 6.07	66.33 ± 39.27	38.15 ± 27.39	20.82 ± 7.72
GMMReg	67.51 ± 29.66	21.21 ± 9.38	27.50 ± 7.03	66.36 ± 35.55	21.95 ± 11.29	26.89 ± 8.68	65.62 ± 36.07	21.28 ± 13.05	26.03 ± 8.07	55.72 ± 33.22	19.62 ± 13.09	21.69 ± 7.98
OGMM	82.55 ± 34.07	21.99 ± 9.65	32.97 ± 5.90	73.60 ± 27.76	16.54 ± 6.63	29.50 ± 5.38	65.40 ± 24.83	14.51 ± 6.61	26.89 ± 5.38	31.14 ± 17.91	9.03 ± 3.57	10.41 ± 5.55
OMNet	23.49 ± 14.14	9.00 ± 4.33	7.20 ± 3.64	15.23 ± 8.36	5.77 ± 2.58	5.29 ± 2.17	11.72 ± 6.45	4.37 ± 1.95	3.82 ± 1.91	10.45 ± 4.81	3.86 ± 1.76	3.73 ± 1.82
Lepard	27.03 ± 39.44	10.94 ± 18.04	10.58 ± 15.79	<u>9.82 ± 8.59</u>	3.91 ± 2.30	4.86 ± 3.04	<u>7.45 ± 5.49</u>	3.21 ± 1.95	<u>3.64 ± 2.08</u>	7.41 ± 6.69	3.27 ± 1.68	3.04 ± 2.26
LiverMatch	18.69 ± 28.58	<u>7.47 ± 6.09</u>	8.41 ± 10.67	7.91 ± 5.90	3.64 ± 2.03	<u>4.68 ± 2.17</u>	6.39 ± 4.54	<u>3.08 ± 1.91</u>	3.73 ± 2.00	4.81 ± 3.21	<u>2.78 ± 1.52</u>	2.78 ± 1.47
Ours	<u>22.36 ± 12.04</u>	6.60 ± 4.14	<u>7.72 ± 4.77</u>	13.01 ± 6.07	<u>3.79 ± 2.25</u>	4.25 ± 2.00	8.49 ± 2.62	2.85 ± 1.32	2.60 ± 1.13	<u>6.78 ± 2.08</u>	2.33 ± 1.06	<u>2.86 ± 0.87</u>
GT	6.45 ± 5.62	0.00 ± 0.00	0.00 ± 0.00	6.17 ± 5.18	0.00 ± 0.00	0.00 ± 0.00	5.37 ± 4.82	0.00 ± 0.00	0.00 ± 0.00	5.01 ± 4.40	0.00 ± 0.00	0.00 ± 0.00

The last table row describes the ground truth error, which for TRE represents the residual non-rigid component. In particular, the TRE remains nonzero because it is measured between preoperative markers and their intraoperative counterparts, which have undergone both a rigid transformation and non-rigid deformation. Since all evaluated algorithms perform solely rigid registration, they inherently cannot correct for the non-rigid component, resulting in a residual error consistent with this ground-truth baseline (same concept of Fig. 7.4(d) representing DePoll dataset). While TRE may not fully capture performance under our scenario, it remains the primary metric in surgical applications. Therefore, we include TRE results for all rigid registration methods to transparently evaluate their performance and highlight their inherent limitations under these challenging conditions.

We first analyze the 15% partiality test set, which uses the same level of partiality as the training data but with different liver anatomies. In this scenario, our method achieves the third-lowest TRE ($8.49mm \pm 2.62mm$) yet remains highly competitive, as evidenced by its lower standard deviation and best performance across other rigid registration metrics (i.e: $MAE(R) = 2.58^\circ \pm 1.32^\circ$ and $MAE(t) = 2.60mm \pm 1.13mm$).

Across the remaining partiality levels, models must generalize not only to new liver anatomies but also to varying crop ratios. Our method remains highly robust, consistently ranking first or second, while conventional rigid techniques struggle with the complete-to-partial registration challenge. In particular, GMMReg at 25% partiality shows $TRE = 55.72mm \pm 33.22mm$, $MAE(R) = 19.62^\circ \pm 13.09^\circ$ and $MAE(t) = 21.69mm \pm 7.98mm$. Even deep learning partial-to-partial approaches exhibit limited generalizability, with OGMM showing the greatest performance drop (at 25% partiality: $TRE = 31.14mm \pm 17.91mm$, $MAE(R) = 9.03^\circ \pm 3.57^\circ$ and $MAE(t) = 10.41mm \pm 5.55mm$). All methods degrade at low visibility (small target area), whereas higher partiality (e.g., 25%) yields noticeably

better registration accuracy. For instance, at 25% partiality, our method achieves $TRE = 6.78mm \pm 2.08mm$, $MAE(R) = 2.33^\circ \pm 1.06^\circ$ and $MAE(t) = 2.86mm \pm 0.87mm$.

Examining the qualitative results in Fig. 7.5, we observe that some predictions closely match the ground truth; however, the models often fail to capture non-rigid deformations and instead bias the target's contours to match the preoperative edges.

7.3.2 Evaluation on the DePoll dataset

To further assess generalization, we evaluated each model across all 13 cases in the DePoll dataset using the same weights trained on the deformed IRCAD data. Differently from previous studies that register the intraoperative CT-derived point cloud to the video-based reconstruction ((Guan et al., 2023), (Zhang et al., 2024), (Dai et al., 2025)), we perform complete-to-partial registration by treating the preoperative CT-derived liver model as the "complete" source and the video-based reconstruction as the "partial" target. The critical difference is the source cloud: the intraoperative CT scans deliver a partial and non-rigidly deformed liver model (Fig. 7.4(c)), whereas the preoperative model (Fig. 7.4(a)) describes the full organ before any deformation. Although these conditions are more demanding, we believe they more accurately reflect surgical practice, where a complete preoperative 3D model is registered to a partial, intraoperative surface reconstruction.

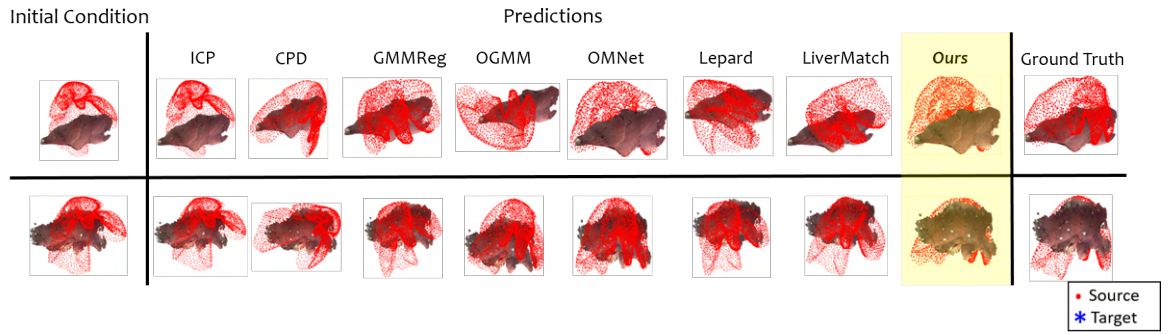


Figure 7.6 Qualitative registration results on the DePoll dataset, each row represents a different case. The inputs used were the preoperative complete point cloud and the intraoperative partial point cloud reconstructed from the video.

Fig. 7.6 shows some qualitative registration results produced by each algorithm tested. Table 7.3 presents the quantitative results as the mean $\pm$ std on the 13 test cases.

Although our approach outperforms existing methods, its performance remains insufficient to overcome this challenge entirely. With a $MAE(R)$ of $14.66^\circ \pm 5.52^\circ$ and $MAE(t)$ of $12.22mm \pm 5.52mm$, residual misalignments are frequently large enough to be visually

Table 7.3 Average Performance Metrics on DePoll dataset. Legend: **best**, second best

Algorithm	TRE (mm)	MAE(R) (°)	MAE(t) (mm)
ICP	60.16 ± 28.96	24.09 ± 7.78	21.61 ± 6.31
CPD	58.09 ± 26.81	25.47 ± 7.90	16.48 ± 6.70
GMMReg	56.23 ± 28.55	21.41 ± 5.45	14.98 ± 7.10
OGMM	68.57 ± 38.05	28.61 ± 22.55	27.20 ± 12.54
OMNet	<u>42.85 ± 25.65</u>	<u>16.05 ± 5.68</u>	11.28 ± 3.94
Lepard	72.47 ± 32.06	24.09 ± 7.78	22.08 ± 8.44
LiverMatch	54.47 ± 28.54	17.87 ± 6.19	17.11 ± 8.44
Ours	42.10 ± 23.74	14.66 ± 5.52	<u>12.22 ± 5.52</u>
GT	29.58 ± 22.38	0.00 ± 0.00	0.00 ± 0.00

perceptible. The first row of Fig. 7.6 shows one of the best examples: both our method and OMNet approximate the ground truth, although a noticeable rotational offset remains. This case also highlights how intraoperative anatomy can differ substantially from the preoperative scan: achieving the target alignment here requires a large leftward displacement of all three liver lobes. In contrast, the second row depicts only minor lobe deformation, leading to better registration accuracy.

7.3.3 Discussion

The motivation for this benchmark is to transparently assess how far rigid registration methods can address the inherently non-rigid challenges of the surgical scenario. This evaluation not only highlights the strengths and limitations of rigid approaches but also establishes a robust guideline for developing more advanced non-rigid algorithms to fill the current gaps in surgical applications.

On the deformed IRCAD dataset, generic computer-vision algorithms stay competitive when trained on small deformations. They also remain robust under various partiality conditions, but performance degrades when partiality is excessive; for example, at 5%, the target point cloud may not contain enough information to perform the registration. Thus, partiality plays a critical role in overall robustness. However, by incorporating self- and cross-attention modules with overlap/visibility scoring, the network can explicitly identify and weigh corresponding regions between the two clouds, making this combination particularly effective for handling variations in point cloud visibility. In contrast, OGMM and traditional registration methods struggle to align complete-to-partial point clouds in low-overlap scenarios. In our current setup, training uses fixed source/target cardinalities. When evaluating on substantially different target sizes, the effective candidate set in self-

and cross-attention changes, which can shift attention sharpness and the conditioned feature distribution consumed by the overlap/visibility head, potentially degrading performance. A straightforward way to improve robustness is to train with variable point-count augmentation (e.g., random subsampling/upsampling), so that the attention modules and overlap head see a broader range of cardinalities during training.

In the DePoll registration task, the required generalization level exceeds the current algorithms' capabilities. This challenge arises from large deformations, anatomical differences, and significant noise. In our experiments, the models trained for rigid registration on deformed data still failed under DePoll conditions. The extreme magnitude of non-rigid deformations severely impairs the models' ability to establish correct correspondences and achieve accurate registration, demonstrating that current algorithms cannot solve this complete-to-partial task. Moreover, the pig liver's three-lobe structure and irregular reconstruction noise differ from the human training data and increase the task's difficulty.

Overall, our results show that when deformations are limited and the exposed intraoperative surface covers more than 10% of the organ, TREs below 10 mm can be achieved, approaching the 5 mm accuracy typically required in surgery (Doornbos et al., 2024). Moreover, when the exposed surface increases to 25%, the TRE decreases to approximately 5 mm, reaching an acceptable level for surgical practice. Conversely, in the presence of large deformations, rigid registration alone cannot achieve low TREs. However, it remains valuable as an initialization step to bring the two point clouds into closer alignment before non-rigid refinement, or as a backbone architecture that can be extended to deformation prediction.

Both our proposed solution and LiverMatch provide a solid foundation for addressing the challenge of large deformations. These architectures predict correspondences between point clouds, which can subsequently be leveraged by external algorithms (e.g., N-ICP (Amberg et al., 2007)) to perform non-rigid registration. However, their current limitation lies in the low quality of the estimated correspondences, which hinders accurate deformation prediction. For example, in our IRCAD experiments with 10% partiality, LiverMatch achieved 37% recall and 51% precision, meaning that out of 525 ground-truth correspondences, the algorithm predicts on average 380 matches, but only 194 of these are correct. In this case, it is crucial to maximize recall and precision because incorrect correspondences severely degrade the registration. To achieve this, future work should focus on improving the modules responsible for correspondence estimation, such as exploring novel feature-processing or extraction strategies. One promising direction could be to leverage semantic cues (e.g., colour information) within the point clouds to generate more discriminative and accurate matches.

An alternative strategy to improve inference performance is patient-specific training. Current state-of-the-art methods typically rely on an agnostic approach, training on large heterogeneous datasets to generalize to unseen cases. However, this paradigm may not be optimal in surgical contexts, where anatomical variability between patients might be substantial. Since preoperative CT images are routinely acquired, synthetic datasets of deformations for each patient’s organ could be generated and used to train or fine-tune the network. In this way, the model is trained and tested on the same anatomy, reducing the burden of inter-patient generalization and requiring it to adapt only to intraoperative factors such as deformation and noise.

7.4 Conclusion

In this chapter, we have benchmarked deep-learning and traditional approaches for point cloud registration, offering a transparent assessment of their generalizability to real-world surgical applications. The approach involves registering two input point clouds: a complete one extracted from a preoperative 3D organ model (derived from CT or MRI scans) and a partial one reconstructed from the intraoperative stereoscopic video. Our comparison covered GMM-based, correspondence-based, correspondence-free, matching-based, and liver-specific methods, aiming to identify the shared modules that lead to top performance.

Secondly, we leveraged the backbone of one of the state-of-the-art partial-to-partial registration models, and we implemented its complete-to-partial version by incorporating the estimation of overlap points only for the complete point cloud, which led to improved performance. We included our algorithm in the benchmark, demonstrating its competitive results under different partialities and deformations.

To stress the experimental setup, we intentionally applied rigid registration algorithms to scenarios in which the underlying anatomy may undergo non-rigid tissue deformations, to evaluate how well rigid approaches generalize beyond their modelling assumptions. Remarkably, despite this mismatch, deep learning-based rigid registration methods remain robust across a wide range of partiality levels when deformations are small. We attribute this resilience to combining self- and cross-attention modules with overlap scoring. However, these same methods struggle to handle large, non-rigid deformations, as the DePoll experiment shows. Addressing such cases will require developing non-rigid registration algorithms that retain the effective components of rigid models while enhancing correspondence estimation to accommodate more extreme deformations.

Chapter 8

Kidney point clouds dataset from KITS segmentation challenge

8.1 Introduction

Deep learning has revolutionized computer vision in recent years, yielding state-of-the-art performance in feature extraction, segmentation and image classification tasks. One of the reasons for its popularity is the wide availability of data in various fields, which are used to train deep learning models and allow them to learn complex patterns and representations. However, in the medical field, comprehensive datasets are not always available. Medical data often require expert annotation, such as identifying specific pathologies or anatomical structures, and this process can be time-consuming and expensive. Moreover, some medical conditions are rare, leading to small sample sizes for training datasets. The scarcity of data poses a challenge in developing robust deep-learning models capable of generalizing effectively across diverse patient populations.

The same issue arises when moving from images to 3D shapes and point clouds. Although there are some datasets of standard shapes (ModelNet40 (Wu et al., 2015), Stanford 3D (Curless and Levoy, 1996)) and datasets dedicated to autonomous driving (KITTI (Geiger et al., 2012)), we have identified a scarcity of datasets associated with human organs point cloud. For example, 3D-IRCADb-01 (Soler et al., 2010) provides a dataset of 20 livers with tumours, while, to the best of our knowledge, the only existing dataset offering 3D models of kidneys is 3D-IRCADb-02, but it only contains 4 kidneys.

Motivated by this gap, this chapter focuses on generating a dataset of 3D organ models and point clouds. Such models are usually derived from segmented CT or MRI scans and thus rely on expert annotations to delineate the relevant anatomy.

As highlighted by Neri et al. (2025b), deep learning methods applied to point cloud registration can be an excellent solution to the problem of augmented reality in surgery. These end-to-end networks are trained on object point clouds and learn how to register them to obtain a precise overlap. With the dataset proposed in this chapter, we aim to resolve this gap, allowing researchers to train or fine-tune their models on kidney point clouds, increasing their performance and promoting transferability in the surgical augmented reality domain. Our dataset derives directly from the segmentation images provided during the Kidney and Kidney Tumor Segmentation Challenge (KITS23) of MICCAI23 (Heller et al., 2021, 2023). The dataset contains 489 CT scans of patients who underwent cryoablation, partial nephrectomy, or radical nephrectomy for suspected renal malignancy. The main contribution lies in designing and developing a semi-automatic pipeline (Fig. 8.1) to process these CT scans to reconstruct the 3D models associated with the kidneys and then extract the point cloud that describes their surface. Our dataset comprises 688 kidney models, including those with tumours ($\approx 50\%$).

8.2 Methods

In medical practice, 3D models are frequently reconstructed from preoperative imaging modalities such as CT scans or MRI. For instance, during a CT scan, a sequence of two-dimensional images (slices) of the organ is obtained. By segmenting the region of interest and combining the information from each slice, it is possible to reconstruct the 3D organ. Physicians undertake the process of segmenting and potentially annotating tumours on each slide, which consumes significant time. For this reason, we decided to leverage the efforts already made in the past by utilizing the KITS dataset, which is publicly available and includes all the kidney segmentations proposed in the MICCAI 2019, 2021 and 2023 KITS challenges. The annotation team consisted of attending radiologists, urologic oncologists, medical students, and undergraduates planning to study medicine. The dataset comprises 489 cases from various patients, with many segmentations available for both kidneys. It is important to specify that KITS contains both "annotations" and "segmentations". In particular, the annotations refer to the raw vectorized interactions the user generates during an annotation session. A "segmentation," on the other hand, refers to the rasterized output of a postprocessing script that uses "annotations" to define regions of interest.

Starting from these KiTS segmentations, we then designed our processing pipeline to obtain high-quality 3D kidney models and corresponding point clouds. Figure 8.1 illustrates this pipeline, from the input segmentations to the final curated dataset.

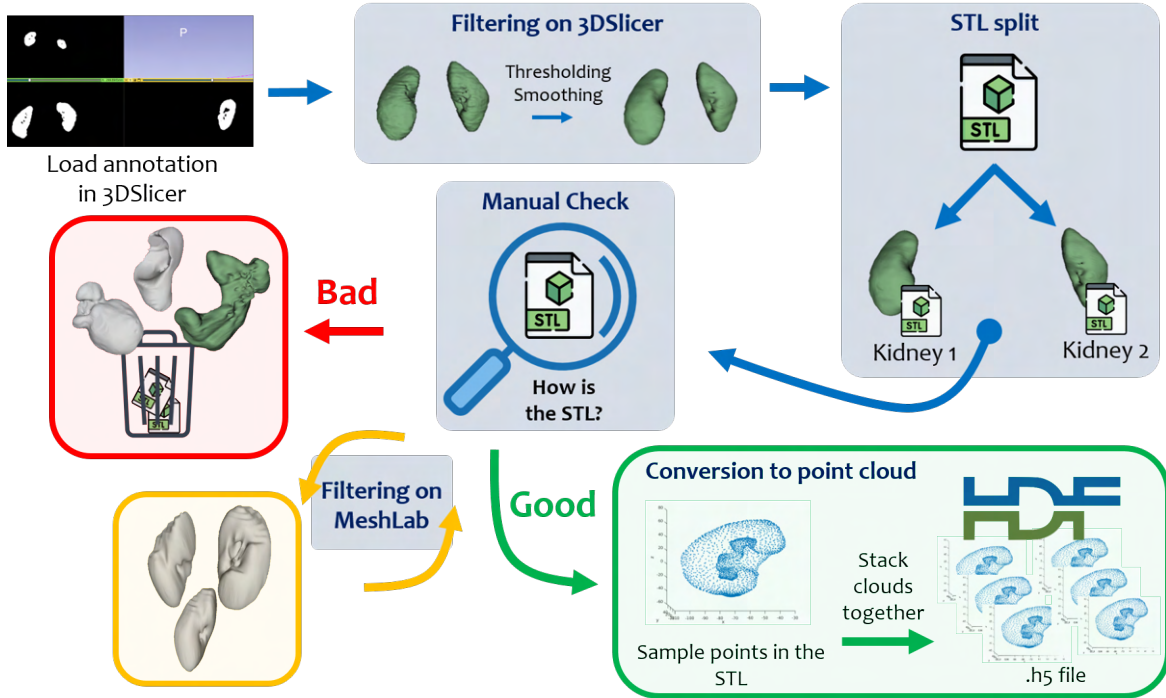


Figure 8.1 The semi-automatic pipeline proposed in this work. The only manual step is the Manual Check, during which the user must inspect the STL models to determine if they are good, bad or need further filtering. All the other steps are automated.

8.2.1 Filtering on 3DSlicer

The first step is loading the segmentations into 3DSlicer (Fedorov et al., 2012; Website, 2024), a widely used, free and open-source software for visualization, processing, segmentation, and analysis of medical images. Given the abundance of segmentations, manually processing each one would be time-consuming. Therefore, we have developed a Python macro to automate this task. The code iterates through each segmentation in the dataset, applying two filters to each; the output is a folder containing an STL file for each segmentation originally present.

Fig. 8.2 shows the salient stages of the macro. For each segmentation, we first apply an empirically chosen threshold filter in the range 1.10–2, identified by sweeping candidate thresholds on a subset of volumes and visually assessing kidney coverage and background suppression. Considering the coarse segmentation and variations between slices, relying solely on this filter may not be adequate, as the surface obtained is wrinkled and non-uniform. Therefore, a second smoothing filter is applied to refine the surface. We selected the Joint Taubin filter with a factor of 0.5, selected as the smallest value that consistently removed high-frequency artefacts without visibly distorting the overall kidney shape. The resulting

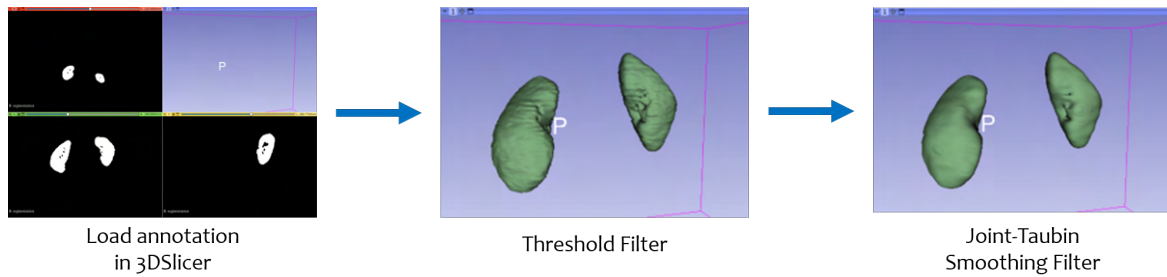


Figure 8.2 Pipeline of the 3DSlicer macro. Initially, the segmentation is loaded into the environment. Then, a threshold filter is applied, but the surface is still wrinkled and irregular. Thus, a second smoothing filter is applied. Finally, the 3D models are saved as an STL file. The pipeline is iterated over every segmentation in the original dataset.

3D models are saved as STL files, and the 3DSlicer environment is cleaned up to allow the code to process subsequent scans.

8.2.2 STL split

Fig. 8.2 shows a segmentation featuring two kidneys; however, in some instances, physicians have segmented only one kidney. Consequently, the macro output may yield STL files containing either one or two kidney 3D models. In this dataset, we focus on individual kidney point clouds; therefore, separating the STL files containing two kidneys becomes necessary. This step is performed automatically through a Python code that uses the VTK library. Specifically, the code cycles through all the STLs generated by the 3DSlicer macro and applies the splitting logic to each one (Fig. 8.3). This logic mimics the manual procedure: if two kidneys are present, a separating cut is defined between them. In practice, we implement this cut as a vertical plane through the origin that partitions the STL into two components. Then, to understand whether the original STL contained one or two kidneys, we calculate the cartesian position of the split centroids. If the distance between the two centroids is greater than an empirical *threshold*, then there are two kidneys, otherwise there is only one. The output of this step is a set of STLs, each containing the 3D model of a kidney. In general, this logic allows the splitting of most STLs effectively, although there are cases where it fails. The most frequent case is when one of the two kidneys has a massive tumour ($>$ double the size of the kidney). In these cases the vertical plane intersects the tumorous kidney and consequently cuts it. Although this only happened 11 times with this dataset, this situation must be handled manually with other software such as MeshLab, which will be discussed in subsequent sections.

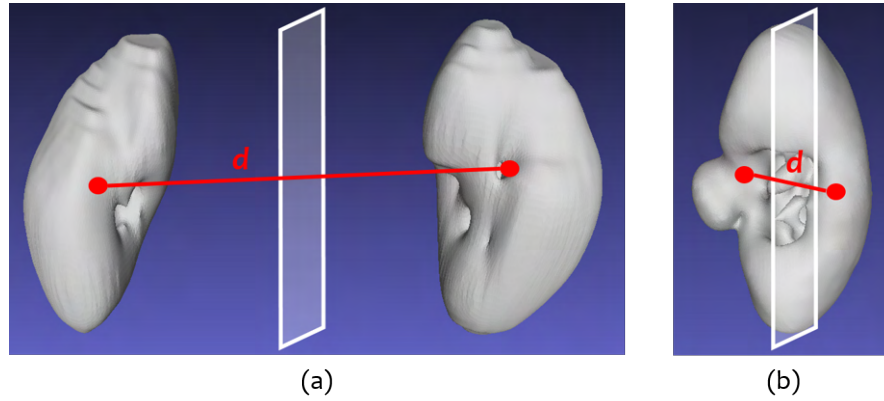


Figure 8.3 STL split logic. In both cases, a vertical plane is drawn at the origin, which serves as a cut region to split the STL. We proceed by calculating the spatial coordinates of the centroids of the two splits and the distance d between them. (a) If $d > threshold$, the splits are saved in two separate STLs. (b) If $d < threshold$, we have only one model and save it in a single STL. The threshold was found empirically.

8.2.3 Manual check

After splitting the STLs, a manual check is required. This operation requires human intervention as the user needs to evaluate the issue mentioned in the previous section and whether the kidney's shape is acceptable, refinable, or needs to be discarded. We performed this task on all the 978 3D models resulting from the split, using a code to plot the STLs and then moving on to visual analysis. We have divided the kidneys into three categories:

- Good. These kidneys have a shape similar to the healthy kidney; furthermore, their surface is smooth and has no significant irregularities, making the model realistic.
- Bad (Fig. 8.4a). They exhibit significantly different shapes compared to typical kidneys. This disparity might arise from either the tumour's dimensions or insufficiently precise segmentation. These kidneys are definitively discarded as their potential inclusion in the training set could adversely impact the model's learning process.
- Refinable (Fig. 8.4b). In these cases, the kidneys resemble the typical kidney shape, yet their surface appears irregular and wrinkled. Upon observing these models, one notable characteristic is the presence of a staircase-like pattern, possibly attributed to coarse segmentations or a low resolution of the original CT scan. Thus, an additional smoothing filter may be advantageous to refine the surface texture.

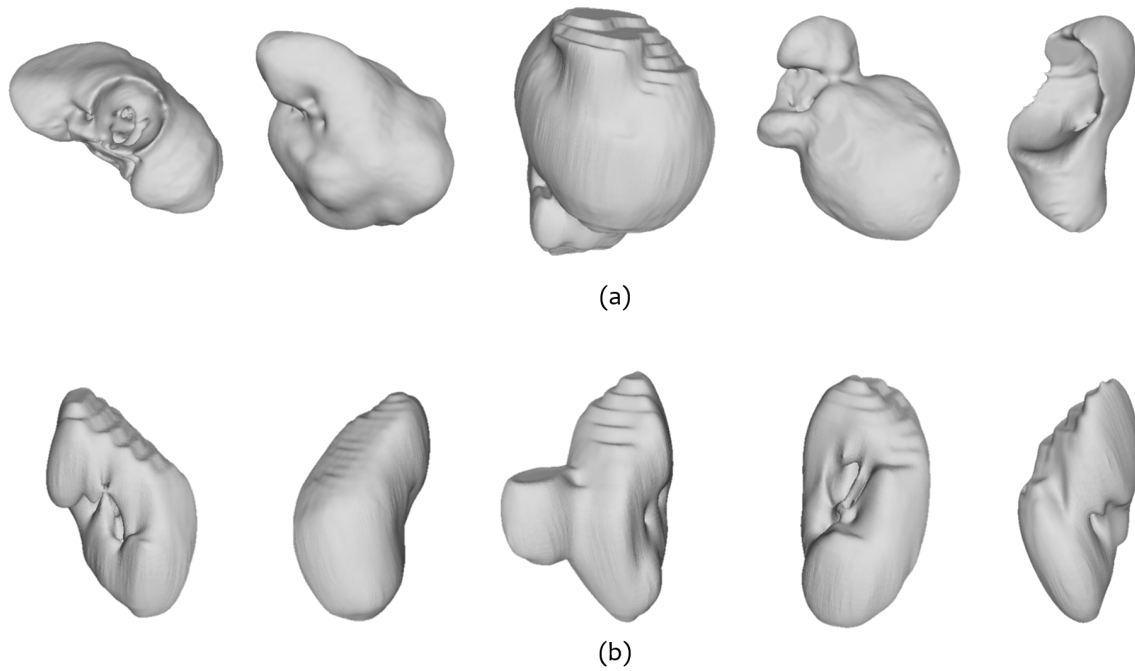


Figure 8.4 (a) Examples of discarded kidneys. It includes kidneys with a very irregular shape due to massive tumours, kidneys in which the tumour is missing, and kidneys resulting from poor segmentation. (b) Examples of refinable kidneys. They are promising as their shape resembles traditional kidneys, but we need to improve their surface quality through an additional smoothing filter. Staircase-like pattern are visible.

Filtering on MeshLab

We can automate the application of the second smoothing filter. To accomplish this, we have created a Python script utilizing the PyMeshLab library (Muntoni and Cignoni, 2021), which enables the utilization of MeshLab (Cignoni et al., 2008), an open source system for processing and editing 3D triangular meshes. In particular, our code applies the Taubin filter (Taubin, 1995) to each STL with the following parameters: $\lambda=0.9$, $\mu=-0.01$, $\text{stepsSmoothNum}=20$ (empirically chosen). Then, it saves the new model obtained. The filter is very similar to the one applied in the second step of the 3DSlicer macro, in fact the result is comparable to that shown in Fig. 8.2. After this filtering, the user must manually recheck the models as previously described. Some models may require further smoothing, while others may have lost too much information and therefore should be discarded.

8.2.4 Conversion to point cloud

Once the final list of acceptable STLs is obtained, the user can proceed to the last step, which involves extracting the surface point cloud from each model. This process is automated using Python and the open3D library (Zhou et al., 2018). Specifically, it utilizes the *sample_points_poisson_disk* function (Yuksel, 2015), which samples N points from the mesh, ensuring that each point is approximately equidistant from its neighbours. Subsequently, the extracted points are normalized with min-max normalization so that each coordinate is between -1 and +1. Normalizing the dataset is crucial to ensure stable training of deep-learning models and to prevent certain samples from dominating the optimization due to larger magnitude values. The point clouds obtained are inserted into the final dataset, which can be partitioned into training, validation and test sets.

8.3 Conclusion

In this chapter, we present a semi-automatic pipeline that converts kidney segmentations into 3D models. Using the kidney segmentations provided in the KiTS (MICCAI) challenge, we generate a dataset of 688 kidney 3D models, including both healthy and tumorous cases. The pipeline relies on filters from 3D Slicer and MeshLab to smooth the reconstructed kidneys while preserving their anatomical shape.

The process is semi-automatic because it requires a manual quality check after the first filtering step. Starting from 978 initial kidney STLs, we visually inspected each model to assess whether the filter removed surface artefacts without losing important details or introducing distortions. As described in Sect. 8.2.3, kidneys were classified as good, bad, or refinable. In total, 393 kidneys were refiltered and rechecked; only a subset of these satisfied the acceptance criteria, while 108 still showed irregular shapes after the second filter or had been over-smoothed. Overall, 276 out of the 978 kidneys were discarded, and 21 STLs required additional manual processing due to an incorrect initial split. The final set of 688 models meets the following requirements: realistic kidney shape, smooth surface, absence of holes, and no major irregularities.

The dataset includes both STL meshes and derived point clouds. The meshes can, for example, be used to generate deformed complete-to-partial datasets, following the procedure described in Chapter 7 Sect. 7.2.3, which enables the training and evaluation of rigid and non-rigid point-cloud registration methods.

Chapter 9

Surgical neural radiance fields from one image

9.1 Introduction

Neural Radiance Fields (NeRF) (Mildenhall et al., 2021) have emerged as robust techniques for 3D view synthesis, reconstruction and registration, gaining traction in surgical applications (Fehrentz et al., 2024; Wang et al., 2022a). While advancements like InstantNGP (Müller et al., 2022) enable rapid NeRF generation, compatible with surgical guidance requirements, they rely on acquiring multiple images from diverse viewpoints during surgery—a process prone to user-dependent variability and potential errors. In this chapter, we propose to generate a NeRF using a single surgical image.

Single-image NeRF has been addressed with different approaches (Gu et al., 2023; Jain et al., 2021; Xu et al., 2022a; Yu et al., 2021a) outside of surgery. For instance, PixelNeRF (Yu et al., 2021a) employs a ResNet-34 encoder and a feature-conditioned NeRF trained on a multi-view dataset of similar scenes to enable single- or few-shot view synthesis. Since PixelNeRF is trained on ShapeNet (Chang et al., 2015), it excels at synthesizing images of objects within similar categories but struggles to generalize to surgical environments. DietNeRF (Jain et al., 2021) uses prior knowledge from a pre-trained image encoder (CLIP) to guide the NeRF optimization process in the few-shot setting. In their single-view setting, they fine-tune PixelNeRF synthesis by augmenting the reconstruction loss with a semantic

This work has been first accepted at **IPCAI 2025** conference and later **extended and published** as: Neri, A., Fehrentz, M., Penza, V., Mattos, L. S., Haouchine, N. (2025). Surgical neural radiance fields from one image. *International Journal of Computer Assisted Radiology and Surgery*, 1-5. Publicly available at: <https://doi.org/10.1007/s11548-025-03447-5>. (This work was carried out during the **abroad period** at *Brigham and Women's Hospital and Harvard Medical School*, in Boston, Massachusetts.)

consistency loss derived from the pre-trained CLIP. This additional supervision enables realistic novel view synthesis and plausible completion of unobserved regions even with few input images. NerfDiff (Gu et al., 2023) is also trained on large datasets containing multiple scenes, each with at least two views. During training, it jointly optimizes an image-conditioned NeRF and a 3D-aware conditional diffusion model (CDM) across a collection of scenes. However, at test time, it fine-tunes the NeRF using a single input image by generating and refining virtual views with the CDM. Finally, SinNeRF (Xu et al., 2022a) is the only model reconstructing scenes from a single RGB and depth input without multi-view pre-training. The authors apply geometric and semantic supervision to enable realistic renderings of unseen views. Depth information from the reference view is propagated to other viewpoints through image warping, generating geometry pseudo-labels that enforce multi-view consistency. Additionally, a pre-trained Visual Transformer network ensures appearance consistency across views through semantic supervision. Overall, these methods suffer from higher computational overhead than InstantNGP. Additionally, apart from SinNeRF, which depends on depth information, the others rely on shape priors from large datasets of common objects, limiting their ability to generalize to surgical scenes.

Contribution. We propose a novel approach illustrated in Fig. 9.1 that requires only a single intraoperative image to train a NeRF. Our method leverages preoperative data to define the set of camera viewpoints and images needed for robust and unobstructed training. Intraoperatively, the appearance of the surgical image is transferred to the pre-constructed training set, enabling instant and fast NeRF training. We present preliminary results from four neurosurgery cases and demonstrate the effectiveness of the proposed approach in synthesizing novel surgical views.

9.2 Methods

Problem formulation

We aim at training a neural radiance field F_θ , where θ are the learned parameters, that maps spatial positions $\mathbf{x} \in \mathbb{R}^3$ and viewing directions $\mathbf{d} \in \mathbb{S}^2$ to volumetric density $\sigma(\mathbf{x})$ and radiance $\mathbf{c}(\mathbf{x}, \mathbf{d})$ so that $F(\mathbf{x}, \mathbf{d}) = (\sigma(\mathbf{x}), \mathbf{c}(\mathbf{x}, \mathbf{d}))$. In order to train F_θ , a dataset of intraoperative poses and images $\{(\mathbf{P}_i, \mathbf{J}_i)\}_{i=1}^N$ of N samples is required. We want to circumvent this requirement and only use a single intraoperative image $\mathbf{J}$ instead of a multi-view set $\{\mathbf{J}_i\}_{i=1}^N$, thus lifting the intraoperative burden from surgeon. To achieve this we create a training dataset from preoperative scans. Using volume rendering to capture brain surface

geometry, and by targeting the area where the craniotomy will most likely be placed, we can sample a dataset $\{(\mathbf{P}_i, \mathbf{I}_i)\}_{i=1}^N$ of poses and images. Because NeRFs are patient-specific, this dataset contains the geometry needed to learn the volumetric density $\sigma(\mathbf{x})$; however, it does not contain the intraoperative appearance to properly train the radiance $\mathbf{c}(\mathbf{x}, \mathbf{d})$. To this end, we define an unpaired image-to-image transfer function that transfers the intraoperative appearance from $\mathbf{J}$ to the preoperative images dataset $\{\mathbf{I}_i\}_i$. Our pipeline is illustrated in Fig. 9.1.

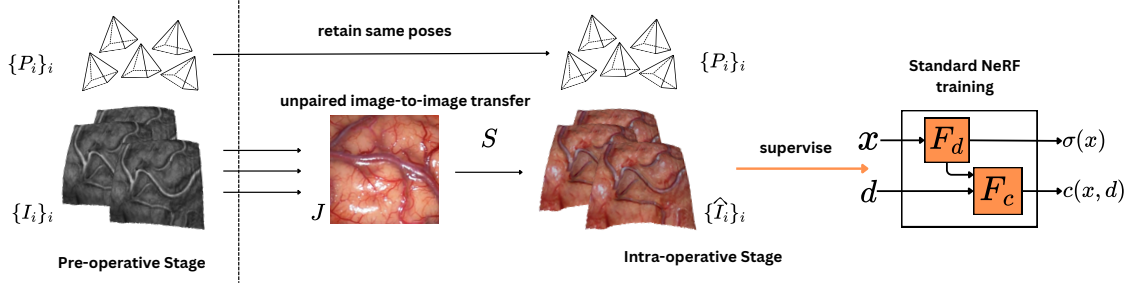


Figure 9.1 Method overview: the key aspect of our approach is to retain the preoperative poses $\{\mathbf{P}_i\}_i$ and transfer the intraoperative appearance $\mathbf{J}$ to the preoperative images $\{\mathbf{I}_i\}_i$ using S , an unpaired image-to-image translation. This process allows for the supervision of the training of a NeRF F_θ using only one image, circumventing the error-prone multi-view intraoperative acquisition.

15

Intraoperative to preoperative appearance transfer

Given the intraoperative image $\mathbf{J}$, we want to transfer its appearance to $\{\mathbf{I}_i\}_i$ while retaining the camera poses. We use Neural Style Transfer (NST) to perform this transformation. We opted for NST instead of adversarial or variational methods for its deterministic/controlable outputs, fast inference and reduced data requirements. Because the preoperative data and intraoperative image are unregistered we need an unpaired image-to-image translation. In unpaired situations, traditional NST methods do not consistently preserve geometrical consistency and are designed for stylistic rendering at the cost of semantic fidelity. We combined WCT²(Yoo et al., 2019), a wavelet-corrected transfer method that preserves the semantic content of images while maintaining photorealism, with STROTSS(Kolkin et al., 2019), which utilizes relaxed optimal transport and self-similarity for stylization with controllable guidance. This hybrid approach results in a more balanced style transfer that avoids over-stylization and is performed in two stages. First, we generate an intermediate image, $\hat{\mathbf{I}}_{\text{int}}$, using Haar wavelet pooling and whitening/coloring transforms applied in the VGG feature space while minimizing the WCT² loss, $\mathcal{L}_{\text{WCT}^2}$. This stage ensures content

preservation and edge detail retention. Then, we apply STROTSS to refine $\hat{\mathbf{I}}_{\text{int}}$ for enhanced style fidelity using the loss function $\mathcal{L}_{\text{STROTSS}}$, which primarily relies on a relaxed Earth Mover’s Distance to match feature statistics and color distributions, leading to a more cohesive and realistic result. The full generation can be formulated as follows:

$$\hat{\mathbf{I}} = \arg \min_{\hat{\mathbf{I}}} \mathcal{L}_{\text{STROTSS}}(\hat{\mathbf{I}}, \mathbf{J}), \quad \text{where} \quad \hat{\mathbf{I}}_{\text{int}} = \arg \min_{\gamma} \mathcal{L}_{\text{WCT}^2}(\mathbf{I}, \mathbf{J}, \gamma) \quad (9.1)$$

This two-stage process is defined by the function S that generates a new image $\hat{\mathbf{I}}$ given the intraoperative image $\mathbf{J}$ and preoperative image $\mathbf{I}$.

NeRF optimization and training

In order to train F_{θ} , we can express the problem so that the rendered pixel intensities approximate the target pixel intensities of the intraoperative image. This amounts to optimizing the following loss:

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{|\Omega|} \sum_{\mathbf{p} \in \Omega} \|S(\mathbf{I}_i, \mathbf{J})(\mathbf{p}) - \mathbf{I}_{\text{rendered}}(\mathbf{p}; \mathbf{P}_i, F_{\theta})\|^2 \quad (9.2)$$

where $\mathbf{p}$ is 2D pixel coordinates in the image $\hat{\mathbf{I}}_i$ and $\mathbf{I}_{\text{rendered}}(\mathbf{p}; \mathbf{P}_i, F_{\theta})$ are the color value predicted for pixel $\mathbf{p}$. We can notice here that we only use a single image $\mathbf{J}$ and that we use the pre-defined set of poses $\{\mathbf{P}_i\}_i$, removing the need for multiple intraoperative acquisition.

To meet the real-time intraoperative requirement, we chose to train NeRF using Instant-NGP (Müller et al., 2022) for computational efficiency. It leverages a multi-resolution hash grid encoding to represent the 3D scene efficiently, mapping spatial coordinates and viewing directions to a compact latent space. This latent representation is then passed to a small neural network to predict the color and density values. Intraoperatively, the training lasts approximately 150s, while the style transfer takes around 30s. The generation of preoperative poses and images, which occurs in the preoperative phase, takes about 5 minutes. Both style transfer and pose generation can be further optimized.

Step-by-step pipeline summary

To make the overall workflow explicit, we summarize the proposed single-image NeRF pipeline as a sequence of concrete steps, highlighting inputs/outputs at each stage.

Inputs. We assume (i) a patient-specific preoperative volume (T1 MRI in our experiments) and (ii) a single intraoperative microscope image acquired during surgery.

1. **Preoperative view synthesis and pose sampling.** Starting from the preoperative volume, we generate a set of rendered images by sampling a collection of plausible camera poses around the expected surgical viewpoint. This produces a synthetic multi-view dataset of images and their associated camera poses.
2. **Single-image style transfer (unpaired appearance matching).** Since the rendered preoperative views look different from the intraoperative image (e.g., in color, contrast, and texture), we adapt their appearance using the single intraoperative frame as a reference. Specifically, we apply WCT² followed by STROTSS to each rendering, producing a stylized multi-view set that resembles the intraoperative image while preserving the underlying anatomical structure and the original camera poses.
3. **NeRF training using stylized supervision.** We train an Instant-NGP NeRF using the fixed camera poses (from step 1) and the corresponding stylized targets (from step 3).
4. **Inference.** After optimization, the trained NeRF can be queried at arbitrary camera poses to synthesize novel views that (i) follow the intraoperative appearance and (ii) remain geometrically consistent with the preoperative patient model.

9.3 Experiments and Results

Dataset and metrics. We evaluated our method on 4 clinical neurosurgical cases, each consisting of a preoperative T1 MRI scan and a corresponding surgical microscope image. Each microscopic image was acquired intraoperatively using a standard clinical microscope setup. For each case, we generated 100 images, along with their respective camera poses from the preoperative T1 MRI. We use volume rendering to obtain the visualization of the surface of the brain. Then, guided by the surgeon’s input on the craniotomy area, we retained only the most relevant portion of the preoperative surface. Thus, we selected camera poses to mimic intraoperative viewpoints realistically.

We used the following performance metrics: the Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) for assessing image quality; the Learned Perceptual Image Patch Similarity (LPIPS), commonly used for evaluating synthesis methods; and the Gram Matrix Score (GMS), which measures texture similarity and is widely used in NST approaches.

Agreement with multi-view NeRF. We trained a NeRF using our method and measured its synthesis agreement with a Multiview NeRF (MV-NeRF) trained on real surgical microscope

images. For evaluation, we synthesized 9 images from random poses to account for the small observed brain region (50 mm diameter). This approach simulates intraoperative viewpoint variability and ensures out-of-distribution testing by avoiding overlap with training poses. Table 9.1 shows the average results across all poses for each case.

Table 9.1 Average synthesise agreement with MV-NeRF across all poses.

	SSIM	PSNR	LPIPS	GMS
<i>Case 1</i>	0.78 ± 0.02	30.26 ± 0.88	0.34 ± 0.06	0.10 ± 0.04
<i>Case 2</i>	0.79 ± 0.03	31.71 ± 0.89	0.30 ± 0.06	0.15 ± 0.04
<i>Case 3</i>	0.74 ± 0.03	32.51 ± 0.70	0.23 ± 0.04	0.28 ± 0.07
<i>Case 4</i>	0.79 ± 0.04	33.39 ± 0.85	0.20 ± 0.04	0.13 ± 0.04

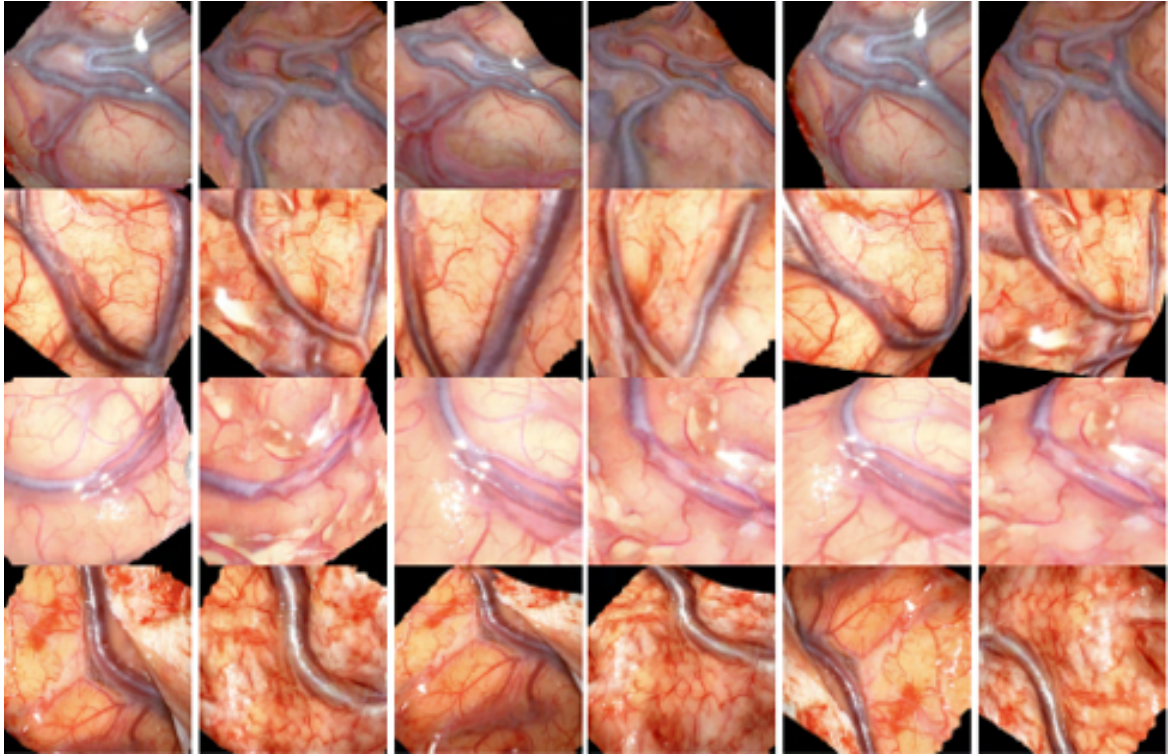


Figure 9.2 Tests on four real cases, one per row with pairs of synthesized images from three random viewpoints using MV-NeRF and our method, respectively.

All experiments achieved SSIM scores above 0.74 and PSNR values greater than 30dB, indicating that our method consistently maintains the structural integrity of the images. The LPIPS values varied between 0.20 and 0.34, suggesting that our method achieves a good balance between perceptual similarity and image quality. Lastly, except for *Case 3*, the GMS values remained below 0.15, indicating that the generated images closely match the

MV-NeRF regarding textural representation. We provide qualitative results with a visual assessment in Fig. 9.2.

Comparison against ground-truth. To compare with the ground-truth surgical image, we manually registered the preoperative MRI’s brain surface with the real surgical image to obtain a ground-truth camera position and orientation. We used this information to synthesize one image per case using our methods and MV-NeRF and compare them with the real surgical image. The plots in Fig. 9.3 show the results for each case. In Cases 1 and 2, SSIM exceeds 0.70, and in Cases 3 and 4 it remains above 0.50, likely due to mis-registration. Overall, the SSIM difference between MV-NeRF and our method is less than 10%. PSNR is consistently around 30dB in all cases, suggesting good reconstruction quality. Although LPIPS indicates some deviations due to the unpaired nature of the NST, GMS confirms that our synthesized images faithfully preserve style and texture—with scores remaining below 0.15, except for Case 3, consistent with the results shown in Table 9.1.

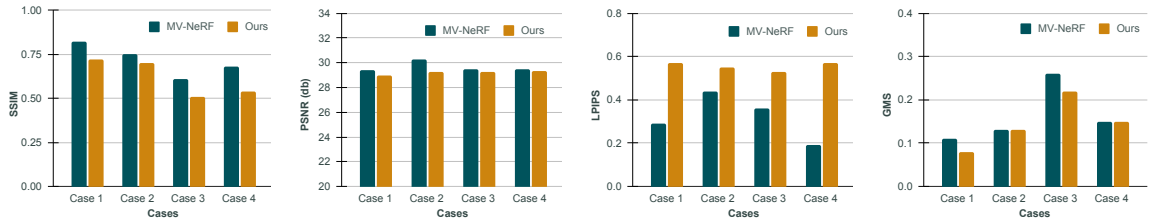


Figure 9.3 Comparison against ground-truth image after registration.

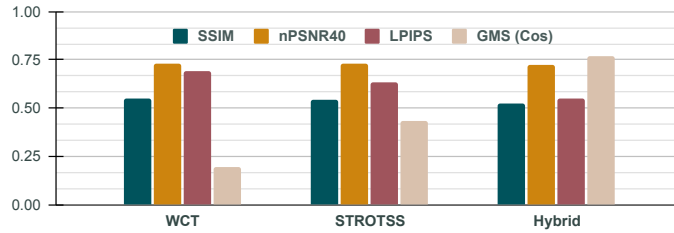


Figure 9.4 Comparison of various NST methods. Our hybrid approach strikes an optimal balance between style and structure. For readability, PSNR and GMS are scaled to the range [0,1]: PSNR is normalized with a maximum value of 40dB, and GMS is computed using cosine similarity.

Impact of neural style transfer methods. We compared the impact of the NST methods on image synthesis by measuring the added value of each method when compared with the real image. We compared the synthesis using WCT², STROTSS and the combination of both. The results reported in Fig. 9.4 show that our hybrid approach maintains the best

balance between style and structure with results comparable to WCT and STROTSS, while outperforming in LPIPS and GMS metrics.

Interactive segmentation on synthesized images. Additionally, we show that our synthesis method can be applied to downstream tasks such as vessel segmentation. Using an interactive semi-automated foundation model (Ke et al., 2023), not fine-tuned on surgical brain dataset, we obtain good segmentation results as illustrated in Fig. 9.5.

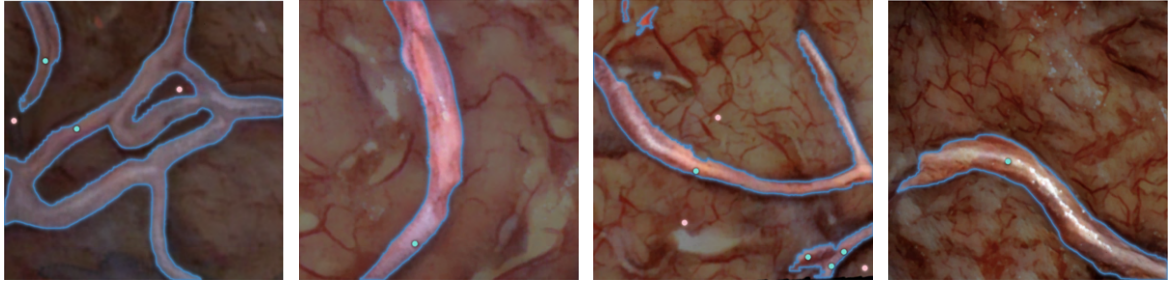


Figure 9.5 Interactive, semi-automated segmentation on synthesized images.

9.4 Conclusions

We introduced a novel approach for intraoperative surgical NeRF training using only a single image. Our preliminary results demonstrate that image synthesis from our single-image NeRF leads to good agreement with a multi-view strategy, and achieves good reconstruction and texture preservation when compared with ground-truth image.

A current limitation of our method is that the intraoperative style is static and does not account for changes in lighting conditions. Future work will focus on extending the approach to downstream registration tasks, where differentiable NeRF representations can be exploited for 3D/2D registration and camera pose estimation, and on exploring variants better suited to deformable laparoscopic scenes.

Finally, given its reliance on a preoperative 3D mesh as a strong prior, the method is particularly well suited to neurosurgery, where the operative field is mostly rigid, and the preoperative anatomy remains representative intraoperatively. In laparoscopy, where organs undergo substantial non-rigid deformation, the same pipeline could still be applied, but is expected to be less effective for surgical AR, since the preoperative mesh diverges from the intraoperative anatomy. In this context, surface reconstruction methods driven primarily by intraoperative images would be more appropriate.

Part III

Deep Learning non-rigid point cloud registration method

Introduction

This part represents the culmination of the work presented in this thesis (Fig. 9.6). Building on the experience gained with traditional rigid registration, deep-learning rigid methods, dataset generation, and intraoperative surface reconstruction, we move to the final and most challenging step: **patient-specific non-rigid registration** for surgical AR.

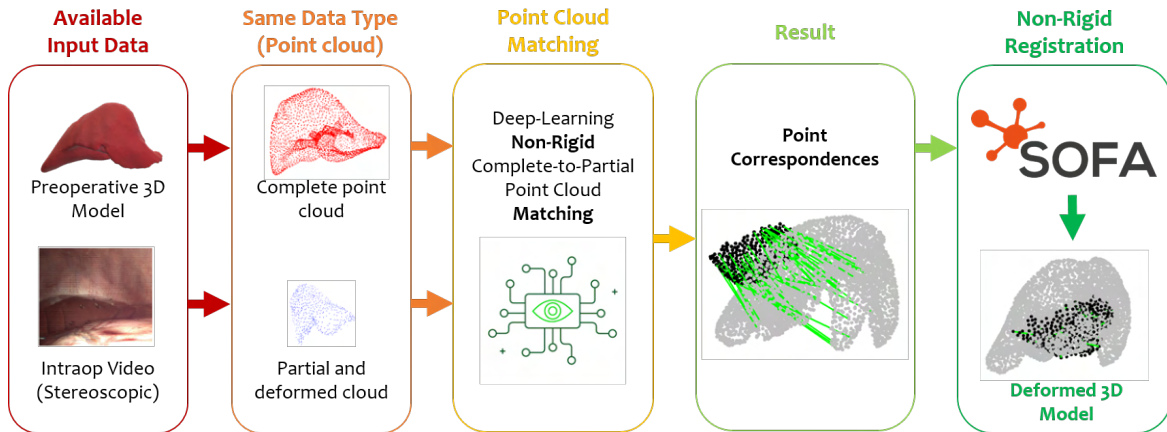


Figure 9.6 Proposed pipeline for deformable surgical augmented reality. A preoperative 3D organ model and an intraoperative laparoscopic image are first converted to a common representation (point clouds). A deep learning–based point cloud matching model estimates dense correspondences between the complete and partial point clouds. These correspondences are then used within a physics-based framework (SOFA) to perform non-rigid registration.

- *Chapter 10* introduces a deep-learning approach that predicts dense correspondences between preoperative and intraoperative point clouds using a Transformer-based architecture with overlap estimation and a dedicated matching module, followed by a physics-based deformation model to enforce biomechanically plausible motion. A key novelty is the **patient-specific** training scheme: for each case, the model is trained and evaluated on the same anatomy using an on-the-fly deformation and visible-crop generation pipeline, rather than relying on an anatomy-agnostic model trained across many patients. To the best of our knowledge, this is the first method to adopt such a strategy in this context, and it yields substantial improvements in performance metrics compared with state-of-the-art approaches, suggesting strong potential for surgical augmented reality.

Chapter 10

Towards patient-specific deformable registration in laparoscopic surgery

10.1 Introduction

As discussed in Chapter 7, point cloud registration methods have been proposed to align the preoperative organ surface point cloud, extracted from imaging modalities like CT scan, with an intraoperative point cloud, obtained from e.g. stereo vision of the surgical field, by estimating the transformation needed to match corresponding anatomical structures. Early non-deep-learning methods rely on finite-element models of liver mechanics: (Mestdagh and Cotin, 2022) cast registration as an optimal control problem on a hyperelastic FEM, solving for the surface forces that reproduce sparse displacement, while (Heiselman et al., 2020) build a linear modal basis from relaxed control-point perturbations and use a Levenberg–Marquardt solver to fit mode weights to surface and vascular landmarks. Moving to deep learning, a state-of-the-art algorithm in non-rigid partial point cloud registration applied to generic objects is Lepard (Li and Harada, 2022a). Their model combines a fully convolutional feature extraction, Transformer-based self cross-attention and differentiable matching to predict point cloud correspondences.

Despite the excellent performances on synthetic data of deep learning-based point cloud registration, challenges remain in real-world medical applications due to the disparity between

This work has been published as: Neri, A., Haouchine, N., Penza, V., Mattos, L. S. (2025, September). Towards Patient-Specific Deformable Registration in Laparoscopic Surgery. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 638-647). Cham: Springer Nature Switzerland. Publicly available at: https://doi.org/10.1007/978-3-032-05114-1_61. (This work was carried out during the **abroad period** at *Brigham and Women’s Hospital and Harvard Medical School*, in Boston, Massachusetts.)

dense, noise-free preoperative point clouds and intraoperative point clouds, which are often partial ($< 20\%$), noisy (Zha et al., 2023), and deformable, making accurate alignment difficult. To this end, recent methods have been developed to improve registration in challenging surgical environments. V2S-Net (Pfeiffer et al., 2020) learns a data-driven deformation basis from simulated tissue deformation to align preoperative volumes with full intraoperative surfaces. (Guan et al., 2023) proposes a rigid registration algorithm for liver surfaces in a low overlap scenario. It combines local and global features to estimate the overlap mask used to filter the non-overlapping region of the point cloud. LiverMatch (Yang et al., 2023) employs a transformer-based network for matching complete and partial liver point clouds in deformed conditions. Furthermore, the authors propose their synthetic dataset starting from 16 livers in the 3D-IRCADb-01 dataset (Soler et al., 2010) and applying deformations and crops to them. Zhang et al. (Zhang et al., 2024) propose KCR-Net, integrating a Neighborhood Feature Fusion Module (NFFM) for robust keypoint registration, even in low-overlap scenarios. Despite these advancements, challenges remain in handling intraoperative deformations, noise, and complete-to-partial registrations.

All these state-of-the-art methods adopt an agnostic strategy, training on large datasets to generalize to new cases. However, we argue this approach may not be ideal in surgical settings, where error margins are extremely low and optimal outcomes are critical for each patient. To address this limitation, we introduce a novel patient-specific non-rigid point cloud registration method that leverages preoperative information to tailor the registration process for each individual. Uniquely, our model is both trained and tested on the same patient’s anatomy. It employs a Transformer encoder-decoder architecture combined with overlap estimation and a dedicated matching module to predict dense correspondences. Additionally, we introduce a novel on-the-fly data generation strategy during training, which includes dynamic patient-specific deformation generation and an innovative visible crop logic. Experimental results on both synthetic and real data confirm that our patient-specific approach significantly outperforms traditional agnostic methods.

10.2 Methods

The architecture of the proposed registration method is presented in Fig. 10.1. The pipeline consists of five main modules: *Keypoint Features Extraction*, *Keypoint Features Conditioning*, *Complete-to-Partial Overlap Prediction*, *Point-to-Node Decoder* and *Matching Non-Rigid Registration*. First, KPConv is utilized to extract keypoints and their corresponding features from the input point clouds. Specifically, let $\mathbf{X} \in \mathbb{R}^{M \times 3}$ represent the complete point cloud

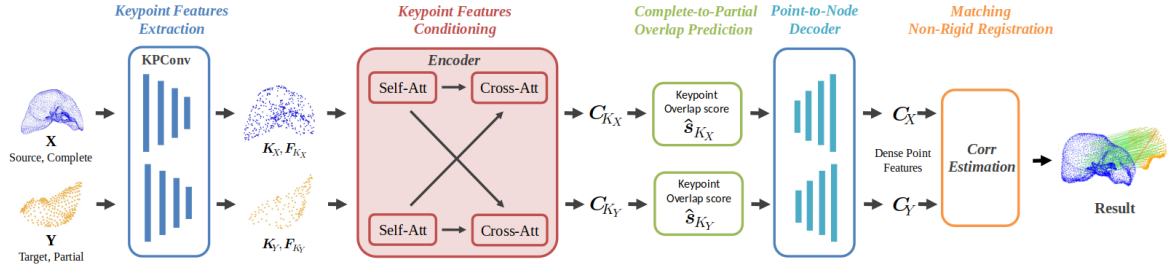


Figure 10.1 The network uses KPConv to detect keypoints and extract their features, refines these features through self- and cross-attention, and then feeds them into the decoder and matching module to predict dense point correspondences.

of the organ of interest (e.g., derived from a 3D model obtained via a CT scan), and let $\mathbf{Y} \in \mathbb{R}^{N \times 3}$ denote a partial point cloud of the same organ (e.g., captured using an endoscopic camera), where $N \ll M$. The extracted features are then processed by the encoder, which employs self-attention and cross-attention mechanisms to enhance feature representation and alignment. The *Complete-to-Partial Overlap Prediction* module predicts the regions of $\mathbf{X}$ and $\mathbf{Y}$ that correspond to the same anatomical area, helping to refine keypoint matching. Keypoint features and overlap scores are extended to the dense representation using a *Point-to-Node Decoder*. Subsequently, the *Matching* module determines correspondences by leveraging the similarity matrix computed between these features and the original 3D coordinates.

Keypoint features extraction module

Kernel Point Convolution (KPConv) (Thomas et al., 2019) is used for feature extraction. The KPConv backbone converts the input point cloud into a reduced set of keypoints $\mathbf{K}_X \in \mathbb{R}^{M' \times 3}$ and $\mathbf{K}_Y \in \mathbb{R}^{N' \times 3}$ through a series of ResNet-like blocks and strided convolutions, effectively downsampling the two clouds to achieve similar densities. In addition to the keypoints, the associated features $\mathbf{F}_{K_X} \in \mathbb{R}^{M' \times D}$ and $\mathbf{F}_{K_Y} \in \mathbb{R}^{N' \times D}$ are also produced.

Keypoint features conditioning (encoder) module

We reduce the dimension of the keypoints features (i.e. $\mathbf{F}_{K_X}$ and $\mathbf{F}_{K_Y}$) to $d = 256$ through a linear projection. Additionally, we apply sinusoidal positional encoding to the keypoint coordinates (i.e. $\mathbf{K}_X$ and $\mathbf{K}_Y$) to provide the transformer with an understanding of their spatial relationships. Both the high-level features and positional encodings are fed into the cross-encoder. High-level features enable the transformer to interpret and compare semantic content within the point clouds during alignment, while positional encoding supplies spatial context by emphasizing the structure and relative arrangement of points.

Following (Yew and Lee, 2022), each transformer cross-encoder layer consists of three sub-layers: (i) a multi-head self-attention layer that processes each point cloud independently; (ii) a multi-head cross-attention layer that updates the features of one point cloud by incorporating information from the other; and (iii) a position-wise feed-forward network. $\mathbf{C}_{K_X} \in \mathbb{R}^{M' \times d}$ and $\mathbf{C}_{K_Y} \in \mathbb{R}^{N' \times d}$ are the conditioned keypoint features produced by the encoder.

Overlap prediction module

The overlap module leverages conditioned features to compute keypoint overlap scores, denoted as $\hat{\mathbf{s}} = [\hat{\mathbf{s}}_{K_X}, \hat{\mathbf{s}}_{K_Y}]$ using a linear fully connected layer with sigmoid activation, following:

$$\hat{\mathbf{s}} = \begin{cases} \hat{\mathbf{s}}_{K_X} = 1/(1 + e^{-(\mathbf{C}_{K_X} \mathbf{W}_3 + \mathbf{b}_3)}) \\ \hat{\mathbf{s}}_{K_Y} = 1/(1 + e^{-(\mathbf{C}_{K_Y} \mathbf{W}_3 + \mathbf{b}_3)}) \end{cases} \quad (10.1)$$

where $\mathbf{W}_3$ and $\mathbf{b}_3$ are learnable weights and biases parameters. These scores quantify the probability that a keypoint lies within the overlapping region. In our low-overlap scenario—where less than 20% of $\mathbf{X}$ is covered by the target—accurate overlap score prediction is vital (Zhao et al., 2024), as it directs the model to focus its correspondence search on the relevant region.

Point-to-node decoder

To establish dense correspondences between the two clouds, we employ a decoder that propagates features and overlap scores across every dense point. Instead of using the conventional KPConv decoder, which relies on k-nearest neighbour search, we adopted a point-to-node grouping strategy similar to (Yu et al., 2021b). This approach offers two main advantages: (1) each point is uniquely assigned to one node (i.e. keypoint), ensuring no point is left unassigned, and (2) it inherently adapts to various scales. After grouping, we obtain dense point features ($\mathbf{C}_X \in \mathbb{R}^{M \times d}$, $\mathbf{C}_Y \in \mathbb{R}^{N \times d}$) and overlap scores ($\hat{\mathbf{s}}_X, \hat{\mathbf{s}}_Y$). The features are concatenated with the point’s 3D coordinates ($\mathbf{C}_X^* \in \mathbb{R}^{M \times d+3}$, $\mathbf{C}_Y^* \in \mathbb{R}^{N \times d+3}$), and this combined vector is fed into a MLP to estimate the final coordinates for $\mathbf{X}$. Our MLP first projects the concatenated input from a $d + 3$ dimensional space into a d -dimensional representation, applies non-linear activations, and ultimately outputs a 3-dimensional vector corresponding to the deformed coordinates ($\hat{\mathbf{X}}, \hat{\mathbf{Y}}$). Specifically: $\hat{\mathbf{X}} = \text{ReLU}(\mathbf{C}_X^* \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2$, where $\mathbf{W}_1, \mathbf{W}_2$ and $\mathbf{b}_1, \mathbf{b}_2$ are learnable weights and biases, respectively. At this stage, the model attempts an initial non-rigid registration. Although the

deformed coordinates it produces are rough, they provide a valuable signal for the chamfer loss to effectively guide the training process. Separately, the overlap scores are refined through a linear layer.

Matching and physics-based non-rigid registration

Within the Matching module, we begin by refining the $\mathbf{C}_X$ and $\mathbf{C}_Y$ using a linear layer. Then, we compute a similarity matrix and convert it into a confidence matrix $\mathbf{M}$ using a dual-softmax operation. Finally, following (Yang et al., 2023), matches $\mathbf{m}$ are selected from $\mathbf{M}$ based on a thresholded Mutual Nearest Neighbor criterion.

Once the matching is obtained, we can formulate the non-rigid registration problem as an energy minimization: $\min_{\delta\mathbf{X}} \frac{1}{2} \delta\mathbf{X}^T \mathbf{S} \delta\mathbf{X} + k \|\mathbf{Y}_m - \mathbf{X}_m\|^2$, where $\delta\mathbf{X}$ is the displacement field to be estimated, $\mathbf{S}$ is the stiffness matrix, which encodes the biomechanical properties of the organ, k is a scalar stiffness parameter that converts displacement to external forces.

The first term $\frac{1}{2} \delta\mathbf{X}^T \mathbf{S} \delta\mathbf{X}$ represents the internal elastic energy, ensuring that deformations follow biomechanical constraints, in particular in non-visible parts of the organ. The second term, $k \|\mathbf{Y}_m - \mathbf{X}_m\|^2$, enforces alignment with the observed correspondences. Since this formulation is static (no acceleration), the system does not include mass or damping terms. This system is solved using the Conjugate Gradient method.

Although dense bijective correspondences do not exist in complete-to-partial matching, our deformation field is not obtained via a pointwise map. In fact, we model the complete organ as a volumetric deformable body, which defines degrees of freedom over the entire domain, and use the predicted correspondences as sparse positional constraints (implemented as spring-like penalty terms between matched surface locations). Each correspondence, therefore, pulls the complete model locally toward the observed partial surface. The elastic regularisation propagates these sparse constraints into non-observed regions, yielding a smooth and well-defined dense deformation field. In this way, we estimate a dense displacement field through a physics-based static solve.

Optimization losses

Our method employs a weighted sum of three losses: (i) a matching loss (ML) (Lin et al., 2017) to supervise the confidence matrix $\mathbf{M}$ with the ground truth matches; (ii) a chamfer loss (CL) weighted on the overlap score, defined as: $CL(\hat{\mathbf{X}}, \mathbf{Y}, \hat{\mathbf{s}}_X) = \frac{1}{|\hat{\mathbf{X}}|} \sum_{\hat{x} \in \hat{X}} \hat{s}_x \min_{y \in Y} \|\hat{x} - y\|^p$. This loss is not focused on finding exact 1-1 correspondences but rather a global geometrical

alignment. Finally, (iii) an overlap loss (OL) similar to (Yew and Lee, 2022) designed to optimize the overlap scores.

Patient-specific training dataset generation

Generating a patient-specific training dataset involves creating pairs of complete and partial 3D point clouds from preoperative CT scans, incorporating both rigid and non-rigid transformations.

We employ the As-Rigid-As-Possible (ARAP) algorithm (Sorkine and Alexa, 2007) to generate two types of deformation: (i) compression deformation, which simulates the effect of CO_2 insufflation by displacing control points along their normal vectors with randomly assigned magnitudes in the range $[0, 0.1]$; and (ii) lobe deformations, introduced by applying random displacements to control points within the lobe region, with magnitudes ranging from $[0, 0.25]$.

To realistically generate partial point clouds similar to an endoscopic scene, we generate a dummy camera position that points toward the visible surface of the organ. We sample random spherical coordinates while constraining the polar and azimuthal angles to a realistic range for intraoperative settings. Using the camera direction and the surface normals of the organ, we compute the dot product to assess visibility. Only 5% of points closest to the camera, with a dot product below 80° , are retained.

Following the criteria above, we apply deformation, and cropping is applied randomly to constitute $\mathbf{Y}$, while a random rigid transformation is applied to $\mathbf{X}$. In addition, both $\mathbf{X}$ and $\mathbf{Y}$ are shuffled. This pipeline is executed on the fly for each intraoperative cloud, with a processing time of 0.5 seconds per sample. Since random parameters govern deformations and cropping, the model is exposed to a continuously varying dataset during training, eliminating the need for a pre-computed offline dataset. Conversely, validation is performed using deterministic seeds to ensure consistency. The test set follows the same generation process as validation but with a different seed. Since our approach is patient-specific, a new dataset must be generated and the model retrained for every patient. By applying early stopping based on the validation loss, training completes in approximately 12 hours.

10.3 Results

Matching performances

To evaluate the matching performance of our model, we employed “liver 1” from the 3D-IRCADb-01 dataset and generated a test set comprising 50 examples, following the procedure described in Section 10.2. We compared our approach with three alternative algorithms: (i) FPFH – where FPFH features (Rusu et al., 2009) are extracted from both complete and partial point clouds, and correspondences are established by matching each target point to its nearest source point in the feature space using Euclidean distance; (ii) Leopard (Li and Harada, 2022a) – by leveraging their pre-trained model on the 4DMatch dataset, a collection of dynamic, real-world, partially overlapping point cloud pairs; and (iii) LiverMatch (Yang et al., 2023) – utilizing the pre-trained model derived from liver data within the 3D-IRCADb-01 dataset. Notably, “liver 1” is absent from the LiverMatch training set, as it was also used for test set generation in their work, we call IRCAD-Liver1. We evaluated performance using the Matching Score (MS)—defined as the number of correctly predicted matches relative to the total available matches—and the Inlier Ratio (IR), which represents the proportion of correct matches among all predictions. Additionally, we report the absolute number of exact matches predicted (#MP) by each algorithm. Table 10.1 demonstrates that our algorithm significantly outperforms its competitors in identifying exact matches. Furthermore, we evaluated our algorithm under various loss configurations. Our results indicate that optimal performance is achieved when all the three losses are active (**Full**), confirming the synergistic relationship between overlap score estimation and correspondence estimation. Figure 10.2 visually compares the quality of the matches predicted by each method.

Table 10.1 Matching Metrics

Methods	MS (%)	IR (%)	Avg. #MP
<i>FPFH</i>	0.03 ± 0.11	0.03 ± 0.11	< 1
<i>Leopard</i>	1.6 ± 1.5	2.4 ± 2.2	4 ± 3.75
<i>LiverMatch</i>	13 ± 9	23 ± 13	39 ± 27
<i>Ours - ML + CL</i>	36 ± 8	90 ± 4	90 ± 20
<i>Ours - Full</i>	45 ± 7	92 ± 3	112.5 ± 17.5

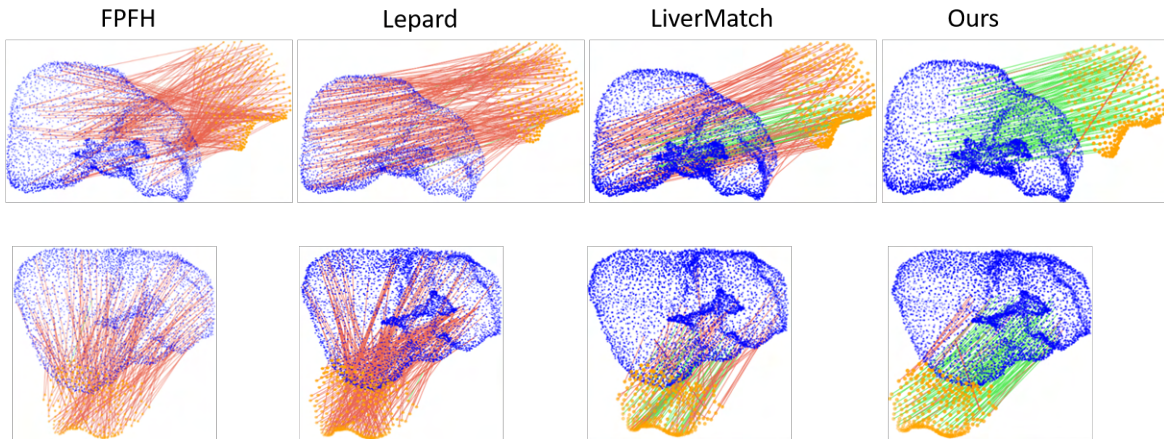


Figure 10.2 Qualitative results of predicted matches. Blue points denote the preoperative point cloud, while orange points represent the partial and deformed intraoperative cloud. Red lines indicate incorrect matches, and green lines highlight exact matches relative to the ground truth.

Non-rigid registration

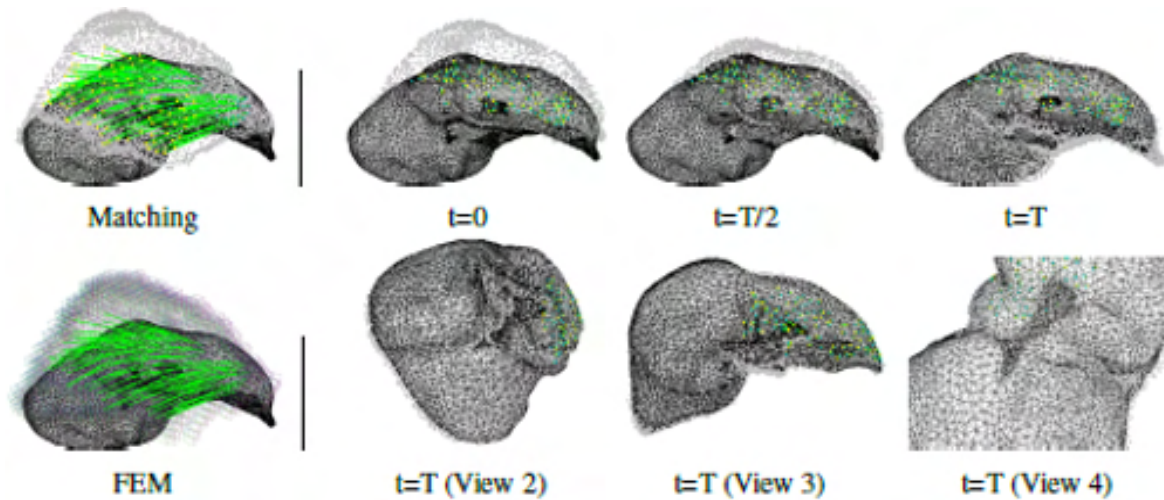


Figure 10.3 Non-rigid registration results on IRCAD-Liver1: Source point cloud is represented with gray points while target point clouds are represented in black wireframe. Green lines represent the matching correspondences and FEM models are represented with grid cells. Registered point clouds are presented at $t=0$, $t=T/2$ and $t=T$ are shown in 4 different views.

We present results on non-rigid registration using biomechanical modeling. We use the framework SOFA (Talbot et al., 2015) to generate a Finite Element model from the liver

mesh and generated around 12000 tetrahedra elements. We use a Young's modulus of 1.5 KPa and a Poisson ratio of 0.45. We compute Target Registration Errors (TREs) as the point-to-point Euclidean distance between vertices of the non-visible part of the preoperative and intraoperative meshes. We also report Fiducial Registration Errors (FREs). Results shown in Table 10.1 using 10 random deformations demonstrate that we obtain very low TREs with an average of 4.82 ± 3.33 mm. FREs are also very low, 1.68 ± 1.11 mm, suggesting few matching outliers and adequate registration. Moreover, our method outperforms LiverMatch, which exhibits a high number of mismatches, resulting in a TRE of 17.36 ± 13.68 mm and FRE of 28.76 ± 21.30 mm. Visualization of the non-rigid registration and FEM model is provided in Figure 10.3.

We then tested our method on the DePoLL porcine dataset (Modrzejewski et al., 2019) with 3D intraoperative point clouds reconstructed from the endoscopic images. We use a Young's modulus of 3 KPa and a Poisson ratio of 0.45. We report in Table 10.2 registration results on 13 cases from DePoll dataset.

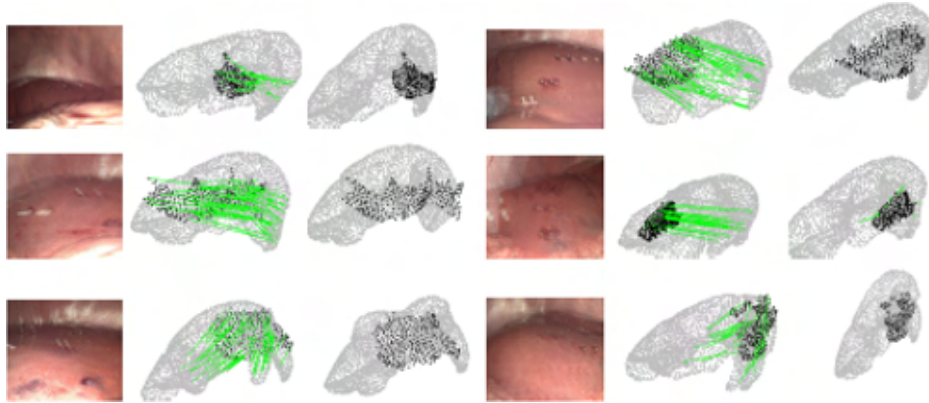


Figure 10.4 Qualitative results of non-registration on six cases from the DePoLL dataset are presented, showing the endoscopic image, the matching, and the registration from left to right. Preoperative point clouds are represented in gray, intraoperative point clouds in black, and correspondences with green lines.

Table 10.2 Registration results on DePoLL dataset (results in mm).

Met.	Case0	Case1	Case2	Case3	Case4	Case8	Case10	Case5	Case6	Case7	Case9	Case11	Case12	Avg
HD	11.12	8.44	6.62	6.23	15.45	10.17	12.04	8.12	5.65	6.93	6.88	5.98	7.75	8.45 ± 3.60
FRE	27.68	18.68	12.53	12.23	32.98	26.68	23.02	10.43	9.46	15.04	8.84	10.54	9.83	15.90 ± 6.41

We obtain an average Hausdorff distance (HD) of 8.45 ± 3.60 mm and an average FRE of 15.90 ± 6.41 mm, suggesting a good surface-to-surface registration and good matching. Visualizations of six cases are provided in Figure 10.4. It is important to note that this test

differs from those proposed in other works (Zhang et al., 2024), as it involves registering the reconstructed point cloud with the preoperative one, rather than with the reconstruction from the intraoperative CT scan.

10.4 Conclusion

We presented a novel method for intraoperative point cloud registration. By leveraging a patient-specific partial and deformable data generation strategy and training a Transformer-based matching network, our approach outperforms traditional agnostic methods, demonstrating improved accuracy and robustness, on synthetic and real surgical dataset.

Our method was trained on the most challenging setting, where the intraoperative point cloud represents only 5% of the full organ surface (5% partiality case). On the synthetic Deformed IRCAD dataset, the proposed patient-specific approach achieves a TRE of 4.82 ± 3.33 mm on the 5% partiality validation set, whereas the rigid method presented in Chapter 7 obtained 22.36 ± 12.04 mm. In other words, the error is reduced by more than a factor of four when only limited intraoperative information is available. For higher partialities, the TRE drops below 3mm in the small-deformation scenario, suggesting that, in this setting, the registration accuracy meets clinical requirements.

On the DePoll dataset, the non-rigid method still improves over the rigid baseline, reducing the TRE from 42.10 ± 23.74 mm to 15.90 ± 6.41 mm. However, this magnitude of error remains too high for surgical standards. As discussed in Chapter 7, DePoll includes large deformations, anatomical variability, and noise; in particular, the extreme non-rigid motion severely limits the model's ability to predict correct correspondences and perform accurate registration. Further developments are therefore necessary to enhance robustness against the large and unpredictable deformations that can occur during real surgical procedures.

Conclusion

Minimally invasive surgery (MIS) has become standard practice in many specialities, offering reduced trauma, fewer complications, and faster recovery compared to open surgery. However, these benefits come at the cost of a limited field of view and loss of direct tactile feedback. In today's practice, surgeons still rely heavily on their own experience and skills to interpret preoperative images, extract the key information, remember it, and mentally transfer it into the operative field. Bridging the gap between preoperative CT/MRI and the intraoperative laparoscopic view is cognitively demanding and error-prone, so outcomes remain strongly dependent on individual expertise. Surgical augmented reality aims to reduce this mental burden by directly overlaying preoperative information, typically in the form of 3D anatomical models, onto the intraoperative scene, thereby increasing the surgeon's situational awareness and visibility.

The central clinical need addressed in this thesis is therefore **non-rigid registration** of preoperative 3D models to intraoperative laparoscopic images, using only data that are routinely available (preoperative CT/MRI and laparoscopic video). Rigid registration alone is not sufficient for soft-tissue surgery, where respiration, insufflation, and surgical manipulation constantly deform the organs. At the same time, purely manual solutions are time-consuming, operator-dependent, and difficult to scale. The work presented in this thesis develops and evaluates a series of methods, from traditional rigid techniques to deep-learning-based rigid and non-rigid approaches, complemented by dataset generation and surface reconstruction, to move surgical AR closer to practical use in MIS.

In Part I, we address traditional rigid registration methods, focusing on manual registration using human–computer interfaces and semi-automatic point-based registration, and highlight a feasible extension of surgical AR to robotic surgery. In Part II, we move to deep-learning-based point cloud registration, still in the rigid setting. This part lays the foundations for non-rigid approaches by revealing the architectural limitations of rigid methods and by introducing dataset generation and surface reconstruction strategies. Finally, in Part III, we

propose our patient-specific non-rigid approach and demonstrate its superior performance on the surgical AR task.

10.5 Thesis contributions

The main contributions of this thesis can be summarised as follows:

- **Traditional rigid registration for surgical AR.** We proposed and evaluated three rigid registration approaches for surgical augmented reality: (i) a comparative study of human–computer interfaces for manual 3D model alignment, showing that manual registration can achieve accurate overlays but it is operator-dependent and requires continuous updates; (ii) an *in vivo* porcine study using the EVA Navi system during conventional laparoscopy, in which we tested a workflow combining point-based initialisation, manual refinement, and optical tracking for a ureter-identification task; the experiment showed that AR guidance with EVA shortened identification time compared with the standard view and was judged functional, efficient, and easy to learn; and (iii) an extension of EVA to robot-assisted surgery, where the registration is updated in real time using robotic kinematics instead of external tracking, achieving accurate AR overlay in simulated procedures.
- **Foundations for non-rigid point cloud registration with deep learning.** We laid the foundations for non-rigid point cloud registration by: (i) providing a structured review of registration and surgical AR methods, from traditional pipelines to hybrid and end-to-end deep-learning approaches; (ii) benchmarking traditional and deep-learning rigid complete-to-partial point cloud registration methods on synthetic laparoscopic data with controlled deformation and visibility, showing that rigid DL methods perform well under limited deformation but degrade sharply in highly deformable and high-partiality settings; (iii) proposing two dataset-generation strategies, including a curated dataset of 688 kidney 3D models derived from KiTS segmentations and a deformation/visibility pipeline for creating realistic complete-to-partial training pairs; and (iv) introducing a single-image NeRF-based intraoperative surface reconstruction method for neurosurgery. It leverages preoperative anatomy and appearance transfer to enable 3D reconstruction under strict data constraints.
- **First patient-specific non-rigid point cloud registration method with deep learning.** Finally, we proposed, to the best of our knowledge, the first *patient-specific* deep-learning framework for non-rigid point cloud registration in surgical AR. The method

uses a Transformer-based correspondence network with overlap estimation, followed by a physics-based deformation model, and is trained and tested on the same patient-specific anatomy using synthetic deformations and partial views generated on the fly. On synthetic data with soft deformations and limited visibility, the approach reduces TRE by more than a factor of four compared with the best rigid baseline (e.g., from ~ 22 mm (Sect. 7.3.1) to ~ 5 mm in the 5% partiality case (Sect. 10.3)) and reaches sub-3 mm accuracy for higher partialities, approaching clinical requirements. On data with very large deformations and anatomical variability, it still improves over rigid methods but does not yet meet surgical accuracy standards, which is not surprising given that the liver represents one of the most deformable abdominal organs in laparoscopy. This suggests that the proposed strategy could already be suitable for anatomies with more moderate deformation (e.g. retroperitoneal organs), while also highlighting the need for further developments to handle extreme soft-tissue motion as seen in liver surgery.

10.6 Impact

The work presented in this thesis has potential impact at clinical, methodological, and data/resource levels.

From a **clinical** perspective, the thesis contributes to bringing surgical augmented reality closer to practical use in minimally invasive surgery. The traditional rigid approaches, evaluated both in phantom and *in vivo* porcine experiments, demonstrate that AR guidance can reduce task time by up to 60% (e.g., ureter identification) and it is perceived as functional and easy to learn within typical laparoscopic workflows. The extension to robot-assisted surgery further demonstrates that existing robotic platforms can be leveraged to maintain AR overlays without external tracking hardware. Although these studies are preliminary and limited in scale, they suggest that well-integrated AR systems can enhance spatial awareness and potentially reduce cognitive load by offloading part of the mental fusion between preoperative and intraoperative information. Building on these results, the deep-learning methods proposed in this thesis directly target the limitations exposed in the rigid experiments: sensitivity to deformation. By introducing dedicated datasets and a patient-specific non-rigid framework, we move towards overlays that can adapt to soft-tissue motion and more challenging visibility conditions. Once integrated into a complete AR system, these advances are expected to further enhance the perceived usefulness and reliability of surgical AR in real procedures.

At a **methodological** level, the thesis clarifies the capabilities and limits of rigid registration in soft-tissue surgery and provides a systematic path toward non-rigid solutions. The benchmarking of rigid complete-to-partial point-cloud registration methods, including deep-learning approaches, quantifies when rigid models are sufficient (small deformations, larger visible surfaces) and when they fail (large deformations, high partiality). This evidence helps to redirect research efforts to genuinely deformable models. The proposed patient-specific non-rigid framework, although not yet ready for clinical deployment, shows that tailoring models to individual anatomy can substantially improve accuracy in realistic deformation scenarios, and thus represents a promising direction for future surgical AR systems.

In terms of **data**, the thesis addresses a key bottleneck for learning-based registration: the lack of suitable organ-specific datasets. The curated kidney dataset derived from KiTS segmentations, together with the deformation and visibility generation strategies, provides a reusable basis for training and evaluating rigid and non-rigid point cloud registration methods. Similarly, the single-image NeRF pipeline illustrates how preoperative data can be combined with minimal intraoperative information to obtain 3D representations under strict data constraints. These contributions support reproducibility and comparative studies, and they can be used by other researchers to test alternative architectures or learning strategies.

Overall, the thesis does not claim to solve non-rigid laparoscopic registration in its entirety, but it narrows the gap between current practice and clinically relevant surgical AR. By combining in-vivo experiments on traditional methods, deep-learning benchmarks, dataset creation, and a patient-specific non-rigid approach, the work provides both building blocks and evidence to guide subsequent research toward robust surgical AR systems.

10.7 Future perspective

Future work should first focus on **clinical validation**. So far, most of the experiments were conducted on phantoms, synthetic data, or *in vivo* porcine models. Future work should move towards human studies, designed to assess not only technical performance but also impact on operative time, error rates, and user acceptance.

A second direction concerns **surgical accuracy requirements and real-time performance**. As discussed in Sect. 1.1, the literature reports accepted TRE values of about 2 mm in spine and neurosurgery, and 5–6 mm in laparoscopy, reflecting the relative simplicity of near-rigid anatomy in the former versus deformable organs in the latter. However, especially when registering exposed anatomy in laparoscopic procedures, we argue that a target TRE closer to 2 mm should also be pursued. The task-specific acceptable accuracy (e.g. for tumour

margins or vessel identification) must therefore be defined together with clinicians, and future evaluations should be framed against these thresholds. This supports the idea of different “modes” of AR guidance, in which stricter accuracy is required for some tasks, while coarser guidance is acceptable for others.

Moving to **real-time performance**, in practical laparoscopic setups, a true real-time AR system should operate at approximately 25-30FPS, leaving on the order of 30-40ms per frame for all processing. If registration were to be recomputed at every frame, this budget would need to accommodate both surface reconstruction and registration. Current deep-learning surface reconstruction methods already require on the order of one second per update (Wen et al., 2025), and while rigid registration can be brought below 10 ms in favourable conditions (Mei et al., 2023a), non-rigid registration remains substantially slower. Our method, for example, ranges from under one second to more than five seconds per update, depending on the number of tetrahedra in the SOFA-based deformation model and the desired local deformation accuracy. This means that, at present, fully recomputing a non-rigid registration at every frame is not yet compatible with strict real-time constraints. Future work could therefore proceed along two complementary directions: on the one hand, investigating lighter-weight models for both reconstruction and deformation; on the other, adopting hybrid update strategies in which a computationally expensive non-rigid registration is performed intermittently, and its result is then propagated between frames using faster components such as camera/organ tracking and rigid registration. For instance, integrating non-rigid registration with optical flow, feature tracking, or SLAM-like methods could help maintain alignment over time and decrease the frequency of expensive non-rigid updates, especially during fast camera motions. Such schemes may offer a practical compromise that leverages current technology while keeping the AR overlay sufficiently up-to-date for intraoperative use.

Methodologically, there is room to **improve accuracy by using richer cues**. Beyond pure geometry, future methods could exploit anatomical landmarks (such as the falciform ligament in liver surgery), tissue appearance, and colour information to refine correspondence estimation, particularly in cases with limited visibility or large deformation. In parallel, new **online quality indicators** are needed to estimate registration reliability when ground truth is not available and to alert the surgeon when the overlay may be unreliable.

Finally, a further open direction is **multi-organ registration**. In this thesis, as in most of the current literature, registration models are trained and evaluated on a single target anatomy (e.g. liver, kidney), assuming that the organ of interest can be treated in isolation. In practice, however, surgeons reason over multiple interacting structures (organs, vessels,

ligaments) whose relative motion and constraints may carry important information. Future work could therefore explore models that register several organs simultaneously, for example, by learning shared representations across anatomies and encoding inter-organ constraints. One possible path is the development of *foundation-style* registration models pre-trained on large, heterogeneous collections of abdominal anatomies and tasks, and then adapted in a patient-specific manner, combining broad anatomical priors with the accuracy of case-specific refinement.

Together, these directions can help transform the methods presented in this thesis from promising prototypes into robust tools for everyday minimally invasive surgery.

References

- Adagolodjo, Y., Golse, N., Vibert, E., De Mathelin, M., Cotin, S., and Courtecuisse, H. (2018). Marker-based registration for large deformations-application to open liver surgery. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4007–4012. IEEE.
- Amberg, B., Romdhani, S., and Vetter, T. (2007). Optimal step nonrigid icp algorithms for surface registration. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE.
- Amparore, D., Checcucci, E., Piazzolla, P., Piramide, F., De Cillis, S., Piana, A., Verri, P., Manfredi, M., Fiori, C., Vezzetti, E., et al. (2022). Indocyanine green drives computer vision based 3d augmented reality robot assisted partial nephrectomy: the beginning of “automatic” overlapping era. *Urology*, 164:e312–e316.
- Aoki, Y., Goforth, H., Srivatsan, R. A., and Lucey, S. (2019). Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7163–7172.
- Atallah, S., Larach, S. W., and Monson, J. R. (2016). Stereotactic navigation for tamis-tme. *Minimally Invasive Therapy & Allied Technologies*, 25(5):271–277.
- Atallah, S., Parra-Davila, E., Melani, A., Romagnolo, L., Larach, S., and Marescaux, J. (2019). Robotic-assisted stereotactic real-time navigation: initial clinical experience and feasibility for rectal cancer surgery. *Techniques in Coloproctology*, 23:53–63.
- Baker, S. and Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework. *International journal of computer vision*, 56:221–255.
- Barcali, E., Iadanza, E., Manetti, L., Francia, P., Nardi, C., and Bocchi, L. (2022). Augmented reality in surgery: a scoping review. *Applied Sciences*, 12(14):6890.
- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.
- Beckmann, S., Rosenthal, J.-C., Wisotzky, E. L., Hilsmann, A., and Eisert, P. (2023). Automatic registration of anatomical structures of stereo-endoscopic point clouds. In *Current Directions in Biomedical Engineering*, volume 9, pages 615–618. De Gruyter.
- Benincasa, A. B., Clements, L. W., Herrell, S. D., and Galloway, R. L. (2008). Feasibility study for image-guided kidney surgery: Assessment of required intraoperative surface for accurate physical to image space registrations. *Medical Physics*, 35(9):4251–4261.

- Bergström, J., Dalsgaard, T.-S., Alexander, J., and Hornbæk, K. (2021). How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In *proceedings of the 2021 CHI conference on human factors in computing systems*.
- Bernhardt, S., Nicolau, S. A., Soler, L., and Doignon, C. (2017a). The status of augmented reality in laparoscopic surgery as of 2016. *Medical Image Analysis*, 37:66–90.
- Bernhardt, S., Nicolau, S. A., Soler, L., and Doignon, C. (2017b). The status of augmented reality in laparoscopic surgery as of 2016. *Medical image analysis*, 37:66–90.
- Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. Spie.
- Birlo, M., Edwards, P. E., Clarkson, M., and Stoyanov, D. (2022). Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review. *Medical Image Analysis*, 77:102361.
- Brollo, P. P., Puggioni, A., Tumminelli, F., Colangelo, A., Biddau, C., Cherchi, V., and Bresadola, V. (2023). Preventing iatrogenic ureteral injury in colorectal surgery: A comprehensive and systematic review of the last 2 decades of literature and future perspectives. *Surgery Today*, pages 1–19.
- Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Bruhn, A., Weickert, J., and Schnörr, C. (2005). Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International journal of computer vision*, 61:211–231.
- Brunet, J.-N., Mendizabal, A., Petit, A., Golse, N., Vibert, E., and Cotin, S. (2019). Physics-based deep neural network for augmented reality during liver surgery. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22*, pages 137–145. Springer.
- Burks, F. N. and Santucci, R. A. (2014). Management of iatrogenic ureteral injury. *Therapeutic advances in urology*, 6(3):115–124.
- Carbone, M., Piazza, R., and Condino, S. (2020). Commercially available head-mounted displays are unsuitable for augmented reality surgical guidance: A call for focused research for surgical applications.
- Chadelon, K., Pitout, A., Souchaud, M., Desternes, J., Margue, G., Peyras, J., Bourdel, N., Bernhard, J.-C., and Bartoli, A. (2025). Landmark-free automatic digital twin registration in robot-assisted partial nephrectomy using a generic end-to-end model. *International Journal of Computer Assisted Radiology and Surgery*, 20(9):1931–1940.
- Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al. (2015). Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

- Chen, G., Wang, M., Yuan, L., Yang, Y., and Yue, Y. (2023a). Rethinking point cloud registration as masking and reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17717–17727.
- Chen, G., Wang, M., Zhang, Q., Yuan, L., and Yue, Y. (2023b). Full transformer framework for robust point cloud registration with deep information interaction. *IEEE Transactions on Neural Networks and Learning Systems*.
- Chen, Z., Chen, H., Gong, L., Yan, X., Wang, J., Guo, Y., Qin, J., and Wei, M. (2022). Utopic: Uncertainty-aware overlap prediction network for partial point cloud registration. In *Computer Graphics Forum*, volume 41, pages 87–98. Wiley Online Library.
- Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., et al. (2008). Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference*, volume 2008, pages 129–136. Salerno, Italy.
- Cofano, F., Di Perna, G., Bozzaro, M., Longo, A., Marengo, N., Zenga, F., Zullo, N., Cavalieri, M., Damiani, L., Boges, D. J., et al. (2021). Augmented reality in medical practice: from spine surgery to remote assistance. *Frontiers in Surgery*, 8:657901.
- Collins, J. A., Weis, J. A., Heiselman, J. S., Clements, L. W., Simpson, A. L., Jarnagin, W. R., and Miga, M. I. (2017). Improving registration robustness for image-guided liver surgery in a novel human-to-phantom data framework. *IEEE transactions on medical imaging*, 36(7):1502–1510.
- Collins, T., Pizarro, D., Gasparini, S., Bourdel, N., Chauvet, P., Canis, M., Calvet, L., and Bartoli, A. (2020). Augmented reality guided laparoscopic surgery of the uterus. *IEEE Transactions on Medical Imaging*, 40(1):371–380.
- Collins, T., Pizarro, D., Gasparini, S., Bourdel, N., Chauvet, P., Canis, M., Calvet, L., and Bartoli, A. (2021). Augmented reality guided laparoscopic surgery of the uterus. *IEEE Transactions on Medical Imaging*, 40(1):371–380.
- Condino, S., Carbone, M., Piazza, R., Ferrari, M., and Ferrari, V. (2019). Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks. *IEEE Transactions on Biomedical Engineering*, 67(2):411–419.
- Coste-Manière, È., Adhami, L., Mourgues, F., and Bantiche, O. (2004). Optimal planning of robotically assisted heart surgery: first results on the transfer precision in the operating room. *The International Journal of Robotics Research*, 23(4-5):539–548.
- Croghan, S. M., Zaborowski, A., Mohan, H. M., Mulvin, D., McGuire, B. B., Murphy, M., Galvin, D. J., Lennon, G., Quinlan, D., and Winter, D. C. (2019). The sentinel stent? a systematic review of the role of prophylactic ureteric stenting prior to colorectal resections. *International Journal of Colorectal Disease*, 34:1161–1178.
- Curless, B. and Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312.

- Dai, Y., Yang, X., Hao, J., Luo, H., Mei, G., and Jia, F. (2025). Preoperative and intraoperative laparoscopic liver surface registration using deep graph matching of representative overlapping points. *International Journal of Computer Assisted Radiology and Surgery*, 20(2):269–278.
- De Paolis, L. T. and De Luca, V. (2019). Augmented visualization with depth perception cues to improve the surgeon’s performance in minimally invasive surgery. *Medical & biological engineering & computing*, 57:995–1013.
- Den Boer, R., de Jongh, C., Huijbers, W., Jaspers, T., Pluim, J., van Hillegersberg, R., Van Eijnatten, M., and Ruurda, J. (2022). Computer-aided anatomy recognition in intrathoracic and-abdominal surgery: a systematic review. *Surgical endoscopy*, 36(12):8737–8752.
- Dhoju, R., Alsadoon, A., Prasad, P., Al-Saiyd, N. A., and Alrubaie, A. (2021). Augmented reality navigation for liver surgery: an enhanced coherent point drift algorithm based hybrid optimization scheme. *Multimedia Tools and Applications*, 80(18):28179–28200.
- Doornbos, M.-C. J., Peek, J. J., Maat, A. P., Ruurda, J. P., De Backer, P., Cornelissen, B. M., Mahtab, E. A., Sadeghi, A. H., and Kluin, J. (2024). Augmented reality implementation in minimally invasive surgery for future application in pulmonary surgery: a systematic review. *Surgical Innovation*, 31(6):646–658.
- Espinell, Y., Calvet, L., Botros, K., Buc, E., Tilmant, C., and Bartoli, A. (2022). Using multiple images and contours for deformable 3d–2d registration of a preoperative ct in laparoscopic liver surgery. *International Journal of Computer Assisted Radiology and Surgery*, 17(12):2211–2219.
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341.
- Fehrentz, M., Azampour, M. F., Dorent, R., Rasheed, H., Galvin, C., Golby, A., Wells, W. M., Frisken, S., Navab, N., and Haouchine, N. (2024). Intraoperative registration by cross-modal inverse neural rendering. In *MICCAI*, pages 317–327.
- Fu, K., Liu, S., Luo, X., and Wang, M. (2021a). Robust point cloud registration framework based on deep graph matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8893–8902.
- Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., and Yang, X. (2020). Deep learning in medical image registration: a review. *Physics in Medicine & Biology*, 65(20):20TR01.
- Fu, Y., Lei, Y., Wang, T., Patel, P., Jani, A. B., Mao, H., Curran, W. J., Liu, T., and Yang, X. (2021b). Biomechanically constrained non-rigid mr-trus prostate registration using deep learning based 3d point cloud matching. *Medical image analysis*, 67:101845.
- Gadoux, E. and Bartoli, A. (2025). Automatic deep deformable registration using domain adaptation and run-time optimisation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 65–74. Springer.

- Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE.
- Geveci, B., Schroeder, W., Brown, A., and Wilson, G. (2012). Vtk. *The architecture of open source applications*, 1:387–402.
- Gojcic, Z., Zhou, C., Wegner, J. D., Guibas, L. J., and Birdal, T. (2020). Learning multiview 3d point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1759–1769.
- Gold, S., Rangarajan, A., Lu, C.-P., Pappu, S., and Mjolsness, E. (1998). New algorithms for 2d and 3d point matching: Pose estimation and correspondence. *Pattern recognition*, 31(8):1019–1031.
- Golse, N., Petit, A., Lewin, M., Vibert, E., and Cotin, S. (2021). Augmented reality during open liver surgery using a markerless non-rigid registration system. *Journal of Gastrointestinal Surgery*, 25:662–671.
- Gouveia, P. F., Costa, J., Morgado, P., Kates, R., Pinto, D., Mavioso, C., Anacleto, J., Martinho, M., Lopes, D. S., Ferreira, A. R., et al. (2021). Breast cancer surgery with augmented reality. *The Breast*, 56:14–17.
- Gu, J., Trevithick, A., Lin, K.-E., Susskind, J. M., Theobalt, C., Liu, L., and Ramamoorthi, R. (2023). Nerfdiff: Single-image view synthesis with nerf-guided distillation from 3d-aware diffusion. In *International Conference on Machine Learning*, pages 11808–11826. PMLR.
- Guan, P., Luo, H., Guo, J., Zhang, Y., and Jia, F. (2023). Intraoperative laparoscopic liver surface registration with preoperative ct using mixing features and overlapping region masks. *International Journal of Computer Assisted Radiology and Surgery*, 18(8):1521–1531.
- Guha, D., Alotaibi, N. M., Nguyen, N., Gupta, S., McFaul, C., and Yang, V. X. (2017). Augmented reality in neurosurgery: a review of current concepts and emerging applications. *Canadian Journal of Neurological Sciences*, 44(3):235–245.
- Guo, M.-H., Cai, J.-X., Liu, Z.-N., Mu, T.-J., Martin, R. R., and Hu, S.-M. (2021). Pct: Point cloud transformer. *Computational Visual Media*, 7:187–199.
- Halabi, W. J., Jafari, M. D., Nguyen, V. Q., Carmichael, J. C., Mills, S., Pigazzi, A., and Stamos, M. J. (2014). Ureteral injuries in colorectal surgery: an analysis of trends, outcomes, and risk factors over a 10-year period in the united states. *Diseases of the colon & rectum*, 57(2):179–186.
- Han, X.-F., Laga, H., and Bennamoun, M. (2019). Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era. *IEEE transactions on pattern analysis and machine intelligence*, 43(5):1578–1604.
- Haskins, G., Kruger, U., and Yan, P. (2020). Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31:1–18.

- Heiselman, J. S., Jarnagin, W. R., and Miga, M. I. (2020). Intraoperative correction of liver deformation using sparse surface and vascular features via linearized iterative boundary reconstruction. *IEEE transactions on medical imaging*, 39(6):2223–2234.
- Heller, N., Isensee, F., Maier-Hein, K. H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., et al. (2021). The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. *Medical image analysis*, 67:101821.
- Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Zhao, Z., Chen, H., Wang, L., Golts, A., Khapun, D., Shats, D., Shoshan, Y., Gilboa-Solomon, F., George, Y., Yang, X., Zhang, J., Zhang, J., Xia, Y., Wu, M., Liu, Z., Walczak, E., McSweeney, S., Vasdev, R., Hornung, C., Solaiman, R., Schoepfhoerster, J., Abernathy, B., Wu, D., Abdulkadir, S., Byun, B., Spriggs, J., Struyk, G., Austin, A., Simpson, B., Hagstrom, M., Virnig, S., French, J., Venkatesh, N., Chan, S., Moore, K., Jacobsen, A., Austin, S., Austin, M., Regmi, S., Papanikolopoulos, N., and Weight, C. (2023). The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct.
- Hey, G., Guyot, M., Carter, A., and Lucke-Wold, B. (2023). Augmented reality in neurosurgery: a new paradigm for training. *Medicina*, 59(10):1721.
- Hu, X., Zhang, D., Chen, J., Wu, Y., and Chen, Y. (2022). Nrtnet: An unsupervised method for 3d non-rigid point cloud registration based on transformer. *Sensors*, 22(14):5128.
- Huang, B., Yang, X., and Jia, F. (2025). A landmark-free 3d–2d rigid liver registration via point cloud matching for laparoscopic surgery. *Healthcare Technology Letters*, 12(1):e70030.
- Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., and Schindler, K. (2021a). Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 4267–4276.
- Huang, T., Li, R., Li, Y., Zhang, X., and Liao, H. (2021b). Augmented reality-based autostereoscopic surgical visualization system for telesurgery. *International Journal of Computer Assisted Radiology and Surgery*, 16(11):1985–1997.
- Huang, X., Li, S., Zuo, Y., Fang, Y., Zhang, J., and Zhao, X. (2022). Unsupervised point cloud registration by learning unified gaussian mixture models. *IEEE Robotics and Automation Letters*, 7(3):7028–7035.
- Huang, X., Mei, G., Zhang, J., and Abbas, R. (2021c). A comprehensive survey on point cloud registration. *arXiv preprint arXiv:2103.02690*.
- Igarashi, T., Moscovich, T., and Hughes, J. F. (2005). As-rigid-as-possible shape manipulation. *ACM transactions on Graphics (TOG)*, 24(3):1134–1141.
- Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211.

- Jain, A., Tancik, M., and Abbeel, P. (2021). Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5885–5894.
- Jha, S. (2015). Ureteric injury: always a guilty verdict? *BJOG: An International Journal of Obstetrics & Gynaecology*, 122(4):499–499.
- Jha, S., Coomarasamy, A., and Chan, K. K. (2004). Ureteric injury in obstetric and gynaecological surgery. *The Obstetrician & Gynaecologist*, 6(4):203–208.
- Jian, B. and Vemuri, B. C. (2011). Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1633–1645.
- Jones J, Walizai T, S. R. (2009). Hepatomegaly. <https://doi.org/10.53347/rID-6213>.
- Ke, L., Ye, M., Danelljan, M., Liu, Y., Tai, Y.-W., Tang, C.-K., and Yu, F. (2023). Segment anything in high quality. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.
- Khodadad, D., Ahmadian, A., Ay, M., Esfahani, A. F., Banaem, H. Y., and Zaidi, H. (2011). B-spline based free form deformation thoracic non-rigid registration of ct and pet images. In *International Conference on Graphic and Image Processing (ICGIP 2011)*, volume 8285, pages 390–396. SPIE.
- Kilkenny, C., Browne, W., Cuthill, I. C., Emerson, M., and Altman, D. G. (2010a). Animal research: reporting in vivo experiments: the arrive guidelines. *British journal of pharmacology*, 160(7):1577.
- Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., and Altman, D. G. (2010b). Improving bioscience research reporting: the arrive guidelines for reporting animal research. *Journal of Pharmacology and Pharmacotherapeutics*, 1(2):94–99.
- Kim, S. and Kim, K. H. (2021). How to use neuronavigation for the brain. *Journal of the Korean Society of Stereotactic and Functional Neurosurgery*, 17(2):126–132.
- Kiyomatsu, H., Ma, L., Wang, J., Kiyomatsu, T., Tsukihara, H., Kobayashi, E., Sakuma, I., and Ishihara, S. (2021). Deformation of the pelvic arteries caused by pneumoperitoneum and postural changes in an animal model. *in vivo*, 35(1):275–281.
- Köhler, H., Pfahl, A., Moulla, Y., Thomaßen, M. T., Maktabi, M., Gockel, I., Neumuth, T., Melzer, A., and Chalopin, C. (2022). Comparison of image registration methods for combining laparoscopic video and spectral image data. *Scientific Reports*, 12(1):16459.
- Kolkin, N., Salavon, J., and Shakhnarovich, G. (2019). Style transfer by relaxed optimal transport and self-similarity. In *IEEE/CVF CVPR*, pages 10051–10060.
- Koo, B., Özgür, E., Le Roy, B., Buc, E., and Bartoli, A. (2017). Deformable registration of a preoperative 3d liver volume to a laparoscopy image using contour and shading cues. In *International conference on medical image computing and computer-assisted intervention*, pages 326–334. Springer.

- Koo, B., Robu, M. R., Allam, M., Pfeiffer, M., Thompson, S., Gurusamy, K., Davidson, B., Speidel, S., Hawkes, D., Stoyanov, D., et al. (2022). Automatic, global registration in laparoscopic liver surgery. *International Journal of Computer Assisted Radiology and Surgery*, 17(1):167–176.
- Krug, K., Satkowski, M., Docea, R., Ku, T.-Y., and Dachsel, R. (2023). Point cloud alignment through mid-air gestures on a stereoscopic display. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–7.
- Kuhlemann, I., Kleemann, M., Jauer, P., Schweikard, A., and Ernst, F. (2017). Towards x-ray free endovascular interventions—using hololens for on-line holographic visualisation. *Healthcare technology letters*, 4(5):184–187.
- Kumari Rani, A. (2020). Urological injury in obstetric and gynecological surgery: a retrospective analysis of five years. *Indian J Obstet Gynecol (IJOG)*, 89(2):18–22.
- Labrunie, M., Ribeiro, M., Mourthadhoi, F., Tilmant, C., Le Roy, B., Buc, E., and Bartoli, A. (2022). Automatic preoperative 3d model registration in laparoscopic liver resection. *International Journal of Computer Assisted Radiology and Surgery*, 17(8):1429–1436.
- Lau, L. W., Luciano, M., Schnermann, M., and Cha, J. (2020). Ureter identification in an inflammatory pig model using a novel near-infrared fluorescent dye. *Lasers in surgery and medicine*, 52(6):537–542.
- Laugwitz, B., Held, T., and Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pages 63–76. Springer.
- Leal Ghezzi, T. and Campos Corleta, O. (2016). 30 years of robotic surgery. *World journal of surgery*, 40:2550–2557.
- Li, Y. and Harada, T. (2022a). Leopard: Learning partial point cloud matching in rigid and deformable scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5554–5564.
- Li, Y. and Harada, T. (2022b). Non-rigid point cloud registration with neural deformation pyramid. *Advances in Neural Information Processing Systems*, 35:27757–27768.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.
- Luo, H., Yin, D., Zhang, S., Xiao, D., He, B., Meng, F., Zhang, Y., Cai, W., He, S., Zhang, W., et al. (2020). Augmented reality navigation for liver resection with a stereoscopic laparoscope. *Computer methods and programs in biomedicine*, 187:105099.
- Makhataeva, Z. and Varol, H. A. (2020). Augmented reality for robotics: A review. *Robotics*, 9(2):21.
- Marques, B., Plantefève, R., Roy, F., Haouchine, N., Jeanvoine, E., Peterlik, I., and Cotin, S. (2015). Framework for augmented reality in minimally invasive laparoscopic surgery. In *2015 17th International conference on E-health networking, application & services (HealthCom)*, pages 22–27. IEEE.

- Medics 3D (2024). Medics 3d official website. Accessed: 2024-08-12.
- Mei, G., Poiesi, F., Saltori, C., Zhang, J., Ricci, E., and Sebe, N. (2023a). Overlap-guided gaussian mixture models for point cloud registration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4511–4520.
- Mei, G., Tang, H., Huang, X., Wang, W., Liu, J., Zhang, J., Van Gool, L., and Wu, Q. (2023b). Unsupervised deep probabilistic approach for partial point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13611–13620.
- Mendes, D., Caputo, F. M., Giachetti, A., Ferreira, A., and Jorge, J. (2019). A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, volume 38, pages 21–45. Wiley Online Library.
- Mestdagh, G. and Cotin, S. (2022). An optimal control problem for elastic registration and force estimation in augmented surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 74–83. Springer.
- Mhiri, I., Pizarro, D., and Bartoli, A. (2025). Neural patient-specific 3d–2d registration in laparoscopic liver resection. *International Journal of Computer Assisted Radiology and Surgery*, 20(1):57–64.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106.
- Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329.
- Modrzejewski, R., Collins, T., Seeliger, B., Bartoli, A., Hostettler, A., and Marescaux, J. (2019). An in vivo porcine dataset and evaluation methodology to measure soft-body laparoscopic liver registration accuracy with an extended algorithm that handles collisions. *International journal of computer assisted radiology and surgery*, 14:1237–1245.
- Molina, C. A., Sciubba, D. M., Greenberg, J. K., Khan, M., and Witham, T. (2021). Clinical accuracy, technical precision, and workflow of the first in human use of an augmented-reality head-mounted display stereotactic navigation system for spine surgery. *Operative Neurosurgery*, 20(3):300–309.
- Monji-Azad, S., Hesser, J., and Löw, N. (2023). A review of non-rigid transformations and learning-based 3d point cloud registration methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:58–72.
- MORELLI, A. (2016). Towards deformable registration for augmented reality in robotic assisted partial nephrectomy.
- Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM TOG*, 41(4):1–15.
- Muntoni, A. and Cignoni, P. (2021). PyMeshLab.

- Mur-Artal, R., Montiel, J. M. M., and Tardos, J. D. (2015). Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163.
- Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275.
- Neri, A., Haouchine, N., Penza, V., and Mattos, L. S. (2025a). Towards patient-specific deformable registration in laparoscopic surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 638–647. Springer.
- Neri, A., Penza, V., Baldini, C., and Mattos, L. S. (2025b). Surgical augmented reality registration methods: A review from traditional to deep learning approaches. *Computerized Medical Imaging and Graphics*, page 102616.
- Neri, A., Penza, V., and Mattos, L. S. (2023). Comparative usability study of human-computer interfaces for 3d model manipulation in surgical augmented reality applications. *Hamlyn Symposium on Medical Robotics 2023*.
- Newcombe, R. A., Lovegrove, S. J., and Davison, A. J. (2011). Dtam: Dense tracking and mapping in real-time. In *2011 international conference on computer vision*, pages 2320–2327. IEEE.
- Nguyen, N. Q., Cardinell, J., Ramjist, J. M., Lai, P., Dobashi, Y., Guha, D., Androutsos, D., and Yang, V. X. (2020). An augmented reality system characterization of placement accuracy in neurosurgery. *Journal of Clinical Neuroscience*, 72:392–396.
- Östrand, E., Rystedt, J., Engstrand, J., Frühling, P., Hemmingsson, O., Sandström, P., Sternby Eilard, M., Tingstedt, B., and Buchwald, P. (2024). Importance of resection margin after resection of colorectal liver metastases in the era of modern chemotherapy: population-based cohort study. *BJS open*, 8(3):zrae035.
- Padovan, E., Marullo, G., Tanzi, L., Piazzolla, P., Moos, S., Porpiglia, F., and Vezzetti, E. (2022). A deep learning framework for real-time 3d model registration in robot-assisted laparoscopic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 18(3):e2387.
- Pang, J. H., Spalding, L., and Pasapula, C. (2023). Minimally invasive surgery: has it come of age? *Orthopaedics and Trauma*, 37(1):62–70.
- Paulus, C. J., Haouchine, N., Kong, S.-H., Soares, R. V., Cazier, D., and Cotin, S. (2017). Handling topological changes during elastic registration: Application to augmented reality in laparoscopic surgery. *International journal of computer assisted radiology and surgery*, 12:461–470.
- Paulus, C. J., Untereiner, L., Courtecuisse, H., Cotin, S., and Cazier, D. (2015). Virtual cutting of deformable objects based on efficient topological operations. *The Visual Computer*, 31:831–841.
- Penza, V., Ciullo, A. S., Moccia, S., Mattos, L. S., and De Momi, E. (2018a). Endoabs dataset: Endoscopic abdominal stereo image dataset for benchmarking 3d stereo reconstruction algorithms. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 14(5):e1926.

- Penza, V., De Momi, E., Enayati, N., Chupin, T., Ortiz, J., and Mattos, L. S. (2017). Envisors: Enhanced vision system for robotic surgery. a user-defined safety volume tracking to minimize the risk of intraoperative bleeding. *Frontiers in Robotics and AI*, 4:15.
- Penza, V., Moccia, S., De Momi, E., and Mattos, L. S. (2020a). Enhanced vision to improve safety in robotic surgery. In *Handbook of Robotic and Image-Guided Surgery*, pages 223–237. Elsevier.
- Penza, V., Moccia, S., Gallarello, A., Panaccio, A., De Momi, E., and Mattos, L. (2018b). Context-aware augmented reality for laparoscopy. In *Sixth national congress of bioengineering, Milan, Italy*, volume 25.
- Penza, V., Neri, A., Koskinopoulou, M., Turco, E., Soriero, D., Scabini, S., Prattichizzo, D., and Mattos, L. S. (2023). Augmented reality navigation in robot-assisted surgery with a teleoperated robotic endoscope. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4621–4626.
- Penza, V., Soriero, D., Barresi, G., Pertile, D., Scabini, S., and Mattos, L. S. (2020b). The gps for surgery: A user-centered evaluation of a navigation system for laparoscopic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 16(5):1–13.
- Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., Browne, W. J., Clark, A., Cuthill, I. C., Dirnagl, U., et al. (2020). The arrive guidelines 2.0: Updated guidelines for reporting animal research. *Journal of Cerebral Blood Flow & Metabolism*, 40(9):1769–1777.
- Pfeiffer, M., Riediger, C., Leger, S., Kühn, J.-P., Seppelt, D., Hoffmann, R.-T., Weitz, J., and Speidel, S. (2020). Non-rigid volume to surface registration using a data-driven biomechanical model. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*, pages 724–734. Springer.
- Pfeiffer, M., Riediger, C., Weitz, J., and Speidel, S. (2019). Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks. *International journal of computer assisted radiology and surgery*, 14:1147–1155.
- Piccinelli, N., Roberti, A., Tagliabue, E., Setti, F., Kronreif, G., Muradore, R., and Fiorini, P. (2020). Rigid 3d registration of pre-operative information for semi-autonomous surgery. In *2020 International Symposium on Medical Robotics (ISMR)*, pages 139–145. IEEE.
- Pieper, S., Halle, M., and Kikinis, R. (2004). 3d slicer. In *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*, pages 632–635. IEEE.
- Pokala, N., Delaney, C. P., Kiran, R. P., Bast, J., Angermeier, K., and Fazio, V. W. (2007). A randomized controlled trial comparing simultaneous intra-operative vs sequential prophylactic ureteric catheter insertion in re-operative and complicated colorectal surgery. *International journal of colorectal disease*, 22:683–687.

- Prasad, K., Fassler, C., Miller, A., Aweeda, M., Pruthi, S., Fusco, J. C., Daniel, B., Miga, M., Wu, J. Y., and Topf, M. C. (2024). More than meets the eye: Augmented reality in surgical oncology. *Journal of Surgical Oncology*, 130(3):405–418.
- Prevost, G. A., Eigl, B., Paolucci, I., Rudolph, T., Peterhans, M., Weber, S., Beldi, G., Candinas, D., and Lachenmayer, A. (2020). Efficiency, accuracy and clinical applicability of a new image-guided surgery system in 3d laparoscopic liver surgery. *Journal of gastrointestinal surgery*, 24:2251–2258.
- Puerto-Souza, G. A., Cadeddu, J. A., and Mariottini, G.-L. (2014). Toward long-term and accurate augmented-reality for monocular endoscopic videos. *IEEE Transactions on Biomedical Engineering*, 61(10):2609–2620.
- Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660.
- Qian, L., Wu, J. Y., DiMaio, S. P., Navab, N., and Kazanzides, P. (2019). A review of augmented reality in robotic-assisted surgery. *IEEE Transactions on Medical Robotics and Bionics*, 2(1):1–16.
- Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Ilic, S., Hu, D., and Xu, K. (2023a). Geo-transformer: Fast and robust point cloud registration with geometric transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9806–9821.
- Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., and Xu, K. (2022). Geometric transformer for fast and robust point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11143–11152.
- Qin, Z., Yu, H., Wang, C., Peng, Y., and Xu, K. (2023b). Deep graph-based spatial consistency for robust non-rigid point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5394–5403.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A. Y., et al. (2009). Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan.
- Reichard, D., Häntsch, D., Bodenstedt, S., Suwelack, S., Wagner, M., Kenngott, H., Müller-Stich, B., Maier-Hein, L., Dillmann, R., and Speidel, S. (2017). Projective biomechanical depth matching for soft tissue registration in laparoscopic surgery. *International Journal of Computer Assisted Radiology and Surgery*, 12:1101–1110.
- Rivas-Blanco, I., Pérez-Del-Pulgar, C. J., García-Morales, I., and Muñoz, V. F. (2021). A review on deep learning in minimally invasive surgery. *IEEE Access*, 9:48658–48678.
- Robu, M. R., Ramalhinho, J., Thompson, S., Gurusamy, K., Davidson, B., Hawkes, D., Stoyanov, D., and Clarkson, M. J. (2018). Global rigid registration of ct to video in laparoscopic liver surgery. *International journal of computer assisted radiology and surgery*, 13:947–956.
- Rosset, A., Spadola, L., and Ratib, O. (2004). Osirix: an open-source software for navigating in multidimensional dicom images. *Journal of digital imaging*, 17:205–216.

- Rucker, D. C., Wu, Y., Clements, L. W., Ondrake, J. E., Pheiffer, T. S., Simpson, A. L., Jarnagin, W. R., and Miga, M. I. (2013). A mechanics-based nonrigid registration method for liver surgery using sparse intraoperative data. *IEEE transactions on medical imaging*, 33(1):147–158.
- Rusu, R. B., Blodow, N., and Beetz, M. (2009). Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE international conference on robotics and automation*, pages 3212–3217. IEEE.
- Rusu, R. B., Blodow, N., Marton, Z. C., and Beetz, M. (2008). Aligning point cloud views using persistent feature histograms. In *2008 IEEE/RSJ international conference on intelligent robots and systems*, pages 3384–3391. IEEE.
- Safety, W. P., Organization, W. H., et al. (2009). *WHO guidelines for safe surgery 2009: safe surgery saves lives*. Number WHO/IER/PSP/2008.08-1E. World Health Organization.
- Schneider, C., Thompson, S., Totz, J., Song, Y., Allam, M., Sodergren, M., Desjardins, A., Barratt, D., Ourselin, S., Gurusamy, K., et al. (2020). Comparison of manual and semi-automatic registration in augmented reality image-guided liver surgery: a clinical feasibility study. *Surgical endoscopy*, 34:4702–4711.
- Seetohul, J., Shafiee, M., and Sirlantzis, K. (2023). Augmented reality (ar) for surgical robotic and autonomous systems: state of the art, challenges, and solutions. *Sensors*, 23(13):6202.
- Sengiku, A., Koeda, M., Sawada, A., Kono, J., Terada, N., Yamasaki, T., Mizushino, K., Kunii, T., Onishi, K., Noborio, H., et al. (2017). Augmented reality navigation system for robot-assisted laparoscopic partial nephrectomy. In *Design, User Experience, and Usability: Designing Pleasurable Experiences: 6th International Conference, DUXU 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II 6*, pages 575–584. Springer.
- Shen, Y., Hui, L., Jiang, H., Xie, J., and Yang, J. (2022). Reliable inlier evaluation for unsupervised point cloud registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2198–2206.
- Singla, R., Edgcumbe, P., Pratt, P., Nguan, C., and Rohling, R. (2017). Intra-operative ultrasound-based augmented reality guidance for laparoscopic surgery. *Healthcare technology letters*, 4(5):204–209.
- Smits, R., Bruyninckx, H., and Aertbeliën, E. (2011). Kdl: Kinematics and dynamics library.
- Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J.-B., Moreau, J., Osswald, A.-B., Bouhadjar, M., and Marescaux, J. (2010). 3d image reconstruction for comparison of algorithm database.
- Soriero, D., Batistotti, P., Malinaric, R., Pertile, D., Massobrio, A., Epis, L., Sperotto, B., Penza, V., Mattos, L. S., Sartini, M., et al. (2022). Efficacy of high-resolution preoperative 3d reconstructions for lesion localization in oncological colorectal surgery—first pilot study. In *Healthcare*, volume 10, page 900. MDPI.

- Sorkine, O. and Alexa, M. (2007). As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, volume 4, pages 109–116. Citeseer.
- Stoyanov, D., Scarzanella, M. V., Pratt, P., and Yang, G.-Z. (2010). Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20–24, 2010, Proceedings, Part I* 13, pages 275–282. Springer.
- Su, L.-M., Vagvolgyi, B. P., Agarwal, R., Reiley, C. E., Taylor, R. H., and Hager, G. D. (2009). Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-ct to stereoscopic video registration. *Urology*, 73(4):896–900.
- Suwelack, S., Röhl, S., Bodenstedt, S., Reichard, D., Dillmann, R., dos Santos, T., Maier-Hein, L., Wagner, M., Wünscher, J., Kennigott, H., et al. (2014). Physics-based shape matching for intraoperative image guidance. *Medical physics*, 41(11):111901.
- Talbot, H., Haouchine, N., Peterlik, I., Dequidt, J., Duriez, C., Delingette, H., and Cotin, S. (2015). Surgery Training, Planning and Guidance Using the SOFA Framework. In Hege, H.-C. and Ropinski, T., editors, *Eurographics 2015 - Dirk Bartz Prize*. The Eurographics Association.
- Tanzi, L., Piazzolla, P., Porpiglia, F., and Vezzetti, E. (2021). Real-time deep learning semantic segmentation during intra-operative surgery for 3d augmented reality assistance. *International Journal of Computer Assisted Radiology and Surgery*, 16(9):1435–1445.
- Taubin, G. (1995). A signal processing approach to fair surface design. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 351–358.
- Taylor, R. H., Menciassi, A., Fichtinger, G., Fiorini, P., and Dario, P. (2016). Medical robotics and computer-integrated surgery. *Springer handbook of robotics*, pages 1657–1684.
- Teatini, A., Brunet, J.-N., Nikolaev, S., Edwin, B., Cotin, S., and Elle, O. J. (2020). Use of stereo-laparoscopic liver surface reconstruction to compensate for pneumoperitoneum deformation through biomechanical modeling. In *VPH2020-Virtual Physiological Human*.
- Teatini, A., Pérez de Frutos, J., Eigl, B., Pelanis, E., Aghayan, D. L., Lai, M., Kumar, R. P., Palomar, R., Edwin, B., and Elle, O. J. (2021). Influence of sampling accuracy on augmented reality for laparoscopic image-guided surgery. *Minimally Invasive Therapy & Allied Technologies*, 30(4):229–238.
- Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., and Guibas, L. J. (2019). Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6411–6420.
- Thompson, S., Schneider, C., Bosi, M., Gurusamy, K., Ourselin, S., Davidson, B., Hawkes, D., and Clarkson, M. J. (2018). In vivo estimation of target registration errors during augmented reality laparoscopic surgery. *International journal of computer assisted radiology and surgery*, 13:865–874.

- Udupa, J. K., Odhner, D., Samarasekera, S., Goncalves, R. J., Iyer, K., Venugopal, K. P., and Furuie, S. S. (1994). 3dviewnix: an open, transportable multidimensional, multimodality, multiparametric imaging software system. In *Medical Imaging 1994: Image Capture, Formatting, and Display*, volume 2164, pages 58–73. SPIE.
- Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(04):376–380.
- Wang, D., Azadvar, S., Heiselman, J., Jiang, X., Miga, M., and Wang, L. (2024). Libr+: Improving intraoperative liver registration by learning the residual of biomechanics-based deformable registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 359–368. Springer.
- Wang, R., Geng, Z., Zhang, Z., Pei, R., and Meng, X. (2017). Autostereoscopic augmented reality visualization for depth perception in endoscopic surgery. *Displays*, 48:50–60.
- Wang, Y., Long, Y., Fan, S. H., and Dou, Q. (2022a). Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In *MICCAI*, pages 431–441.
- Wang, Y. and Solomon, J. M. (2019a). Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3523–3532.
- Wang, Y. and Solomon, J. M. (2019b). Prnet: Self-supervised learning for partial-to-partial registration. *Advances in neural information processing systems*, 32.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5):1–12.
- Wang, Y., Yan, C., Feng, Y., Du, S., Dai, Q., and Gao, Y. (2022b). Storm: Structure-based overlap matching for partial point cloud registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1135–1149.
- Website (2024). 3dslicer. <https://www.slicer.org/> [Accessed: 7/5/2024].
- Wei, R., Li, B., Mo, H., Lu, B., Long, Y., Yang, B., Dou, Q., Liu, Y., and Sun, D. (2022). Stereo dense scene reconstruction and accurate localization for learning-based navigation of laparoscope in minimally invasive surgery. *IEEE Transactions on Biomedical Engineering*, 70(2):488–500.
- Wen, B., Trepte, M., Aribido, J., Kautz, J., Gallo, O., and Birchfield, S. (2025). Foundation-stereo: Zero-shot stereo matching. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5249–5260.
- White, L. A., Joseph, J. P., Yang, D. Y., Kelley, S. R., Mathis, K. L., Behm, K., and Viers, B. R. (2021). Intraureteral indocyanine green augments ureteral identification and avoidance during complex robotic-assisted colorectal surgery. *Colorectal Disease*, 23(3):718–723.

- Wilson, M. R., Poolton, J. M., Malhotra, N., Ngo, K., Bright, E., and Masters, R. S. (2011). Development and validation of a surgical workload measure: the surgery task load index (surg-tlx). *World journal of surgery*, 35:1961–1969.
- Wu, Y., Yao, Q., Fan, X., Gong, M., Ma, W., and Miao, Q. (2023a). Panet: A point-attention based multi-scale feature fusion network for point cloud registration. *IEEE Transactions on Instrumentation and Measurement*.
- Wu, Y., Zhang, Y., Fan, X., Gong, M., Miao, Q., and Ma, W. (2022). Inenet: Inliers estimation network with similarity learning for partial overlapping registration. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(3):1413–1426.
- Wu, Y., Zhang, Y., Ma, W., Gong, M., Fan, X., Zhang, M., Qin, A., and Miao, Q. (2023b). Rornet: Partial-to-partial registration network with reliable overlapping representations. *IEEE Transactions on Neural Networks and Learning Systems*.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. (2015). 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920.
- Xu, D., Jiang, Y., Wang, P., Fan, Z., Shi, H., and Wang, Z. (2022a). Sinnerf: Training neural radiance fields on complex scenes from a single image. In *European Conference on Computer Vision*, pages 736–753. Springer.
- Xu, H., Liu, S., Wang, G., Liu, G., and Zeng, B. (2021). Omnet: Learning overlapping mask for partial-to-partial point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3132–3141.
- Xu, H., Ye, N., Liu, G., Zeng, B., and Liu, S. (2022b). Finet: Dual branches feature interaction for partial-to-partial point cloud registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2848–2856.
- Yang, G., Manela, J., Happold, M., and Ramanan, D. (2019). Hierarchical deep stereo matching on high-resolution images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5515–5524.
- Yang, J., Li, H., Campbell, D., and Jia, Y. (2015). Go-icp: A globally optimal solution to 3d icp point-set registration. *IEEE transactions on pattern analysis and machine intelligence*, 38(11):2241–2254.
- Yang, Z., Simon, R., and Linte, C. A. (2023). Learning feature descriptors for pre-and intra-operative point cloud matching for laparoscopic liver registration. *International journal of computer assisted radiology and surgery*, 18(6):1025–1032.
- Yang, Z., Simon, R., Merrell, K., and Linte, C. A. (2024). Boundary constraint-free biomechanical model-based surface matching for intraoperative liver deformation correction. *IEEE Transactions on Medical Imaging*.
- Yasuda, J., Okamoto, T., Onda, S., Futagawa, Y., Yanaga, K., Suzuki, N., and Hattori, A. (2018). Novel navigation system by augmented reality technology using a tablet pc for hepatobiliary and pancreatic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 14(5):e1921.

- Yew, Z. J. and Lee, G. H. (2020). Rpm-net: Robust point matching using learned features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11824–11833.
- Yew, Z. J. and Lee, G. H. (2022). Regtr: End-to-end point cloud correspondences with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6677–6686.
- Yoo, J., Uh, Y., Chun, S., Kang, B., and Ha, J.-W. (2019). Photorealistic style transfer via wavelet transforms. In *IEEE/CVF ICCV*, pages 9036–9045.
- Yu, A., Ye, V., Tancik, M., and Kanazawa, A. (2021a). pixelnerf: Neural radiance fields from one or few images. In *CVPR*, pages 4578–4587.
- Yu, H., Li, F., Saleh, M., Busam, B., and Ilic, S. (2021b). Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34:23872–23884.
- Yuan, W., Eckart, B., Kim, K., Jampani, V., Fox, D., and Kautz, J. (2020). Deepgmr: Learning latent gaussian mixture models for registration. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 733–750. Springer.
- Yuksel, C. (2015). Sample elimination for generating poisson disk sample sets. In *Computer Graphics Forum*, volume 34, pages 25–32. Wiley Online Library.
- Zeng, Y., Qian, Y., Zhu, Z., Hou, J., Yuan, H., and He, Y. (2021). Corrnnet3d: Unsupervised end-to-end learning of dense correspondence for 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6052–6061.
- Zha, R., Cheng, X., Li, H., Harandi, M., and Ge, Z. (2023). Endosurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos. In *International conference on medical image computing and computer-assisted intervention*, pages 13–23. Springer.
- Zhang, X., Otoo, E.-M., Fan, Y., Tao, C., Wang, T., and Rhode, K. (2022). Autostereoscopic 3d augmented reality navigation for laparoscopic surgery: A preliminary assessment. *IEEE Transactions on Biomedical Engineering*, 70(4):1413–1421.
- Zhang, X., Wang, J., Wang, T., Ji, X., Shen, Y., Sun, Z., and Zhang, X. (2019). A markerless automatic deformable registration framework for augmented reality navigation of laparoscopy partial nephrectomy. *International journal of computer assisted radiology and surgery*, 14:1285–1294.
- Zhang, X., Wang, T., Zhang, X., Zhang, Y., and Wang, J. (2020). Assessment and application of the coherent point drift algorithm to augmented reality surgical navigation for laparoscopic partial nephrectomy. *International journal of computer assisted radiology and surgery*, 15(6):989–999.
- Zhang, Y., Zou, Y., and Liu, P. X. (2024). Point cloud registration in laparoscopic liver surgery using keypoint correspondence registration network. *IEEE Transactions on Medical Imaging*.

-
- Zhao, Y., Zhang, J., Xu, S., and Ma, J. (2024). Deep learning-based low overlap point cloud registration for complex scenario: The review. *Information Fusion*, 107:102305.
- Zhou, J., Gao, B., Wang, K., Pei, J., Heng, P.-A., and Qin, J. (2025). Landmark-free preoperative-to-intraoperative registration in laparoscopic liver resection. *IEEE Transactions on Medical Imaging*.
- Zhou, Q.-Y., Park, J., and Koltun, V. (2018). Open3D: A modern library for 3D data processing. *arXiv:1801.09847*.