Javier Pérez de Frutos

Registration techniques for navigation in laparoscopic surgery

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Javier Pérez de Frutos

Registration techniques for navigation in laparoscopic surgery

Thesis for the Degree of Philosophiae Doctor

Trondheim, December 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

ABSTRACT

In minimally invasive surgery, where a direct sight of the treated organ or structures is not feasible but with a endoscope, image guided systems have been developed to provide the surgeon with real time information of the surgical area, in the form of images of the anatomy of the patient, and surgical navigation. In order to do so, image-to-patient and image-to-image registration techniques are required to bring physical and image information on to the same system of reference. Essentially, both methods rely on point-to-point correspondence, using a collection of selected features on both the fixed and target domains. In the case of image-to-patient, position landmarks are used. Whereas in image-to-image, features related to the intensity information of the image, or the topology of segmentations, are the usual approach. Nevertheless, these procedures require time and are susceptible to human error. In image-to-patient in particular, when sampling landmarks on the patient anatomy. Image-to-image methods on the other hand, operate between different images and virtual models, being complex to implement and to achieve good results, especially in multi-modal setups.

The presented research is framed in the field of hepatocellular carcinoma surgical treatment and colorectal liver metastasis, the most common types of liver cancer. Wedge resection is the usual choice to remove the cancerous cells, aiming to spare as much healthy tissue as possible while ensuring a complete extraction of the tumour. However, during the intervention, the organ undergoes major deformation due to mobilisation and detachment of the abdominal wall. This deformation is not reflected in the pre-operative images, showing an inaccurate location of relevant anatomical structures as well as the target lesion. Through a laparoscope and laparoscopic ultrasound, the surgeon can build a mental image of the anatomy of the

patient during the intervention, adding extra strain on the practitioner. Registration techniques can leverage the procedure outcome and ease the surgery by aligning the virtual models of the patient to the situation on the surgical table. Furthermore, by updating the pre-operative model with the real-time information acquired when inspecting the organ.

In particular, the present study focuses on the use of registration for navigation in minimally invasive interventions, including image-to-patient and image-to-image registration. Main contributions include the evaluation of tracking technologies for surgical navigation; a novel image-to-patient registration method (single landmark registration method) which can take advantage of the laparoscopic ultrasound to improve the registration of the pre-operative images; the use of deep learning for image-to-image registration in medical applications, and development of new training methods; and the research on the influence of human accuracy in landmark sampling during image-to-patient registration, in augmented reality applications for surgical navigation.

ACKNOWLEDGEMENTS

I would personally like to express my deepest gratitude to my supervisors. First and foremost, to Dr. Thomas Langø, for his guidance and friendship during these almost-never-ending years. Dr. Langø is a brilliant scientist, great mentor, and valuable friend. His unwavering optimism and skill to play-down any situation, no matter how daring or challenging it could be, inspired me to fulfil this work. To Prof. Frank Lindseth, whose advice and discussions brought inspiring ideas which culminated in publications. And last but not least, to Prof. Ole Jakob Elle, who led the HiPerNav project, promoting fruitful collaboration between the ESRs, and between the institutions. Also I would like to thank my co-supervisor Dr. Ingerid Reinertsen, for always keeping and eye on my progress, and taking the time to share new interesting publications or challenges she finds. Lastly to my mentors Dr. Erlend F. Hofstad and Ronald Mårvik, MD, much obliged for the collaboration.

To my fellow colleagues, both ESRs and supervisors, of the HiPerNav project, with whom I had the pleasure of having embolden discussions and fruitful exchange of ideas. To Dr. Andrea Teatini, Dr. Benjamin Eigl, Egidijus Pelanis, MD, Dr. Iwan Paolucci, Dr. Rafael Palomar, Prof. Arjo Loeve, Prof. Joaquín Olivares,

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI Artificial intelligence.
ANTs Advanced normalisation tools.
AR Augmented reality.

BL Baseline.

CC Cross-correlation.
CLM Colorectal liver metastases.
CNN Convolutional neural network.
CPD Coherent point drift.
CT Computed tomography.

DDMR Deep deformation map registration.
DOF Degrees of freedom.
DSC Dice similarity coefficient.

EMTS Electromagnetic tracking system.

FLBR Fiducial landmark-based registration method.
FLE Fiducial localisation error.
FRE Fiducial registration error.

GMM Gaussian mixture model.

GPU Graphical processing unit.

HCC Hepatocellular carcinoma.
HD Hausdorff distance.
HD95 95th percentile of the Hausdorff distance.
HMD Head-mounted display.
HUD Heads-up display.

ICP Iterative closest point.
IGS Image-guided system.
IXI Information extraction from images dataset.

LUS Laparoscopic ultrasound.

MAE Mean absolute error.
ML Machine learning.
MLP Multilayer preceptron.
MR Magnetic resonance.
MSE Mean squared error.
MS-SSIM Multi-sale structural similarity index metric.

NCC Normalised cross-correlation.

Oslo-CoMet Oslo-CoMet randomised controlled trial dataset.

OTS Optical tracking system.

PBR Point-based registration method.

RANSAC Random sample consensus.

ReLU Rectified linear unit.

SG Segmentation guidance.

SLRM Single landmark registration method.

SSIM Structural similarity index metric.

STN Spatial transformer network.

SVD Singular value decomposition.

SyN Symmetric normalisation.

SyNN Symmetric normalisation with normalised cross-correlation.

TPS Thin plate spline.

TRE Target registration error.

US Ultrasound.

UW Uncertainty weighting.

CHAPTER 1 INTRODUCTION

Mortality of liver cancer has risen in time, being the third most lethal cancer according to the 2020 Global Cancer Observatory database[1] summing to 830,180 cases worldwide (ages from 0 to 85). Preceded by colorectal cancer (935,173) and lung cancer (1,796,144). The preferred treatment for liver cancer is resection [1, 2], which can be done through open surgery or minimally invasive surgery. Minimally invasive surgery is becoming the preferred approach for hepatocellular carcinoma (HCC) and colorectal liver metastases (CLM) [3, 4], proven to be a reliable alternative to the traditional open surgery [5, 1, 2], with fewer complications and inconveniences for the patient. Through small incisions on the abdominal wall, the surgeon can access the liver and perform a complete resection without the need of open surgery. Main advantages of this approach include, but are not limited to, short recovery time, no organ adhesion, short hospitalisation, and fewer blood loss [6, 5, 7].

However, the eligibility of a patient for minimally invasive treatment is dependant on the size, number, and distribution of the lesions, the accessibility and risk of conversion to open surgery, or the experience of the surgeon in minimally invasive surgery [4, 8]. From a procedural point of view, the lack of direct sight of the surgical area is the biggest challenge of minimally invasive treatment. A common approach to overcome this difficulty is the use of a laparoscope inserted through one of the trocars, enabling the surgeon to see what is happening within the abdomen [4]. Nevertheless, this camera provides a monocular view of the area. More advanced cameras provide an stereoscopic view with depth information, yet

---

[1]https://gco.iarc.fr/

from a single point of view. This situation poses an additional burden to the surgeon, as the clinician must make a mental image with the location of the tools, the anatomical structures, and the view given by the camera. Underlying structures are also hidden from the laparoscope view, relying on intra-operative imaging techniques like laparoscopic ultrasound (LUS), intra-operative computed tomography (CT), or fluoroscopy.
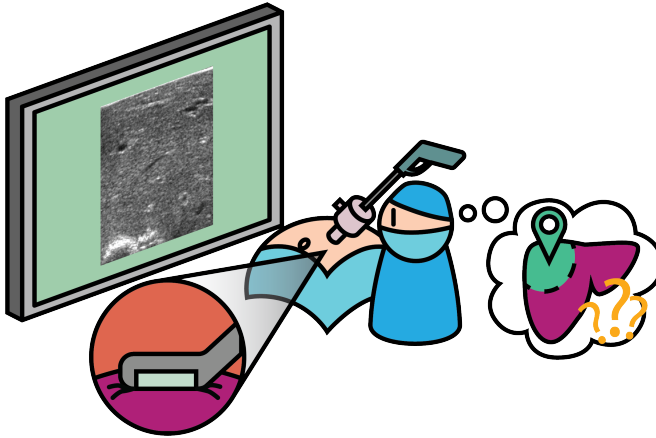
Figure 1.1: Clinical setup with no surgical navigation. The surgeon is presented with the laparoscopic ultrasound (LUS) image and the location of the probe must be estimated by the surgeon on the fly.

In order to ease the surgical tasks to the surgeons, image-guided systems (IGS) were developed to provide location information required during the intervention [9]. These systems bring together the virtual models of the patient and the tracked location of the tools through the registration of landmarks or images, to give the surgeon a virtual representation of the situation within the surgical area. To achieve IGS navigation, registration techniques are used bringing together the tools and virtual data onto the same reference system. Hence, the clinician now has a direct view of the relative placement of the tools with respect of the patient. Even more, the surgeon can navigate to the desired location using the information provided by the IGS [9].

Registration algorithms for clinical applications can be classified into two major groups: image-to-patient and image-to-image. Whereas the former relates image and physical coordinates, aligning the virtual model of the patient with the location
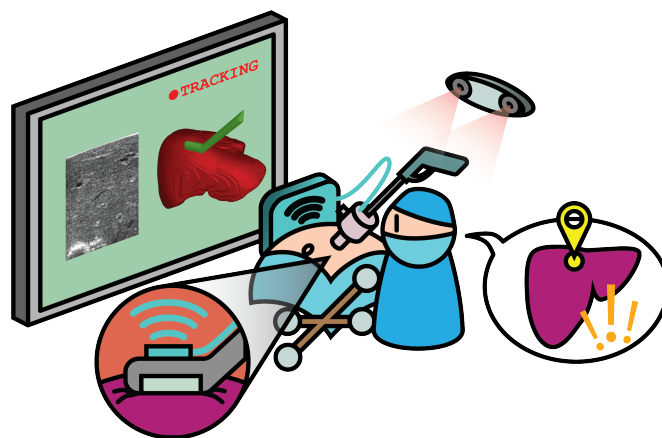
Figure 1.2: Clinical setup with surgical navigation. Both electromagnetic and optical tracking systems are available for tracking the flexible laparoscopic ultrasound (LUS) probe and the patient respectively. The surgeon is presented with both the LUS image and the tracked location of the probe, displayed together with a 3D virtual model of the patient.

of this in the surgical theatre; image-to-image registration line up images of the same anatomical structure taken at different time or by various techniques. Medical imaging includes several modalities which provide complementary information: computed tomography, X-ray, angiography, magnetic resonance (MR) imaging , video from a laparoscope, fluoroscopy, ultrasound (US), or segmented 3D virtual models. Image-to-patient registration algorithms are then used to align and match the images and the virtual model of the patient with the corresponding physical landmarks on the patient, enabling intra-operative navigation. Image-to-image registration methods are employed instead to combine the different image modalities onto the same reference frame, so that corresponding points match.

Unlike image-to-patient registration, image-to-image techniques can be employed at several stages of the surgical workflow. For instance, in the planning stage to align diagnostic images, like, diagnostic CT and MR scans to generate a more detailed model of the patient. During the intervention, so as to add information happening during the procedure to the pre-operative plan, adjust the resection margins, and locate neighbouring relevant anatomical structures. Or after the surgery, to assess and validate the outcome of the intervention by comparing the treated

area. [10, 11, 12]. In Figure 1.3 a schematic workflow showing the aforementioned stages is depicted, indicating where the two types of registration methods are most relevant.
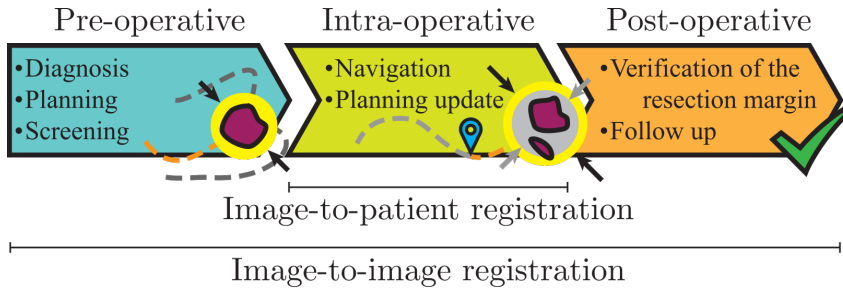
Figure 1.3: Schematic surgical workflow showing on which steps the registration techniques are more relevant.

Especially during surgery, registration methods are expected to provide a high level of precision to enable accurate navigation, assessment of the lesion and surrounding relevant structures. Resection margins for liver resection can be as small as 5 mm to 10 mm [11, 13, 14]. However, resection margins of under 1 mm with no recurrence within 5-years after the resection have been reported in the literature [15, 3]. Achieving this level of precision is a a challenging endeavour for registration algorithms, as mobilisation of the organ, detachment of the ligaments, and the effect of the pneumoperitoneum can add up to a deformation of 44.6 mm in average [16].

1.1 Research scope and contributions

The presented research dwells into the use of registration during liver resection surgery. Exploring the different aspects of registration in surgical applications, starting with image-to-patient registration, and following with image-to-image, where deep learning solutions have been studied. In addition, the influence of human error in the registration process was observed, applied to the use of augmented reality for surgical navigation. Special emphasis was put on the image-to-image registration, as these techniques can greatly improve the navigation experience by showing an accurate picture of the anatomy within the abdominal area.

Therefore, the following were the questions addressed in this research.

RQ 1: What key elements can be addressed to improve registration for surgical navigation?

RQ 2: What bottlenecks are to be addressed in deep learning for registration in medical applications?

RQ 3: How can registration be seamlessly integrated in the surgical workflow?

RQ 4: Is deep-learning image-to-image registration ready to be integrated in an IGS?

The articles included in this thesis, which aim to clarify the aforementioned questions, are the following:

Paper I [17]: A. Teatini, **J. Pérez de Frutos**, T. Langø, B. Edwin, O. J. Elle. "Assessment and comparison of target registration accuracy in surgical instrument tracking technologies". In: *Proceedings of the Annual International Conference of the IEEEEngineering in Medicine and Biology Society, EMBS*. IEEE, July 2018, pp. 1845–1848. doi: 10.1109/EMBC.2018.8512671

To enable navigation in a surgical setup, it is required to bring together both the physical space where the intervention takes place and the virtual representation of the patient. In this study, we compare the most extended tracking technologies for IGS navigation platforms. RQ1 and RQ3 are addressed in this paper, by studying the tracking technologies employed in modern IGS (RQ 1 and 3).

Paper II [18]: **J. Pérez de Frutos**, E. F. Hofstad, O. V. Solberg, G. A. Tangen, F. Lindseth , T. Langø, O. J. Elle, R. Mårvik. "Laboratory test of Single Landmark registration method for ultrasound-based navigation in laparoscopy using an open-source platform". In: *International Journal of Computer Assisted Radiology and Surgery* 13.12 (Dec. 2018), pp. 1927–1936. doi: 10.1007/s11548-018-1830-7

Image-to-patient registration is conventionally performed by sampling several pre-defined landmarks on the patient anatomy, resulting in a time-consuming step and prone to low accuracy. Single landmark registration method (SLRM) is proposed as a simpler and quicker technique to enable navigation for minimally invasive surgery, allowing for further refinement of the registration when locating the centre of the lesion. Therefore, RQ 1 is partially considered when comparing

the SLRM and the point-based registration (PBR). As well as RQ 3 when proposing the simpler registration procedure offered by the SLRM.

Paper III [19]:A. Teatini, **J. Pérez de Frutos**, B. Eigl, E. Pelanis, D. L. Aghayan, M. Lai, R. P Kumar, R. Palomar, B. Edwin, O. J. Elle. "Influence of sampling accuracy on augmented reality for laparoscopic image-guided surgery". In: *Minimally Invasive Therapy & Allied Technologies* (Mar. 2020), pp. 1–10. doi: 10.1080/13645706.2020.1727524

The accuracy by which the image and physical spaces are aligned is dependant on the skills of the surgeon when recording the landmarks used for the registration. Within the scope of augmented reality for surgical procedures, we evaluate the effect of the human factor during registration. Again, the analysis of the error components in registration aims to shred light onto RQ 1.

Paper IV [20]: **J. Pérez de Frutos**, A. Pedersen, E. Pelanis, D. Bouget S. Survarachakan, T. Langø, O. J. Elle, & F. Lindseth (2022). "Learning deep abdominal CT registration through adaptive loss weighting and synthetic data generation". PLOS ONE 18(2): e0282110. doi: 10.1371/journal.pone.0282110

Deep learning-based image registration remains a challenging task, not only due to the complexity of the problem itself, but also due to the lack of large enough training datasets. In this study, we explore training strategies for deep learning models, and propose weakly-supervision combined with uncertainty weighting (UW) for enhancing the learning of registration models. In addition, transfer learning is shown to enable the training on limited datasets starting from an already trained model. In this study, both RQ 2 and RQ 4 are answered by showing the possibilities of deep learning and the discussion on the limitations of these methods for clinical usage.

Other dissemination carried out during the time of this research includes:

- 32nd International Congress and Exhibition Computer Assisted Radiology and Surgery (Berlin, Germany, June 2018): presentation on Single Landmark Registration Method. Title "Laboratory test of Single Landmark registration method for ultrasound-based navigation in laparoscopy using an open-source platform".

Table 1.1: Contributions of J. Pérez de Frutos on each of the publications included in this thesis.

	Paper I	Paper II	Paper III	Paper IV
Experiment design	Minor	Main	Minor	Main
Software	Minor	N/A	Major	Main
Data collection	Major	Main	Minor	Minor
Data analysis	Major	Main	Minor	Major
Discussion	Major	Main	Minor	Main
Writing	Major	Main	Major	Main
Review	Major	Main	Major	Major
Other	Poster	Presentation	N/A	N/A

- 40th International Conference of the IEEE Engineering in Medicine and Biology Society (Honolulu, Hawaii, USA, July 2018): poster session. Title "Assessment and comparison of target registration accuracy in surgical instrument tracking technologies".
- Society Medical Image Technology congress (Heilbronn, Germany, October 2019): abstract and presentation on Deep Learning based image-to-image registration. Title "Deep Deformation Map Regressor".
- European Computer Assisted Liver Surgery Society meeting (Bern, Switzerland, October 2019): poster session. Title "Up-to-Date Anatomical Model".
- Society Medical Image Technology congress (Oslo, Norway, May 2022): presentation on Deep Learning image-to-image registration. Title: "Image-to-image registration: teaching AIs to deform heads".

1.2 Thesis structure

The present thesis is organised as follows:

- Chapter 1: introduction to the scope of the presented research.
- Chapter 2: introduces mathematical and conceptual background relevant for the content of the thesis.
- Chapter 3: explores the tracking technologies used to enable navigation in IGS (Article I [17]).

- Chapter 4: introduces image to patient registration and the Single Landmark Registration Method (Article II [18]).
- Chapter 5: contains the research on image-to-image registration, and in particular on deep learning based image registration (Article IV [20]).
- Chapter 6: explores the use of augmented reality for surgical interventions, where registration plays a major role (Article III [19]).
- Chapter 7: discussion over the research outcomes.
- Chapter 8: closing remarks.
- Appendix **??**: the collection of articles produced as part of the presented research.

CHAPTER 2 FUNDAMENTALS

This section provides an introduction to relevant tools and concepts used throughout the presented research. Further insight on the different topics can be found in the provided references.

2.1 Linear transformations

In IGSs, every tracked element has a coordinate system associated to it. Linear transformations are thus used to relate the coordinates of one tool with respect to other coordinate system. These transformations map vector spaces onto vector spaces [21], this is, coordinates in a given coordinate system onto a different one, and thus, allows to represent different objects e.g., the location of surgical tools, and image coordinates, onto the same space. In IGS, this transformation is considered to be a linear transformation between subspaces. A Cartesian coordinate system will always be used to represent the physical location of the objects, and it will be noted as $\mathbb{R}^n$ i.e., a n-dimensional Cartesian space of real coordinates. The notation to describe a linear transformation followed in this document is shown in Equation 2.1.

$$\mathrm{T}_\beta^\alpha : \mathbb{R}^{d_\alpha} \to \mathbb{R}^{d_\beta} \tag{2.1}$$

where α is the origin reference frame of dimensionality d_α and β is the target reference frame with dimensionality d_β. This transformation can be read as the coordinates of objects in α as seen from the reference frame β. Rigid transformations are commonly represented in the form of augmented matrices which describe rotation, translation, scaling, sheering and perspective transformations, as shown in

Equation 2.2. This form is usually adopted together with the use of homogeneous coordinates, as it eases the computation by using matrix multiplications to apply these transformations [22, 21].

$$T_\beta^\alpha = \left(\begin{array}{c|c} A_{3x3} & t_{3x1} \\ \hline p_{1x3} & s_{1x1} \end{array} \right) \tag{2.2}$$

where A is the affine transformation matrix, t is a translation vector, p represents the perspective transformation and s is the uniform global scale factor.

Though in general no perspective nor global scaling is considered, simplifying the notation as follows:

$$T_\beta^\alpha = \left(\begin{array}{c|c} A_{3x3} & t_{3x1} \\ \hline 0_{1x3} & 1_{1x1} \end{array} \right) \tag{2.3}$$

The affine matrix A is further simplified, as it is common to limit the transformations to rotation and scaling along the two or three coordinate axes [21, 22].

As aforementioned, the location of every element e.g., the patient and the surgical tools, is represented by vectors with respect to the origin reference frame or world coordinates. This reference frame is usually associated with the tracking sensor i.e., the device used to track all the other objects in the operating room. Figure 2.1 shows the transformation matrices that relates different tracked objects, with respect to the tracking sensor, in this case, an optical tracking system (see Section 3.1).

Next is to bring the image data into the operating room. To do so, image-to-patient and image-to-image registration algorithms are required. These techniques are used to obtain the transformation matrix that relates the virtual image models to its corresponding physical location (T_I^P) (see Figure 2.2).

Once the transformations are known, these can be combined by regular matrix multiplication to relate the location of the different objects in the operating room, and hence, enable surgical navigation [23, 24].

For example, in the situation shown in Figure 2.2, the coordinates of the tool tip (T) in the reference frame of the virtual model (I) can be defined as shown in Equation 2.4.
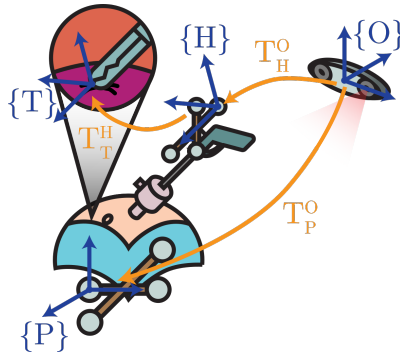
Figure 2.1: Transformations between reference frames before image-to-patient registration. The optical markers allow the sensor (O) to track the location of the patient (P) and the handler of the rigid surgical tool (H). Through calibration, the location of the tool tip (T) can also be tracked.
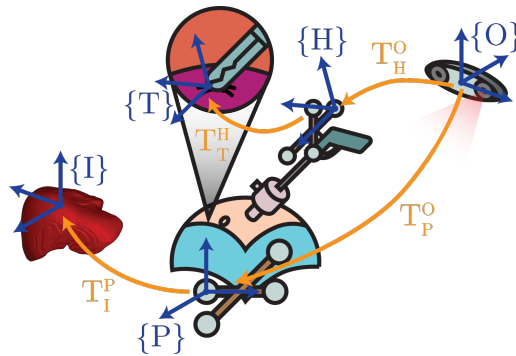
Figure 2.2: Transformations between reference frames after image-to-patient registration. Now the transformation between the patient (P) and the image data (I) is known, and the location of the tool tip (T) can be shown on the navigation platform relative to the virtual model.

$$T_T^I = \left(T_I^P\right)^{-1} \left(T_P^O\right)^{-1} T_H^O T_T^H \tag{2.4}$$

2.2 Registration error

Depending on the registration algorithm (image-to-patient or image-to-image), the accuracy of the method can be evaluated based on different metrics. While image-

to-image algorithms are evaluated in terms of geometrical properties of the images, but not limited to, e.g., region overlap or intensity distribution; image-to-patient algorithms are assessed mainly in terms of distances between the image and patient landmarks.

Point-based registration (PBR), or fiducial based registration, was explored for medical applications by Maurer *et. al.,* [25, 26], who also coined the terms fiducial localisation error (FLE), fiducial registration error (FRE) and the target registration error (TRE). These three metrics are used to assess the performance and accuracy of registration algorithms. FLE refers to the error in locating the fiducials i.e., error when pointing at the centre of a fiducial or when selecting points on images (Figure 2.3a); FRE is the error between corresponding fiducial points in the physical and virtual model, after correcting the latter as a result of the registration (Figure 2.3b); and TRE, equivalent to the FRE but for other points e.g., targets, than the fiducials used for the registration (Figure 2.3c). Out of the three measures, FRE is the easiest and most direct one to measure during an intervention [27, 26]. But it is TRE the one that matters the most, as it refers to the registration accuracy on the location of the lesion.

(a)   (b)   (c)
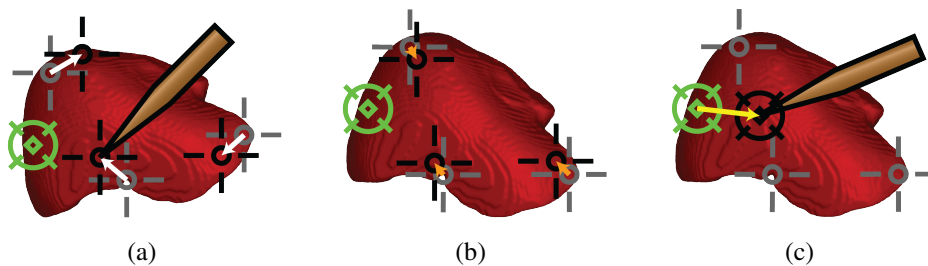
Figure 2.3: Fiducials in the virtual model are marked with grey cross hairs. The physical location, obtained through a tracked surgical marker, of those same fiducials are marked with the black cross hairs. The target in the virtual model is shown in green and its physical location with the same cross hairs but in black colour. (a) Fiducial localisation error (FLE) (white arrows), (b) fiducial registration error (FRE) after registration (orange arrows) and (c) target registration error (TRE) (yellow arrow).

For a set of fiducial points in the physical model (patient) $P = \{p_0, p_1, \ldots, p_n\}$ and the corresponding points on the virtual model $V = \{v_0, v_1, \ldots, v_n\}$, and a rigid transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ that relates these two sets, the FRE can be directly

computed as [25, 26]:

$$\text{FRE} = \sqrt{\frac{1}{n} \sum_{i=0}^{n} ||v_i - \text{T}(p_i)||^2} \tag{2.5}$$

Then, the FLE can be then estimated using Equation 2.6 [28].

$$\langle \text{FLE}^2 \rangle = \left(\frac{2n}{2n-1} \right) \langle \text{FRE}^2 \rangle \tag{2.6}$$

TRE can be computed in two different ways. The direct form would be as the distance between the location of the target point on the volume, after registration, and the sampled location. However unless a tracked tool is used to sample the location of the target point, during the intervention the TRE cannot be directly computed. Hence, the second form is through the Equation 2.7. This equation shows an approximation of the TRE using the FLE and the principal axes (j) of the spatial distribution of the fiducials [27, 29]. However, it should be noted that these two metrics TRE and FLE are uncorrelated [28], and thus, the direct approach should be preferred instead.

$$\langle |\text{TRE}(x)|^2 \rangle = \langle \text{FLE}^2 \rangle \left(\frac{1}{N} + \frac{1}{3} \sum_{j=1}^{3} \frac{\text{t}_j(x)^2}{\text{f}_j^2} \right) \tag{2.7}$$

where $\text{t}_j(x)$ is the distance of the target point x from the principal axis j and f_j is the root mean squared distance of the fiducials to that same principal axis.

2.3 Registration evaluation metrics

Both image-to-patient and image-to-image registration methods explained as an optimisation problem, where the aim is to correctly align two or more data sets. An objective or loss function is then used to correctly drive this alignment during the optimisation phase. This function describes the nature of the transformation e.g., affine, rigid, non-rigid, and how to evaluate the resulting registration. A well-defined objective function is of great importance when designing a machine learning (ML) pipeline, as the learning agent will rely on the output of the objective function to update the learning weights. Not only the objective function but a

regularisation term will drive the learning process of the algorithm. In addition to the objective function, the regularisation term is meant to impose restrictions on the output of the algorithm and/or the learning weights e.g., smoothness, normalisation, range of values, etc; to improve the chances of finding an optimal solution.

A thorough and exhaustive collection of objective and regularisation functions is out of the scope of this research, as it is highly dependent on the type of data and application; but a collection of operators that were considered relevant for this research. More information about this topic can be found in the following references [30, 31, 32, 33, 34, 35, 36].

2.4   Point set and image similarity metrics

Mean square error (MSE): measure of the distance between the elements of the sampled vectors using the squared Euclidean distance. As it is a point-wise operator, MSE is highly sensitive to the presence of outliers and to transformations between the images.

$$\text{MSE}(X,Y) = \frac{1}{N} \sum_{n=1}^{N} (x_i - y_i)^2 \tag{2.8}$$

Mean absolute error (MAE): equivalent to MSE but using the absolute distance between the vectors, i.e., the Manhattan distance.

$$\text{MAE}(X,Y) = \frac{1}{N} \sum_{n=1}^{N} |x_i - y_i| \tag{2.9}$$

Structural similarity index (SSIM): similarity measure between images that takes into account three components: luminance, contrast, and structure [37, 38]. Unlike MSE, SSIM is not a distance as it can have negative values and also it does not satisfy the identity of indiscernible axiom i.e., $\text{SSIM}(X,Y) = 0 \nleftrightarrow X = Y$, in fact $\text{SSIM}(X,Y) = 1 \leftrightarrow X = Y$. The metric evaluates three aspects of the images: luminance, contrast and structure. Luminance refers to the similarity in intensity (Equation 2.11), contrast to the distribution of the intensity (Equation 2.12), and structure to the correlation of between the intensity information of the images

(Equation 2.13).

$$\text{SSIM}(X,Y) = \text{l}(X,Y)^{\alpha} \cdot \text{c}(X,Y)^{\beta} \cdot \text{s}(X,Y)^{\gamma} \qquad (2.10)$$

where

$$\text{l}(X,Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \qquad (2.11)$$

$$\text{c}(X,Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \qquad (2.12)$$

$$\text{s}(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \qquad (2.13)$$

C_1, C_2 and C_3 are constants that take the form of: $C_1 = K_1L$, $C_2 = K_2L$ and $C_3 = C_2/L$, where L is the dynamic range of the image. The parameters α, β and γ weight the contribution of each of the three components, and are specific to the application. For simplicity these are commonly set to one, resulting in the usual formulation shown in Equation 2.14.

$$\text{SSIM}(X,Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \qquad (2.14)$$

In [37], Wang. Z., *et al.*, propose a multi-scale formulation of the SSIM (MS-SSIM), which is relevant, for example, in multi-scale deep learning architectures. For M levels, the MS-SSIM is shown in Equation 2.15

$$\text{MS-SSIM}(X,Y) = \text{l}_M(X,Y)^{\alpha_M} \prod_{k=1}^{M} \text{c}_k(X,Y)^{\beta_k} \cdot \text{s}_k(X,Y)^{\gamma_k} \qquad (2.15)$$

Cross-correlation (CC) or correlation coefficient: quantifies the correlation of two signals (f and g) as a function of the shift between these signals [30, 39]. Provides information of which delay or offset makes both signals best overlap, matching with a peak value of the CC function. It is calculated as the convolution of one signal over the other as shown in Equation 2.16 [40]. This formulation makes it convenient to be analysed in the Fourier domain, as convolution operations turn into regular multiplications.

$$\text{CC}_{\text{f,g}}(\tau) = (f * g)(\tau) = \int_{-\infty}^{\infty} \overline{f(t)}g(t + \tau)dt \qquad (2.16)$$

With two discrete signals $f[t]$ and $g[t]$, the discrete cross-correlations is as shown in Equation 2.17 [40].

$$dCC_{f,g}(\tau) = \frac{\sum_{t=-\infty}^{\infty} f[t]g[t-\tau]}{\sqrt{\sum_{t=-\infty}^{\infty} f[t]f[t-\tau]}\sqrt{\sum_{t=-\infty}^{\infty} g[t]g[t-\tau]}} \qquad (2.17)$$

Normalised cross-correlation (NCC): the main issue of the CC is that the signals to compare are assumed to have the same range, hence signals of different nature cannot be directly compared with normal CC. NCC normalises the signals so the ranges of these are the comparable. The formulation for discrete signals, relevant for image processing, is as follows:

$$dNCC_{f,g}(\tau) = \frac{\sum_{t=-\infty}^{\infty} (f[t] - \bar{f})(g[t-\tau] - \bar{g})}{\sqrt{Var(f)}\sqrt{Var(g)}} \qquad (2.18)$$

Hausdorff distance (HD) measures how far two sets (X and Y) are from each other, according to a certain distance metric ($d(\cdot, \cdot)$), typically the Euclidean (Equation 2.19). The HD is useful to evaluate the proximity of two shapes or point sets, specially in registration algorithms, because it also provides information about the direction in which the two sets will match. However, it is limited to binary images e.g., segmentations [41, 42].

$$h(X,Y) = \max_{x \in X}(\min_{y \in Y} d(x,y)) \qquad (2.19)$$

This metric can be understood as the greatest distance between a set and the closest point to the second set. Because this function is asymmetric t i.e., $h(X,Y) \neq h(X,Y)$, the generalised form shown in Equation 2.20 is used instead [43, 42].

$$HD(X,Y) = max\{h(X,Y), h(Y,X)\} \qquad (2.20)$$

However, because this formulation is quite sensitive to outliers, the 95th percentile of the HD (HD95) is a preferred metric [44, 45, 46, 47].

The original HD requires knowing the coordinates of the points at the boundary, unless this information is known from before, the operation of getting these coordinates cannot be used in a artificial intelligence (AI) training pipeline, as it prevent

the backpropagation of the gradients back to the network weights. A morphological erosion-based HD is proposed in [48] and shown in Equation 2.21. Here, the distance is approximated by the size of the difference of the two regions, after a series of erosion operations. In [48] HD and erosion-based HD were compared using two datasets of segmented prostate and brain white matter MR scans, reporting a Pearson correlation coefficient for the fitted linear function of the sampled values of 0.83 with respect to the exact HD, in both datasets.

$$\mathrm{HD}_{\mathrm{erosion}}(X,Y) = \frac{1}{||\Omega||} \sum_{k=1}^{K} \sum_{\Omega} \left((X-Y)^2 \ominus_k \mathscr{E} \right) k^{\alpha} \qquad (2.21)$$

were

Ω is the grid where X and Y are defined, and $||\Omega||$ is the size of this grid,

K is an hyper-parameter representing the number of successive erosion operations ($\ominus_k$),

$\mathscr{E}$ is the cross-shaped erosion kernel whose elements add to one; the erosion is implemented as a convolution with the kernel followed by a soft threshold,

and α is the second hyper-parameter that weights the effect of the largest difference $X - Y$.

After testing, the hyper-parameters K is set to one, whereas α is set to two, following [48]. Notice that this approximation represents a ratio, as it is scaled to the size of the image and dimensionless.

Dice score coefficient (DSC): measure of the similarity between two samples based on the degree of overlap between these. DSC is mostly used in segmentation tasks to evaluate how much the predicted mask (X) matches that of the ground truth (Y).

$$\mathrm{DSC}(X,Y) = \frac{2||X \cap Y||}{||X|| + ||Y||} \qquad (2.22)$$

were $|| \cdot ||$ represents the size (number of elements) of the mask. The coefficient ranges from zero (no intersection) to one ($X = Y$). Is relevant to note that small changes on the boundaries of the sets would yield minor changes in DSC. Hence why a combination of DSC and HD is desired, as the latter can recover the goodness of overlap between the boundaries of the two sets.

2.5   Introduction to deep learning

Mimicking the human brain, made of neurons, the building block of AI models are artificial neurons or preceptrons [49]. These are non-linear operators which react to input signals in a similar fashion as human neurons. In fact, artificial neurons have the capacity to adapt their threshold, by updating their input weight ($\{\omega_0, \omega_1, ..., \omega_N\}$) and bias ($b$) based on the backpropagated error signal generated during the training phase [50, 51]. Finally, the activation function ($f(\cdot)$) yields a numeric value based on the weighted sum of the inputs and the bias (see Figure 2.4). There exists a myriad of activation function, each with its pros and cons, being as simple as a linear function, to more complex like the sigmoid, the hyperbolic tangent, or the rectified linear unit (ReLU).
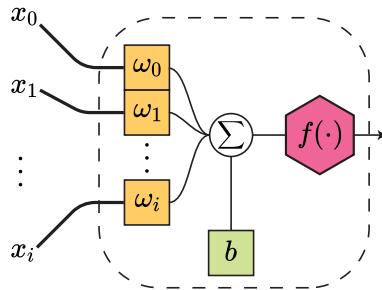
Figure 2.4: Inner structure of an artificial neuron. The trainable parameters are shown in boxes, whereas the activation function is depicted inside the hexagon.

These neurons on their own represent a linear regression, which is further processed by the activation function. Combining multiple of these operators in layers, also known as multilayer preceptron (MLP) [49], the decision boundary becomes more sophisticated. Furthermore, by pilling multiple of these layers, the model becomes sensitive to more complex features extracted from those detected in the previous layer [51]. Because of how each perceptron is feeding from all the preceptrons on the previous layer, MLPs models are not suitable to process higher dimensional data such as images. Hence the need to pre-process the data (and eventually post-process the output), onto a format suitable to be analysed by the MLP.

Convolutional neural networks (CNN) on the other hand, were conceived to operate over high dimensional data. Unlike MLPs, convolutional layers use a filter

or *kernel* of trainable weights, which is applied to the input data in a sequential fashion, as in the convolution operator (cross correlation [49]) as the one shown in Equation 2.23 [51, 49]. Here the discrete convolution for two bidimensional functions e.g., and image and a kernel, where the overlap between the operators I and K is evaluated across the $\mathbb{R}^2$ space, by shifting (and mirroring) g with respect to $f(a,b)$.

$$[I * K](i,j) = \sum_a \sum_b I(a)K(i-a, j-a) \tag{2.23}$$

The result is then fed to the activation function, resulting in a lower resolution version of the input. Convolutional layers usually have several kernels, reacting to different features e.g., vertical edges, round patters, textures, etc. Furthermore, due to the limited receptive field (fixed by the kernel size) deeper convolutional layer will learn from the filtered data produced by the previous layer, gaining insight on more complex features and increasing its receptive field. Figure 2.5 shows an example of a CNN with two convolutional layers. The first layer features a kernel of size $k \times k$, depicted in the figure. In this case, the bias has been omitted, but as with perceptrons a bias is added prior the activation function. The classification task is carried out by the fully connected layer, made using a MLP. Notice how the second convolutional layer learns from the filtered images made by the first layer.
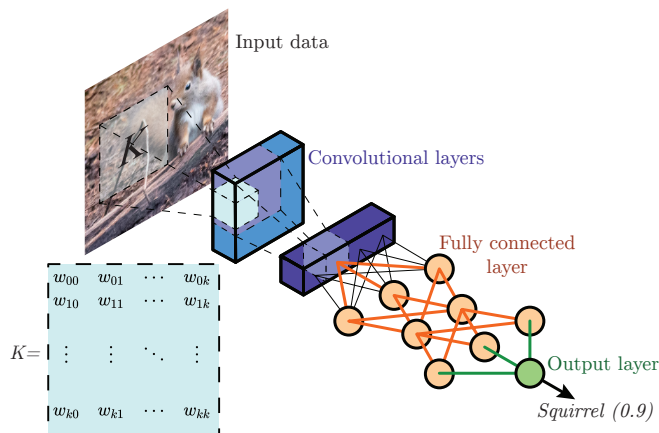
Figure 2.5: Example of a convolutional neural network (CNN), featuring two convolutional layers (blue) and a fully connected layer (orange).

Deep learning models are the result of stacking a large number of convolutional layers, resulting in models with a massive amount of parameters. For instance, the object recognition model ResNet50 features 48 convolutional layers, plus a max pooling and a average pooling layers, adding over 23 million trainable parameters. This achievement was motivated by the improvements in hardware, with the development of more powerful graphical processing units (GPU), as well as techniques like back-propagation and mini-batch training [50] which enabled quick feeding of data to the model and short training stages.

Commonly, there are three strategies for training ML models: supervised, unsupervised, and reinforcement learning [51]. Supervised training assumes the existence of a ground truth against which the model is tested during training, resulting in an error value used to stir the values of the trainable weights of the algorithm. On the other hand, unsupervised training strategies are used when no ground truth is available. In this case, different metrics are used to assess the goodness of the predictions, for example, when clustering the data samples based on a priori information of the data. Lastly, reinforcement learning is perhaps the closest technique to how we humans learn. The model is given a goal or set of targets, and a collection of rules or available actions to choose from, and at every step, the model is given a positive or negative feedback based on the given goals.

Lately, a variation over the unsupervised approach was presented for image registration: weakly-supervision [52]. This is specifically tailored for image registration in that the ground truth is also available as the training data, i.e., is used as both training and ground truth as depicted in Figure 2.6. This methods are further explored later on in Section 5.1.
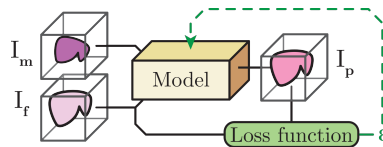
Figure 2.6: In weakly-supervision, the predicted image (I_p), result of transforming the moving image (I_m), is evaluated against the fixed image (I_f). The resulting error (ε) is backpropagated to train the artificial intelligence (AI) model.

2.6 Multitask learning

This field dwells on the optimisation of models on various different task. This usually involves the use of multiple loss functions L_i to train a model [53, 54, 55]. The simplest approach for multi-objective loss functions is the use of weights ω_i, to balance the contribution of each function (see Equation 2.24).

$$\mathscr{L}\left(\mathbf{y}_t,\mathbf{y}_p\right) = \sum_{i=1}^{N} \omega_i L_i\left(\mathbf{y}_t,\mathbf{y}_p\right) \tag{2.24}$$

Optimising the values of the weights is a delicate process, which can dramatically affect the convergence of the model and may require an extended period to finds the optimal values.

In A. Kendall, *et al.,* [53] propose a maximum likelihood based multiple-loss function formulation. The output of the network is defined as a probabilistic model, relating the predicted values with the weight of the loss functions. Through *maximum likelihood*, it is possible to optimise this probabilistic model with respect to the collection of weights ($W = \{\omega_i\}, i \in \{0, 1, \ldots, N\}$). Hence, finding the optimal values for Equation 2.24 [53]. For instance, in classification problems, the softmax function is applied to the output to produce a probability value. Following A. Kendall, *et al.,* approach, the multiple-loss function for a classification model can be defined as shown in Equation 2.25.

$$\mathscr{L}\left(W,\sigma_i;i \in \{1,\ldots,N\}\right) \cong \sum_{i=1}^{N} \frac{1}{2\sigma_i} L_i\left(W\right) + \log\sigma_i \tag{2.25}$$

By setting the losses weights as trainable parameters, the network is given the flexibility to balance the contribution of each loss function dynamically during training. This strategy is further explored in Article IV and Section 5.1.

CHAPTER 3

TRACKING
TECHNOLOGIES

As briefly introduced in Section 2.1, before any navigation can be done, the
IGS requires the localisation information of the different elements in the operating
theatre. In order to do so, the navigation platform makes use of different tracking
systems which can compute the location and attitude i.e., orientation, of the tracked
elements, and feed this information to the navigation platform. This information
is later used to move a virtual representation of the surgical equipment around the
virtual model of the patient, hence providing navigation information.

For surgical navigation, the most relevant and extended technologies are the
optical tracking systems (OTS) and electromagnetic tracking systems (EMTS).
Both technologies will be further described in the following sections.

3.1   Optical tracking systems

As the name implies, OTSs rely on optics to track the objects. The hardware typic-
ally comprises two or more cameras, whose relative location is known and fixed; a
set of reflective markers, rigidly attached to the tool in a predefined configuration;
and a computer, to process the information and return the spatial location of the
tool. Applying epipolar geometry, the system can triangulate the location of the
markers located within the tracking volume and recorded by the cameras.

The markers can be of two types: passive or active. The former requires an
external illumination source for the cameras to detect the reflected radiation. This
source of light is usually mounted on the same frame as the cameras to guarantee a
proper illumination of the scene. Whereas the active markers are equipped with a

Figure 3.1: Optical tracking system (OTS). From left to right and top to bottom: passive markers, active markers, frame or marker shield with passive markers and infrared cameras.

source of light, needing no external illumination. In order to improve the detection of the markers, OTSs usually operate in the range of infrared light. At this frequency, the images recorded are mostly black, due to the low ambient IR radiation, with bright marks corresponding to the optical markers.

Once the markers are located, the system is able to recover the orientation of the object being tracked. To do so, the markers not only are rigidly attacked to the tool, but they are mounted on a frame, or marker shield, in a non-arbitrary and non-symmetrical distribution. The spatial disposition of the markers is known to the system, and thus, the orientation can be estimated by matching the recorded configuration to the known model.

Lastly, because of the geometry of laparoscopic tools and their use, it is impractical to place the large marker shield directly on the tip of the tool, which is what the surgeon needs to track. Instead, the marker shield is normally placed on the handler of the tool, in such a way that is visible to the cameras of the OTS and does not interfere with the movements of the surgeon. Therefore, for the OTS to be able to return the location of the tip of the tool, the geometrical information of the tool must be provided in advanced. This is done through calibration. There are several calibration procedures, being the most commonly used the *pivot calibration*. This procedure, consist of fixing the tool tip and moving the instrument around it. The spatial location of the markers is recorded and, through a least square estimation, the location of the tip is computed.

Examples of commercially available solutions of OTSs are: NDI POLARIS Spectra® and NDI POLARIS Victra® (NDI, Toronto, Canada) (see Figure 3.2) or Kick® (Brainlab, Munich, Germany). An example on non-clinical applications

are the SMARTTRACK3® and TRACKPACK/E® (Advanced Realtime Tracking, Weilheim in Oberbayern, Germany) devices.

(a)   (b)

Figure 3.2: (a) NDI POLARIS Spectra® (NDI, Toronto, Canada) and (b) a marker shield with a gripper for ultrasound (US) probes and passive reflective markers.

3.2   Electromagnetic Tracking Systems

While OTSs provide accurate information in a large area i.e., frustum, their major flaw is the need of a direct line of sight of the markers. Also, the inability to track flexible tools unless a marker shield is directly attached to the point of interest makes it impractical for flexible surgical tools. On the other hand, EMTSs are based on electromagnetic radiation to locate the sensor, or probe, within the electromagnetic field generated by the field generator. With a relative magnetic permeability close to that of the free space, the human skin is invisible to the low intensity generated magnetic field, and does not disrupt it. Hence, EMTSs can track the sensor, and thus, the tools to which this is attached, without the need of a direct line of sight.

EMTSs usually consist of three elements: the field generator, which is placed close to the area where to do the tracking and generates a variable magnetic field; the probe or sensor, rigidly attached to the point to track; and the control unit, which interfaces with the computer, operates the field generator, and reads and digitises the sensor readings. The generated magnetic field changes along time, inducing currents on the coils built within the sensor, following Faraday's Induction Law. These currents are measured by the control unit and used to compute the location and orientation of the sensor within the field.
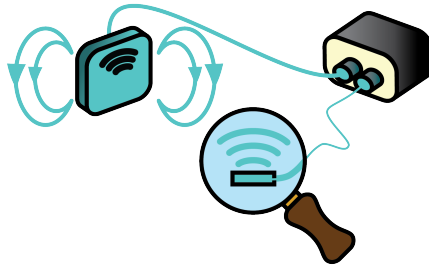
Figure 3.3: Electromagnetic tracking system (EMTS). From left to right: the field generator, the electromagnetic probe (magnified) and the control unit.

Because the probe can be placed close to the point of the tool to be tracked, there is no need for a calibration step. Also, because of the small size of the sensor (see Figure 3.4b it is suitable to be placed close to the tip of flexible instruments.

Even though EMTSs may show robustness to stainless steel and titanium made surgical instrument, ferromagnetic objects e.g., the surgical table, can distort the generated magnetic field, and thus, decrease the tracking accuracy. Therefore, the proximity to this type of elements must be taken into consideration when placing the field generator. In addition to this, the EMTSs have a small tracking volume, forcing the user to place the field generator as close as possible to the patient. To do so, the field generator is rigidly fixed using an articulated arm attached to the table, allowing various orientations and locations without interfering with the movements of the surgeon. There are also larger field generators that can be placed underneath the patient and are designed to minimise the effect of the bed on the magnetic field e.g., NDI Aurora® Tabletop field generator[1].

Examples of commercial products on EMTS are NDI Aurora® (NDI, Toronto, Canada) (see Figure 3.4a) or Kick EM® (Brainlab, Munich, Germany). In 2017, the Cork University released a research EMTS named Anser EMT, being the first open-source device of its kind [56] (see Figure 3.4c).

---

[1]`https://www.ndigital.com/products/aurora/aurora-field-generators/`
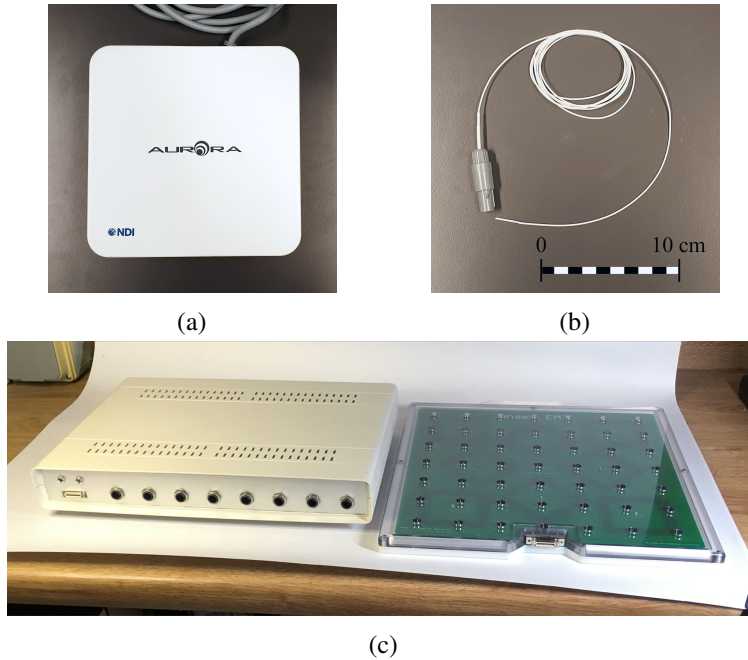
Figure 3.4: (a) NDI Aurora® V2 (NDI, Toronto, Canada) field generator, (b) a electromagnetic sensor and (c) Anser control unit (left) and field generator (right), courtesy of Hernan Alex Jaeger (University College Cork, Cork, Ireland)

3.3 Tracking in surgical navigation

Article I [17] studies the accuracy of these two technologies for surgical navigation, as well as evaluating its applicability. In order to emulate similar conditions to those the tracking systems will be used in, the experiments were carried out in a laparoscopic suite at St. Olavs Hospital (Trondheim, Norway), on a MAQUET® surgical table (Mediatek, Winnipeg, Canada). A custom-made phantom, which allows to test the accuracy of optical and electromagnetic tracking systems, was place in a configuration and location similar to that of the liver. The OTS was placed at the head end of the table, while the EMS was located close to the phantom. The OTS tested was a NDI Polaris Spectra®, and the EMS was a Planar Field NDI Aurora® V2 (NDI, Toronto, Canada). An OTS and a EMTS reference frames were rigidly attached to the phatom and calibrated. A custom-made wooden tool was manufactured to serve as surgical instrument, able to hold both OTS and EMTS

markers. To ascertain the repeatability of the measurements, the markers were fixed using Lego® pieces. Both the tool and the phantom were accurately measured using a CNC machine before conducting the experiments.

A picture of the experimental setup can be seen in Figure 3.5. The EMTS sensor used for the experiments was a NDI 6DOF flex Tube 1.3 mm® Part Number 610060 (NDI, Toronto, Canada). Whereas the optical marker frame was custom build and equipped with NDI passive infrared markers (NDI, Toronto, Canada).

(a)   (b)
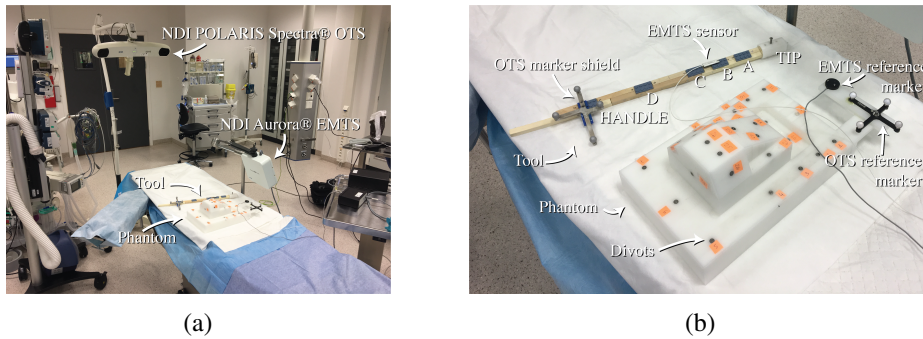
Figure 3.5: (a) Experimental setup showing the optical tracking system (OTS), the electromagnetic tracking system (EMTS), the phantom, and the tool. (b) Detail of the phantom and the tool. The locations on the tool, ordered by increasing distance from the tip of the tool, are named as TIP, A, B, C, D and HANDLE.

Two experiments were conducted to evaluate the TRE (see Section 2.2), consisting on the recording of the location of the metal divots using the custom made tool.

(i) Distance to tip Both the OTS and EMTS markers were placed at different locations along the tool (A to D), increasing the distance to the tool tip.

(ii) Real case scenario The OTS markers was located the farthest from the tool tip (HANDLE), whereas the EMTS sensor was placed close to the tool tip (TIP).

Figure 3.6 shows the average measurements with the standard deviation error for each tracking technology and sensor configuration. And a summary of the measurements can be seen in Table 3.1. A complete table with the measurements can be found at [17].

The factorial ANOVA test yielded significant influence of the tracking technology ($F_{(1, 204)} = 201.596$, $p < 0.01$), existing a significant difference between the
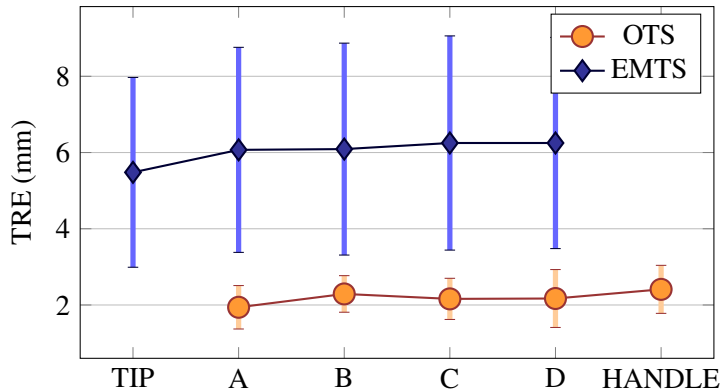
Figure 3.6: Average target registration error (TRE) for both the optical tracking system (OTS) (orange) and the electromagnetic tracking system (EMTS) (blue), and sensor position along the tool (ordered by increasing distance from the tip of the tool: TIP, A, B, C, D, HANDLE).

Table 3.1: Summary of the target registration error (TRE) (mean $\pm$ standard deviation) measurements for the optical tracking system (OTS) and the electromagnetic tracking system (EMTS).

Tracking technology	EMTS	OTS
TRE	6.03 ± 2.71	2.19 ± 0.59

OTS and the EMTS. However, no statistically significant influence was shown for the sensor position along the tool ($F(3, 204) = 0.113$, $p = 0.952$).

For the second experiment, a T-test was performed instead and statistically significant difference was found between the measurements obtained from the OTS and the EMTS ($t(29.439) = -6.081$, $p < 0.01$). Hence there is a significant difference in TRE for each technology in the second scenario. As it was previously shown in the first experiment and shown in Figure 3.6.

From Figure 3.6 and Table 3.1 it is clear to see how the OTS outperforms the EMTS in terms of TRE. However, the use of visual information to triangulate the location of the tools requires a direct line of sight, which means, the OTS sensor must be placed so the surgeon nor the equipment disrupts its visual field. Also, due to the location of the optical markers, OTS is useless when tracking flexible tools

such as LUS. Is in this niche where the EMTS, even though the expected TRE is worse than that of the OTS, becomes the only feasible tracking technology.

Therefore, IGS platforms should primarily rely on OTS whenever possible, but should also integrate electromagnetic-based tracking for minimally invasive surgical navigation.

CHAPTER 4

IMAGE-TO-PATIENT REGISTRATION

Image-to-patient registration, also known as image-to-model registration, comprises the registration techniques that are used to relate physical and virtual coordinates or corresponding objects, e.g., image data and anatomical structures. This is, taking the patient or the tracking system as the reference frame, place the virtual model so that locations on the patient correspond to equivalent locations on the virtual model.

4.1 Image-to-patient registration for surgical navigation

In order to complete the process, the surgeon must locate a set of points or landmarks in the patient that correspond to pre-defined locations in the virtual model. This step is usually performed with a tracked and calibrated tool e.g., a surgical pointer. Coined in the robotics field, point-based registration (PBR) or fiducial landmark-based registration method (FLBR) is the most extended algorithm for image to patient registration. This method is based on the correspondence of two cloud points, one from the virtual model and the second sampled from the patient [57, 30, 58, 10].

Higher-dimensional features are also used in image-to-patient registration, for example, surfaces. In surface matching or surface registration algorithms, the surgeon collects a large set of points from the surface or the organ [30], e.g., by swiping the tracked tool over a section of the surface of the organ. An alternative would be to take advantage of stereoscopic laparoscopes to extract the organ surface and find the match with the virtual model [59].

In general terms the PBR methods, either using points or higher dimensional features, minimise the fiducial localisation (FLE) (see Equation 2.6) i.e., finds the best transformation that minimises the distance of all registered points to their corresponding ones in the virtual model, after registration. Rigid transformations are used to model the registration transformation for which the analytical solution is well-known and has been extensively studied [60, 61]. The rigid PBR problem can be solved with Procrustes problem, where an orthogonal matrix (R) that maps two given matrices (A and B) has to be found. The solution includes the singular value decomposition (SVD) of the matrix BA^T, being R equal to UV^T, where U and V are the orthogonal matrices resulting from the SVD. However, this approach assumes a known one-to-one correspondence between the points in the two cloud points.

Alternatively, the iterative closest point (ICP) algorithm [62, 63] overcomes the lack of one-to-one known correspondence by following a two steps iteration approach. First, the ICP locates the corresponding points in both source and target sets based on distance; then computes the optimal rigid transformation that best aligns these sets. This process is repeated until a target error is reached. However, the original implementation of ICP shows high sensitivity to the initial point matching and the sampling noise. Also the presence of outliers as well as high discrepancy between the geometry of the two sets might difficult the convergence of the method [64]. Multiple alternatives have been proposed to overcome these issues as well as speed up the process, like improving the point matching and transformation regression using other methods like the RANdom SAmple Consensus (RANSAC) [65].

However soft tissues, like organs, are flexible and deformable. Coherent point drift (CPD) [66] is a PBR method which performs a non-rigid registration between the source and target sets. It is based on Gaussian mixture model (GMM). By associating the target points as the *centroids* of Gaussian distributions and the source set as points drawn from these distributions. The algorithm enforces regularisation on the solution as well as coherence in the movement of the source points, to preserve the topology of the moving set, hence the name of the method. In the original implementation, the algorithm follows the likelihood Expectation-Maximisation (E-M) strategy to solve the problem. During the E-step, the membership probability $P(m)$ to the GMM of each target point is estimated. Then, in the M-step, the objective

function is minimised, optimising the GMM in the process, which is then used again in the E-step. The transformation is obtained from the posterior probability of the resulting GMM [66].

4.2 Single landmark registration method

Article II [18] describes the test and validation of a novel image-to-patient registration method, as the prior step for the image-to-image registration. Once the tracking system is setup and calibrated, the next step before the image guided system (IGS) can be used for surgical navigation is the patient-to-image registration. In this phase, the physical location of the patient, and hence, the previously screened anatomical structures, are registered together by the platform. Using algorithms like the aforementioned e.g., PBR or ICP. The main inconvenience of PBR methods is the requirement to register a large enough set of anatomical landmarks on the patient for the algorithm to provide a good solution.

In Article II [18] the single landmark registration method (SLRM) is presented and proposed a quicker but reliable alternative to the aforementioned methods. The SLRM has two phases, an initial registration phase where the virtual model is aligned and registered to the patient, to enable intra-operative navigation. And a second step where the user i.e., the surgeon, can refine the registration transformation close to the lesion. In the first phase, the user is requested to register a single anatomical landmark, as well as the orientation of the vertical axis of the patient. Whereas in the second phase, a single point is enough to refine the registration of the lesion, as the orientation is kept constant between the two steps. An affine transformation is used to register the two model i.e., a rotation and a translation, so the registration will not be as accurate as with other techniques that use deformable models. Again, the aim of SLRM is to provide a quick procedure to enable intra-operative navigation minimising the impact on the surgical pipeline.

The main contribution of this study was testing the feasibility of the method and the demonstration of it. A comparison in terms of target registration error (TRE) and required user time was done with the FLBR, as gold standard method. The accuracy test was carried out using the Ultrasound Phantom IOUSFAN (Kyoto Kagaku Co., Ltd., Japan) [67]. The tracked tool used for the initial registration phase was a surgical pointer with optical markers. A SonixMDP™ ultrasound (US)

scanner (Ultrasonix Medical Corp., Richmond, Canada) was used to capture the US images from a L14-5/38 linear transducer (Prosonic Gyeongbuk, South Korea). A NDI Polaris® optical tracking system was used to locate the tools and the phantom within the experimental setup. The SLRM algorithm run on the open-source image-guided therapy platform CustusX [68]. The communication between the US scanner and CustusX was enabled through OpenIGTLink. The experimental setup can be seen in Figures 4.1a to 4.1c.

(a)   (b)   (c)
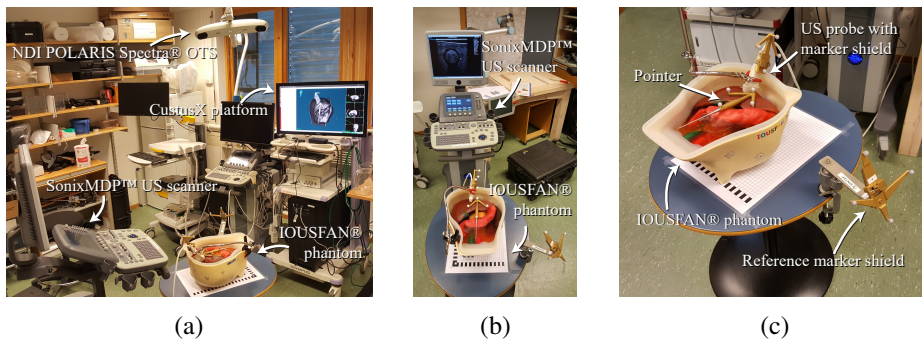
Figure 4.1: Experimental setup with the ultrasound (US) scanner and probe, the optical tracking system (OTS), the electromagnetic tracking system (EMTS), the CustusX platform, and the phantom, used to evaluate the single landmark registration method (SLRM).

While the surgical pointer was calibrated using the pivoting calibration method available in NDI Toolbox® (NDI, Toronto, Canada), the US probe was calibrated using the wire phantom described in [69].

A total of 60 TRE samples were computed for each registration method, being the TRE the Euclidean distance between the centre of the tumour (found on the US image) and the location of this point in the virtual model. For each method, the model was moved 10, 50, and 100 mm in the frontal and longitudinal axes, taking ten samples on each location. In addition, the time required to perform the registration was measured and compared.

Table 4.1 shows a summary of the TRE obtained with the two registration methods for the different displacements of the phantom. Whereas Table 4.2 shows a summary of the time required to complete the SLRM and the FLBR methods during this experiment.

Table 4.1: Target registration error (TRE) (mean ± standard deviation) for the single landmark registration method (SLRM) and fiducial landmark-based registration (FLBR) methods

Registration method	FLBR		SLRM	
Axis	Frontal	Longitudinal	Frontal	Longitudinal
Displacement				
10 mm	4.6 ± 0.2	4.7 ± 0.4	11.3 ± 0.4	11.1 ± 0.7
50 mm	4.6 ± 0.3	5.1 ± 0.4	11.3 ± 0.5	11.1 ± 0.4
100 mm	4.4 ± 0.3	5.2 ± 0.3	10.7 ± 0.4	11.3 ± 0.5

Table 4.2: User time in seconds (mean ± standard deviation) for the single landmark registration method (SLRM) and fiducial landmark-based registration (FLBR) methods

Registration method	FLBR	SLRM
Time	19.63 ± 1.68	7.62 ± 0.63

The TRE measurements were analysed using a one-way ANOVA, clustering the measurements in six groups based on the direction and magnitude of the displacement e.g., X10, X50, X100 for the displacements along the longitudinal axis, and Y10, Y50 and Y100 along the frontal axis. Levene's test yield no difference between the variance of the groups ($p = 0.065$), assumption required to perform a one-way ANOVA test, but the ANOVA test showed statistical difference between the groups ($p = 0.046$). A multiple composition analysis using Tukey's hones significant difference and Scheffe's method yield no statistical difference between pairs of groups, with a significance level of 95%. Hence, the SLRM shows robustness against displacements of the patient.

The TRE results for each registration method were compared using a one-way ANOVA, grouping the samples according to the registration method i.e., SLRM and FLBR. Levene's variance test showed statistical similarity between the variances of the groups ($p = 0.039$). Therefore, instead of the one-way ANOVA, a Welch test of equal means was used instead. The groups were shown to be statistically dissimilar ($p = 0.000$), being the FLBR the method with the best accuracy in terms of TRE.

For the time analysis, the samples were analysed using a T-test analysis. Variances of the groups was found statistically different ($p = 0.000$) and the T-test showed difference between the two groups ($p = 0.000$). Hence, the execution time required for the SLRM is statistically less than that of the FLBR.

CHAPTER 5

IMAGE-TO-IMAGE REGISTRATION

Image-to-image registration refers to the process of aligning sets of images so that corresponding points have the same coordinates. Typically, the images used as a reference is named *fixed* or *target* image (I_f), whereas the set meant to be transformed, so as to resemble I_f, is named *moving* image (I_m). The alignment of the two image sets is defined by a transformation which is in fact the outcome of the registration algorithm.

In a formal way, the image-to-image registration problem can be stated as: given I_f and $I_m \subset \mathbb{R}^n$, find the transformation $\mathrm{T}: \mathbb{R}^n \to \mathbb{R}^n$ so that $I_f(\mathbf{x}) = I_m(\mathrm{T}(\mathbf{x}))$. This same problem can be reformulated as an optimisation problem (see Equation 5.1), as it is usually found in the literature: given I_f and $I_m \subset \mathbb{R}^n$, the transformation model $\mathrm{T}(\boldsymbol{\theta}): \mathbb{R}^n \to \mathbb{R}^n$, where $\boldsymbol{\theta}$ is the set of parameters of the transformation model, and the similarity (dissimilarity) function $\mathrm{S}: \mathbb{R}^n \to \mathbb{R}$; find the set of parameters that maximise (minimise) the similarity (dissimilarity) between $I_f(\mathbf{x})$ and $I_p = I_m(\mathrm{T}(\boldsymbol{\theta}, \mathbf{x}))$, the predicted target image [30, 70].

$$\arg\max_{\boldsymbol{\theta}}\{\mathrm{S}(I_f(\mathbf{x}), I_m(\mathrm{T}(\boldsymbol{\theta}, \mathbf{x})))\} \tag{5.1}$$

The complexity of the aforementioned transformation can vary from that of a rigid affine transformation, with six degrees of freedom (DOF) i.e., three rotations and a translation vector, to more elaborated models like displacement maps. The selection of the model greatly depends on the application, being more usual the affine transformation for registering rigid or semi-rigid structures, like bones. Whereas soft tissues, like lungs or liver, require models with a larger number of

DOF to represent complex deformations i.e., the local movement of consecutive segments of the tissue. However, the larger the number of DOF, the more complex the optimisation strategy, and thus, the harder it will be to find an optimal solution.

It is standard practice to add a regularisation term $(R(\theta))$ that comprises *a priori* information of the problem, or enforces desirable properties onto the transformation, can be added to Equation 5.1, so the problem can be reformulated as follows [30, 70, 71]:

$$\arg\max_{\theta}\{S(I_f(\mathbf{x}), I_m(T(\theta, \mathbf{x}))) + \lambda R(\theta)\} \qquad (5.2)$$

There exists several classifications in the literature attending to the nature of the used images: *monomodal*, when the images are of the same nature, and *multimodal*, when these image were acquired using different technologies. Or based on the transformation model: *rigid*, using affine transformations, and *deformable*, using displacement maps or other non-linear operators. More recently, together with the development of artificial intelligence (AI), it is possible to distinguish between non-deep learning and deep learning registration methods (referred to as deep registration in this text). This distinction is based on the application of convolutional neural networks (CNN) to calculate the registration transformation.

Deep registration will be further introduced in the following section.

5.1   Deep registration

Since beginning of 2010 together with the emergence of advanced graphical processing units (GPU), the research on CNNs architectures for image-to-image registration has gained substantial momentum. In Wu *et al.* [72], a simple two-convolutional-layered network for the alignment of brain magnetic resonance (MR) scans was proposed. It fancied a principal component analysis for dimensionality reduction. Later, the work done by Jaderberg *et al.* [73] became a turning point in the development of deep image registration. The proposed spatial transformer network (STN) allowed for the backpropagation of gradients through the interpolation operations. Combined with auto-encoder architectures like U-Net [74], known for its good performance when trained with small datasets, it was possible to develop new training strategies like weakly supervision [52, 75], and deep image registra-

tion frameworks like VoxelMorph [76]. Further developments include Mok and Chung [77], who proposed a Laplacian pyramid network for multi-resolution-based MR registration, enforcing diffeomorphic non-rigid transformations between the images.

In deep learning, a considerable challenge is the lack of datasets with relevant ground truth. That is, a large collection of images showing the anatomy of the same patient at different times, showing real displacements and artifacts like resections, bleeding, or different poses. This circumstance can be worsened in the case of multi-modal registration, where corresponding images are of different nature. In order to compare images of different modalities and guide the training phase, Hu *et al.* [52] proposed a weakly-supervised strategy [52, 75]. Instead of comparing the images, this approach uses the segmentation masks on both image sets, which can be compared using dice score coefficient (DSC) or Hausdorff distance (HD). Unsupervised learning has also been explored for abdominal and lung computed tomography (CT) in Fu *et al.* [78] and Lei *et al.* [79]. As well as reinforcement learning for multi-modal registration [80].

Article IV further researched on training strategies for cases with limited data, applied to registration of abdominal CT. Focusing on mono-modal registration, and employing well-known deep learning concepts like transfer learning or multi-task learning. Furthermore, transfer learning across image modalities and anatomies was explored to overcome the limited data available for abdominal CT registration. On a technical perspective, this study brought an augmentation layer for on-the-fly data augmentation and ground truth generation was introduced, as well as a uncertainty weighting loss layer for adaptive multi-task learning. Both were packaged in the openly available framework Deep Deformation Map Registration (DDMR)[1].

Two dataset were selected for conducting the experiments: the Laparoscopic Versus Open Resection for Colorectal Liver Metastases: The Oslo-CoMet randomised controlled trial dataset (Oslo-CoMet) [81, 1], and the information extraction from images dataset (IXI)[2].

Four different experiments were conceived for this study:

---

[1] `https://github.com/jpdefrutos/DDMR`
[2] `https://brain-development.org/ixi-dataset/`

(i) Ablation study. To identify the key components in deep image registration, different training strategies and loss function combinations were evaluated. Three training strategies were considered, all using weakly-supervised learning: 1) the baseline (BL) using only intensity information, 2) adding segmentation guidance (SG) to the baseline, and 3) adding uncertainty weighting (UW) to the SG approach. All experiments share the same input size and CNN backbone, described in Figure 5.1, and were evaluated on both the IXI and Oslo-CoMet datasets. Six loss combinations were tested, including intensity and segmentation-based loss functions, as shown in Table 5.1.

Table 5.1: Configurations trained on both the IXI and the Oslo-CoMet datasets.

Design	Model	Loss function
BL-N	Baseline	NCC
BL-NS	Baseline	NCC, SSIM
SG-ND	Segmentation-guided	NCC, DSC
SG-NSD	Segmentation-guided	NCC, SSIM, DSC
UW-NSD	Uncertainty weighting	NCC, SSIM, DSC
UW-NSDH	Uncertainty weighting	NCC, SSIM, DSC, HD

BL: baseline, SG: segmentation guidance, UW: uncertainty weighting, N: normalised cross correlation, S: structural similarity index metric, D: Dice similarity coefficient, H: Hausdorff distance.

(ii) Transfer learning. The entire model was finetuned directly or in two steps, i.e., by first finetuning the decoder, keeping the encoder frozen, and then finetuning the full model. A learning rate of 10^{-4} was used when performing transfer learning. The aim was to assess the benefit of finetuning for deep image registration to applications with little data samples available, e.g., abdominal CT registration.

(iii) Baseline comparison. Advanced Normalisation Tools (ANTs) was chosen as a method against which to compare the deep image registration models. Because this same method was used to generate the segmentation labels on the IXI dataset, this experiment was only conducted on the Oslo-CoMet dataset. Two different configurations were tested: symmetric normalisation (SyN), with mutual information as optimisation metric, and SyN with NCC as metric (SyNCC).

(iv) Training runtime.   Finally, to assess the impact of the augmentation layer on the training phase, the GPU resources were monitored during a full training epoch. Only the second epoch was considered, as the first one served as warm-up.

Because the TRE is the gold standard in surgery, this metric was chosen to run the statistical tests, as described in [20].

The evaluation was carried out on the test sets of the IXI and Oslo-CoMet datasets, for which the fixed-moving image pairs were generated in advance, fixing the evaluation images for all the models. The predicted displacement maps were upsampled to isotropic resolution, at which the four sets of metrics were evaluated. Image similarity metrics: normalised cross-correlation (NCC) and structural similarity index metric (SSIM). Segmentation metrics: DSC, HD, and 95th percentile HD (HD95). For image registration, target registration error (TRE) was estimated using the centroids of the segmentation masks of the fixed image and the predicted image. Lastly, the methods were compared in terms of inference runtime, only measuring the prediction and application of the displacement map. All other operations were the same between the methods.

As depicted in Figure 5.1 the pipeline consists of the augmentation layer, the deep learning model (consisting of an VoxelMorph model [76]), the spatial transformer block, and the loss computation step.
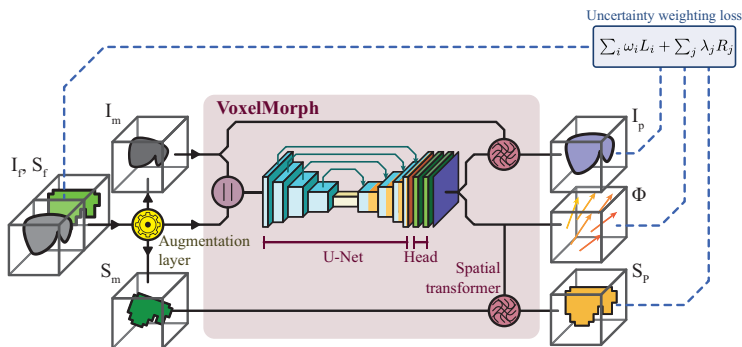
Figure 5.1: Proposed pipeline generating artificial moving images on-the-fly, predicting the displacement map using a modified U-Net and finding the optimal loss weighting automatically using uncertainty weighting (UW) loss.

Training was carried out using Adam optimiser and gradient accumulation. The latter allowed to overcome memory constraints and enable larger batch training. With a batch of one and eight mini-batch gradients, the effective batch size was artificially increased to eight. Learning rate was set to 10^{-3}, with a scheduler to decrease by 10 whenever the validation loss plateaued, with a patience of 10. The training was limited to 10^5 epochs, and manually stopped if convergence did not improve. The model with the lowest validation loss was saved.

Table 5.2 and Figure 5.5 show quantitative and qualitative results of the model evaluated on the IXI test dataset. Whereas Tables 5.3 to 5.5 and Figures 5.4 and 5.5 show the results for the Oslo-CoMet test dataset. And Table 5.6 and Figure 5.7 show the results for the ANTs registration method. Figure 5.8 depicts the results from the Training runtime experiment, with the memory consumption and runtime of models trained with and without the augmentation layer. The best performing methods in terms of individual performance metrics, i.e., most optimal mean and lowest standard deviation, were highlighted in bold.

A sample of the IXI and Oslo CoMet test sets is shown in Figure 5.2.

| Fixed image | Moving image | Fixed segms. | Moving segms. |

(a)

| Fixed image | Moving image | Fixed segms. | Moving segms. |

(b)
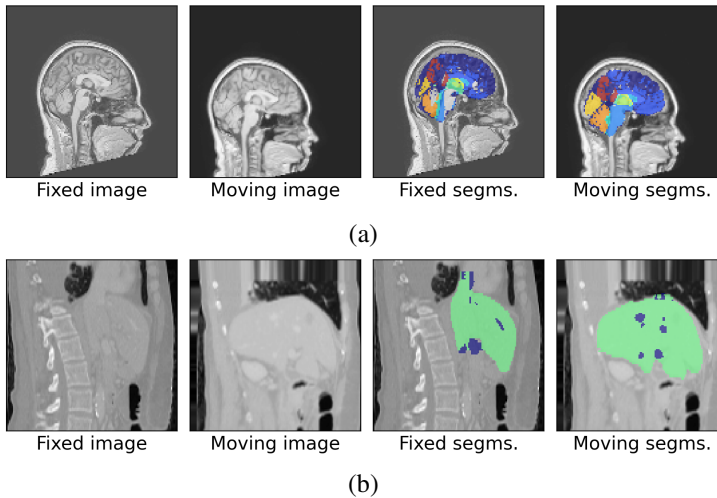
Figure 5.2: Sample test image from the (a) IXI and (b) Oslo-CoMet test sets.

The combination of NCC and SSIM on the IXI dataset, resulted in a better performance in terms of intensity-based metrics for the baseline models, the opposite behaviour was observed on the segmentation metrics (see Table 5.2). SG improved

the performance across all metrics, compared to the baseline. Minor improvement was observed using uncertainty weighting, and Hausdorff loss was not beneficial.

Table 5.2: Evaluation of the models (mean $\pm$ standard deviation) trained on the IXI dataset. The last row shows the metrics evaluated on the original image pairs prior registration (unregistered).

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime (s)
BL-N	0.23±0.16	0.52±0.12	0.03±0.01	109.71±26.19	100.26±27.91	29.47±8.46	0.83±0.77
BL-NS	0.25±0.16	**0.53±0.12**	0.02±0.01	145.36±22.41	138.19±23.48	30.06±9.07	0.73±0.56
SG-ND	0.45±0.24	0.46±0.10	0.61±0.08	4.64±1.37	2.15±0.54	1.08±0.39	0.82±0.58
SG-NSD	0.46±0.24	0.46±0.11	0.61±0.07	4.54±1.42	2.10±0.49	1.07±0.37	0.74±0.64
UW-NSD	**0.47±0.24**	0.46±0.11	**0.63±0.08**	**4.44±1.40**	**2.03±0.51**	**0.97±0.36**	**0.72±0.59**
UW-NSDH	**0.47±0.24**	0.46±0.11	0.61±0.07	4.63±1.49	2.14±0.52	1.06±0.36	0.75±0.59
Unregistered	0.45±0.21	0.24±0.07	0.07±0.06	21.77±5.15	18.73±4.88	11.53±3.01	-

SSIM: structural similarity index metric, NCC: normalised cross-correlation, DSC: dice score coefficient, HD: Hausdorff distance, HD95: 95th percentile of the HD, TRE: target registration error.
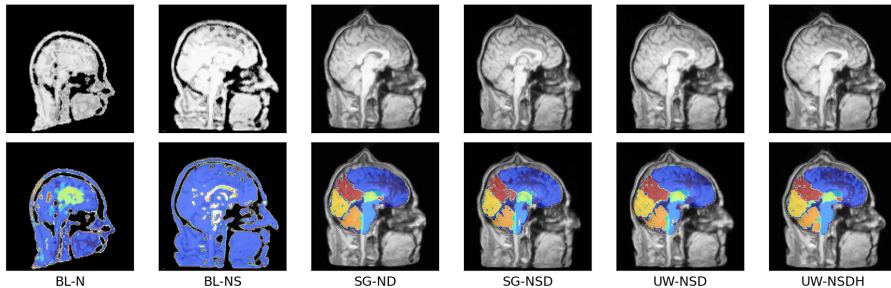
BL: baseline, SG: segmentation guidance, UW: uncertainty weighting, N: normalised cross-correlation, S: structural similarity index metric, D: dice score coefficient, H: Hausdorff distance.

Figure 5.3: Predictions of the models trained on the IXI dataset, on the IXI sample image (see Figure 5.2).

A similar trend was observed on the Oslo-CoMet dataset (see Table 5.3). Nonetheless, the baseline yielded better results than that of the IXI dataset, especially in terms of intensity-based metrics. Again, SG boosted the performance, and uncertainty weighting resulted in a minor improvement.

Compared to the models shown in Table 5.3, fine-tuning the entire model trained on the IXI dataset to the Oslo-CoMet yielded similar intensity metrics overall, but the segmentation metrics dramatically improved with the application of uncertainty weighting (see Table 5.4).

Table 5.3: Evaluation of the models (mean ± standard deviation) trained on the Oslo-CoMet dataset. The last row shows the metrics evaluated on the original image pairs prior registration (unregistered).

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime (s)
BL-N	0.52±0.10	**0.20±0.07**	0.23±0.09	54.10±7.22	30.36±3.58	18.11±7.62	0.78±1.50
BL-NS	**0.62±0.13**	0.17±0.07	0.29±0.07	37.69±8.04	22.06±5.17	13.95±4.78	**0.76±1.44**
SG-ND	0.55±0.15	0.16±0.06	0.38±0.14	**22.03±8.27**	**12.74±6.12**	**7.60±3.96**	0.76±1.46
SG-NSD	0.58±0.13	0.12±0.07	**0.35±0.07**	25.22±7.92	14.49±4.22	8.91±3.08	0.77±1.49
UW-NSD	0.54±0.13	0.11±0.06	0.26±0.07	25.08±6.67	18.47±5.34	11.52±3.32	0.77±1.49
UW-NSDH	0.59±0.14	0.14±0.06	0.35±0.11	24.49±8.67	14.57±5.93	8.34±4.31	0.78±1.50
Unregistered	0.60±0.13	0.09±0.05	0.24±0.08	24.60±5.56	19.06±4.89	11.86±2.75	-

SSIM: structural similarity index metric, NCC: normalised cross-correlation, DSC: dice score coefficient, HD: Hausdorff distance, HD95: 95th percentile of the HD, TRE: target registration error.

BL: baseline, SG: segmentation guidance, UW: uncertainty weighting, N: normalised cross-correlation, S: structural similarity index metric, D: dice score coefficient, H: Hausdorff distance.

Figure 5.4: Predictions of the models trained on the Oslo-CoMet dataset, on the Oslo-CoMet sample image (see Figure 5.2).

Further improvement was observed on the models fine-tuned in two steps when using uncertainty weighting, as shown in Table 5.5. The ANTs methods (SyN and SyNCC variants) performed the poorest on the Oslo-CoMet dataset. Both yielded similar results, but the SyNCC was considerably slower. Unlike the deep image registration models, which had similar inference runtimes of less than one second. On average, the CNN-based methods were $\sim 12\times$ and $\sim 324\times$ faster than SyN and SyNCC, respectively.

The deep learning models struggled with image reconstruction, unlike ANTs (see the online resource Fig S5). For instance, anatomical structures outside the segmentation masks were poorly reconstructed in the predicted image, e.g., the spine of the patient.

Table 5.4: Evaluation of models (mean ± standard deviation) trained on the Oslo-CoMet dataset from fine-tuning the entire architecture. The last row shows the metrics evaluated on the original image pairs prior registration (unregistered).

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime (s)
BL-N	0.52±0.08	**0.17±0.07**	0.23±0.07	57.98±5.36	33.00±5.14	24.09±5.92	0.77±1.45
BL-NS	**0.61±0.09**	0.16±0.07	0.14±0.03	82.91±6.96	59.94±6.41	34.41±13.03	0.77±1.46
SG-ND	0.56±0.13	0.14±0.07	0.43±0.09	15.81±5.56	9.05±3.18	5.89±3.10	0.79±1.56
SG-NSD	0.58±0.13	0.14±0.07	0.42±0.10	16.26±6.37	9.50±3.51	5.84±3.01	0.76±1.48
UW-NSD	0.58±0.12	0.14±0.06	**0.48±0.11**	15.53±5.80	**7.84±3.17**	4.05±2.41	**0.76±1.47**
UW-NSDH	0.59±0.12	0.14±0.06	0.47±0.10	**15.29±5.65**	7.91±2.82	**3.95±2.09**	0.78±1.51
Unregistered	0.60±0.13	0.09±0.05	0.24±0.08	24.60±5.56	19.06±4.89	11.86±2.75	-

SSIM: structural similarity index metric, NCC: normalised cross-correlation, DSC: dice score coefficient, HD: Hausdorff distance, HD95: 95th percentile of the HD, TRE: target registration error.

BL: baseline, SG: segmentation guidance, UW: uncertainty weighting, N: normalised cross-correlation, S: structural similarity index metric, D: dice score coefficient, H: Hausdorff distance.

Figure 5.5: Predictions of the models trained on the Oslo-CoMet dataset from fine-tuning the entire architecture, on the Oslo-CoMet sample image (see Figure 5.2).

Table 5.5: Evaluation of the models (mean ± standard deviation) trained on the Oslo-CoMet dataset from fine-tuning in two steps. The last row shows the metrics evaluated on the original image pairs prior registration (unregistered).

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime (s)
BL-N	0.52±0.07	**0.19±0.07**	0.24±0.06	60.92±26.06	39.96±30.25	22.34±8.60	0.79±1.52
BL-NS	**0.62±0.10**	0.17±0.07	0.14±0.04	85.71±6.40	60.93±4.15	32.84±11.90	0.76±1.45
SG-ND	0.56±0.12	0.14±0.07	0.44±0.09	16.12±5.29	8.87±2.94	5.12±2.52	**0.77±1.48**
SG-NSD	0.58±0.12	0.15±0.07	0.43±0.08	16.93±6.50	9.17±3.02	5.21±2.40	0.77±1.49
UW-NSD	0.60±0.11	0.15±0.06	**0.53±0.13**	15.13±5.68	**6.97±2.83**	**3.40±1.91**	**0.77±1.48**
UW-NSDH	0.60±0.12	0.15±0.06	0.50±0.12	**14.79±5.79**	7.37±2.99	3.55±2.14	**0.77±1.48**
Unregistered	0.60±0.13	0.09±0.05	0.24±0.08	24.60±5.56	19.06±4.89	11.86±2.75	-

SSIM: structural similarity index metric, NCC: normalised cross-correlation, DSC: dice score coefficient, HD: Hausdorff distance, HD95: 95th percentile of the HD, TRE: target registration error.
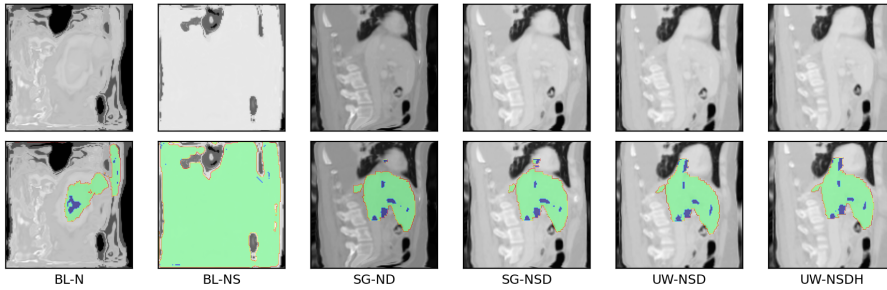
BL: baseline, SG: segmentation guidance, UW: uncertainty weighting, N: normalised cross-correlation, S: structural similarity index metric, D: dice score coefficient, H: Hausdorff distance.

Figure 5.6: Predictions of the models trained on the Oslo-CoMet dataset from fine-tuning in two steps, on the Oslo-CoMet sample image (see Figure 5.2).

Table 5.6: Evaluation of ANTs (mean ± standard deviation) on the Oslo-CoMet dataset. The last row shows the metrics evaluated on the original image pairs prior registration (unregistered).

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime (s)
SyN	0.61±0.13	0.20±0.07	0.49±0.01	17.93±3.44	9.62±1.57	22.34±4.96	10.01±3.69
SyNCC	0.63±0.13	0.20±0.07	0.49±0.01	18.59±2.99	9.64±1.61	22.31±5.04	323.81±87.13
Unregistered	0.60±0.13	0.09±0.05	0.24±0.08	24.60±5.56	19.06±4.89	11.86±2.75	-

SSIM: structural similarity index metric, NCC: normalised cross-correlation, DSC: dice score coefficient, HD: Hausdorff distance, HD95: 95th percentile of the HD, TRE: target registration error.

SyN: symmetric normalisation, SyNCC: SyN with normalised cross-correlation as metric.

Figure 5.7: Predictions of the ANTs (SyN and SyNCC) algorithms, on the Oslo-CoMet sample image (see Figure 5.2).

The use of the augmentation layer resulted in a negligible increase in training runtime of 7.7% per epoch and 0.47% ($\sim$ 74 MB of 16 GB) increase in GPU memory usage (see Figure 5.8).

(a) GPU usage with augmentation layer

(b) GPU usage without augmentation layer

(c) Memory usage with (blue) and without using the augmentation layer (orange)
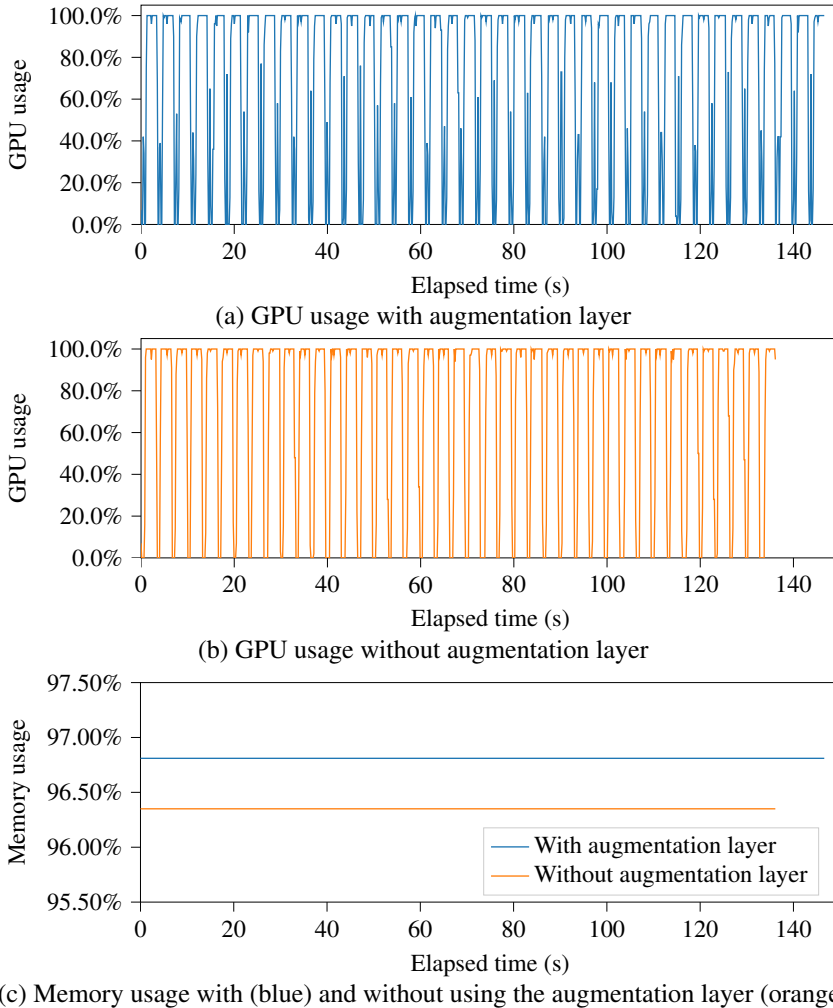
Figure 5.8: Impact of the augmentation layer on the graphical processing unit (GPU) resources (percentage of memory and GPU usage) during one training epoch. The blue line corresponds to the impact on the resources when using the augmentation layer, whereas the orange line corresponds to the same training without using the augmentation layer.

CHAPTER 6   INTRA-OPERATIVE AUGMENTED REALITY

Augmented reality (AR) technology is said to be at the *plateau of productivity* stage of Gartner's hype cycle [82]. Market solutions like HoloLens™and HoloLens 2™developed by Microsoft (Redmond, WA, USA), can provide with the design and hardware configuration to bring augmented reality into the operating room. Though the application of AR can also be found on the enhancement of the video feed of a laparoscope. This section explores the use of augmented reality for intra-operative navigation as a direct result of image-to-patient registration.

6.1   Augmented reality for surgical applications

Augmented reality refers to the introduction of artificially generated images into the natural view of the user, through holographs and projections. Unlike virtual reality, where everything that is shown to the user has been generated by a software, in AR what the user sees is kept constant and software generated graphics are overlaid on top. The aim of AR is indeed to enhance the visual perception of the user by providing additional information about what is viewed.

Commercially available AR solutions are either head-mounted displays (HMD), such as HoloLens™[1] (Microsoft©, USA) or Moverio™[2] (Seiko Epson Corp©, Japan), or heads-up display (HUD), like hand-held devices or external static monitors [83, 84]. Depending on the display technology, there are optical see-trough screens in which the line of sight is not interrupted while projecting the virtual

---

[1] https://www.microsoft.com/hololens
[2] https://tech.moverio.epson.com/

models, and virtual see-through screens, where both the real image and the virtual rendering are shown to the user on a physical screen [85]. Regardless of the approach, in order to correctly render the virtual models, it is required to know the spatial location and orientation of the user (in HMD) or camera (in HUD), i.e., the point of view. HMD and some hand-held HUD e.g., smartphones, are equipped with sensors, like inertial measurement units, and cameras to monitor and track the movement of the head, hence recovering the correct point of view of the user with respect to the virtual environment. Optical and electromagnetic trackers can be used to follow the movement of the user or the camera, using an external sensor such as an optical tracking system (OTS) or electromagnetic tracking system (EMTS) [83].

In surgical applications, AR can be used to show the surgeon anatomical structures that would be hidden otherwise e.g., bones [86], or for surgical planning [87, 88]. Concerning navigation in surgery, the aforementioned AR solutions are integrated in the image-guided system (IGS) bringing together the registered virtual model of the patient to the virtual scenario user by the AR system. This is depicted in Figure 6.1, where the HUD-AR system including the laparoscope camera (H) and the screen (C) are tracked by the IGS through the OTS (O), hence getting access to the registered model (I).
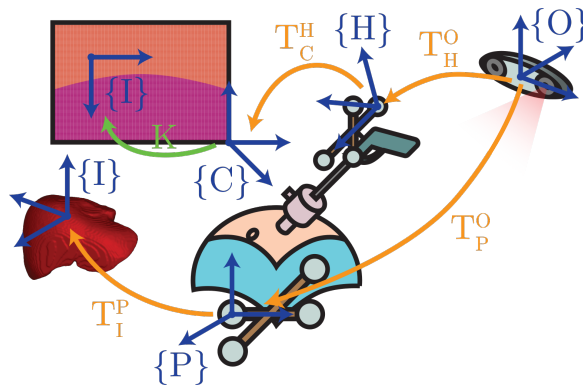
I: image data (2D/3D), P: patient, C: screen, H: laparoscope camera, O: optical tracking system, K: intrinsic matrix, T_b^a: homogeneous transformation from reference frame a to reference frame b.

Figure 6.1: Transformations between reference frames. The hand-eye calibration yields the transformation T_C^H allowing, together with the intrinsic matrix of the camera, the location of the video frames in the patient coordinates space (P).

In deformable environments, as is the case of laparoscopic surgery, the scene needs to be tracked to compensate for movements which could prevent the correct alignment of the projected information [85]. This can be achieved through tracking of visual features in the camera feed as seen, for example, in Simpfendörfer *et al.* [89], where coloured needles were used to track the motion of the scene, or Haouchine *et al.* [90], where the surface of the liver is recovered through stereoscopic vision and used to update the deformation on the virtual model.

AR applications for minimally invasive surgery involves the use of HUD and a laparoscope as the user's point of view. This setup requires the calibration of the camera to correctly project the virtual models, resulting in two sets of parameters: the extrinsic and intrinsic parameters of the optical system. The extrinsic parameters relate to the physical location and orientation of the camera, this is the transformation T_C^H in Figure 6.1 (see Section 2.1). Unless the physical dimensions of the laparoscope and location of the tracked markers is known, N-pose hand-eye calibration [91, 92] enables the computation of this transformation in a quick and easy way. Also known as the $AX = XB$ problem, hand-eye calibration is used in robotics to recover the transformation between the robot end effector and the camera [93]. This method requires the recording of a calibration plate e.g., a calibrated chess-pattern picture, while tracking the camera.

Lastly, the intrinsic parameters are unique to each optical system and relate the 2D pixel space of the image with the 3D spatial coordinates of the scene. These are represented in the intrinsic matrix (K) shown in Equation 6.1. This transformation is the result of the camera calibration [92, 94], which is done using the calibration plate images acquired during the hand-eye calibration

$$K = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{6.1}$$

where

f_x and f_y are the focal distances along the image axes, and scaled with the pixel size,

γ is the skewing factor of the optical sensor,

c_x and c_y are the coordinates of the optical centre in pixels.

When both sets of parameters are combined, it is possible to relate the locations in the registered patient model to the pixels in the laparoscope video, as shown in Equation 6.2.

$$
\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot (\mathrm{T}_C^H)^{-1} \cdot (\mathrm{T}_H^O)^{-1} \cdot \mathrm{T}_P^O \cdot \mathrm{T}_I^P \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}
\tag{6.2}
$$

Notice that homogeneous coordinates are used here, to relate locations in different dimensional spaces i.e., 3D spatial coordinates and 2D pixels.

6.2   Image-to-patient registration on augmented reality

Article III [19] explores the influence of human accuracy in sampling landmark locations, for image-to-patient registration. In particular, for AR in surgery. A HUD system was used to enable AR, using a ENDOEYE FlLEX 3d (Olympus, Tokyo, Japan) laparoscope and an external screen, as seen in Figure 6.2. The spatial location and orientation of both the patient and the camera (extrinsic parameters) were tracked using the NDI POLARIS Spectra® (NDI, Toronto, Canada) OTS. A reference frame with optical markers was attached to the handler of the laparoscope, to track its location and orientation, and to the surgical table. Hand-eye calibration was done following Lee *et al.* [91] algorithm and using the calibration device developed and manufactured by Cascination AG (Bern. Switzerland). The intrinsic parameters were computed from the video frames using OpenCV computer vision library[3].

Three scenarios were considered to evaluate the re-projection accuracy. First, the calibrated phantom presented in the publication [17] (see Section 3.3), with 28 metallic divots spread over a volume of $4320\,\mathrm{cm}^3$. Second, a flexible patient-specific liver phantom built by ARTORG research centre (Bern, Switzerland) [95], with 14 metallic M6 washers and a size of $1882\,\mathrm{cm}^3$. And lastly, a *in-vivo* porcine liver, with an approximate size of $2393\,\mathrm{cm}^3$. It is noteworthy the fact that, unlike with the validation phantom and the patient-specific phantom, the scans of the

---

[3]https://www.opencv.org/

Figure 6.2: Clinical setup for augmented reality-enhanced laparoscopic surgery [19].

in-vivo porcine liver did not show the landmarks used for the experiment, as these were done using a cauterizer through laparoscopy. Later, during the analysis of the data, these same locations were annotated by a surgeon on the virtual models, based on the laparoscope images of the intervention, to serve as ground truth to evaluate the re-projection error. As it will be discussed later, this had a major impact on the numerical results.

The same experiment was conducted in the three cases. Using the laparoscope to observe the scene and a tracked surgical tool, five randomly chosen landmarks are sampled and registered to the virtual model. The registration procedure was repeated ten times on each scenario. Accuracy was then assessed based on the fiducial registration error (FRE), computed from the landmarks used to perform the registration, and target registration error (TRE), using the remaining landmarks (see 2.2). Figure 6.3 shows a summary of the measurements taken during the study. The resulting FRE and TRE values were computed on 100 manually annotated AR frames.

Unlike with the validation and patient-specific liver phantoms, where the landmarks are visible in the computed tomography (CT) scans, the annotations in the *in-vivo* porcine liver are based on the anatomical understanding of the surgeon. Hence, both the FLE and TRE in the *in-vivo* case are significantly larger than in the two other scenarios.

Nonetheless, limitations of the presented study includes:
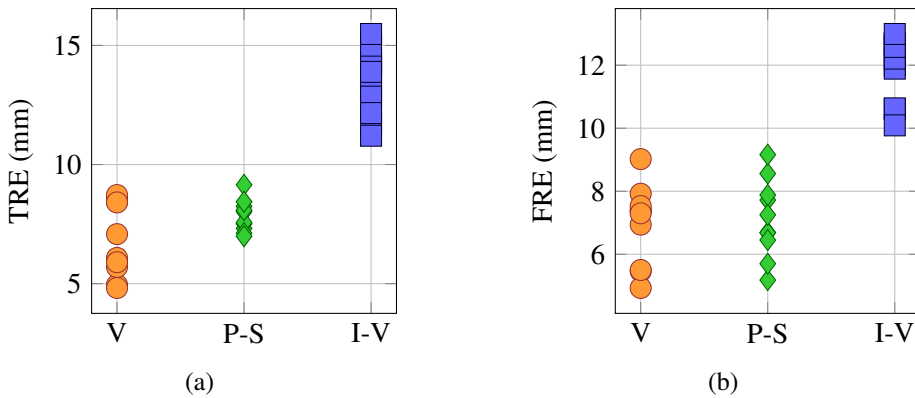
Figure 6.3: Average (a) target registration error (TRE) and (b) fiducial registration error (FRE) for the ten registration experiments computed on the validation phantom (V), patient-specific phantom (P-S), and *in-vivo* porcine liver (I-V).

- the lack of multiple surgeons to perform the annotations in the *in-vivo* case, to validate the influence of the annotations on the AR re-projection accuracy;

- the camera and hand-eye calibration, which were carried out as it would be in a surgical environment and could lead to inaccuracies;

- no analysis of the TRE inside the organ, all the landmarks were placed on the liver surface, no ground truth landmarks were placed underneath the parenchyma;

- and the use on an intra-operative CT scan to obtain the *in-vivo* porcine liver scans. Unlike pre-operative scans, the intra-operative virtual model obtained included the deformations due to pneumoperitoneum and mobilisation of the organ.

CHAPTER 7 DISCUSSION

In the presented research, the use and applications of registration in surgical navigation were explored. Starting with the hardware which enables the tracking and merging of the physical and digital realm, following with how these techniques are used in the operating room to power new tools like augmented reality; and finally exploring the use of machine learning, and more specifically, deep learning.

In Chapter 3 optical and electromagnetic tracking technologies we compared and tested to provide performance and tracking error baselines to be considered in registration. In this regard, the optical tracking system (OTS) yielded better results (p-value < 0.01) than the electromagnetic tracking system (EMTS), with a target registration error (TRE) 30% smaller than the latter (see Table 3.1). Whereas EMTS is based on the induction of the coils in the sensor, prone to be affected by metallic objects in the operating room, OTS uses a stereoscopic viewings system to triangulate the location of the reflective markers. Hence the improved tracking accuracy. Nevertheless, both technologies should be considered to enable surgical navigation, overcoming the shortcomings of each other. For instance, EMTS is well suited to track instruments inserted through the trocars, or flexible instruments such as a laparoscopic ultrasound. On the other hand, OTS should be considered as the main tracking technology due to its higher accuracy. Though it is limited to rigid instruments with reflective markers close to the handler i.e., visible to the stereoscopic cameras, or reflective markers on the surface of the patient, or surgical table, e.g., world reference frames.

Chapter 4 describes image-to-patient registration, the main registration method used for surgical navigation. This algorithm allows the alignment of the physical environment and digital images of the patient. The most extended method to

achieve this alignment is point-based registration or fiducial landmark registration method, in which pre-defined landmark are recorded on the patient using the aforementioned tracking technologies. The need of multiple landmarks (at least three), their accessibility and distribution, can result in a time-consuming task and result in subpar TRE. The Single Landmark Registration Method was proposed and tested to serve as an alternative method. As shown in Table 4.2, the proposed pipeline enables surgical navigation 40% quicker than the fiducial landmark-based registration (FLBR). However, a high TRE was observed after the initial registration, being the FLBR more accurate in terms of TRE (p-value < 0.001) (see Table 4.1). Nonetheless, this value is greatly improved after the refinement step when locating the lesion using the laparoscopic ultrasound (LUS).

Single landmark registration method (SLRM) is suited to be used with OTS and EMTS, using the former for the initial registration to enable navigation, and EMTS for tracking the imaging tool e.g., LUS, used to find the target lesion. Despite its ease of use and integration in the surgical workflow, SLRM is highly dependent on the initial registration. The orientation of the patient is kept constant throughout both registration steps. Hence, a poor acquisition can yield inaccurate visualisation of structures surrounding the lesion between the intra-operative and the pre-operative images. Further corrective techniques should be explored to take advantage of relevant landmarks in the vicinity of the tumour to refine the registration e.g., vascular structures.

As complex as the registration problem is, deep learning has been studied to solve the alignment of images, taking advantage of the capacity shown by convolutional neural networks (CNN) to extract and use large quantities of features extracted from the input images. The current trend shows a preference for autoencoder architectures, trained in a unsupervised or weakly-supervised fashion. Nonetheless, the biggest challenge when developing deep registration methods is, as with any other deep learning model, the access to large amounts of data samples. In clinical applications, this data can be challenging to acquire, even more to made available. Therefore, in Section 5.1 the deep registration training is analysed in two fronts: first, the training pipeline, where different combinations of loss functions are tested; and secondly, the use of transfer learning across modalities and anatomical domains, to extend deep registration models on small training datasets. Furthermore,

an augmentation layer which generates training pairs of fixed and moving images on-the-fly is made available to train the models on artificial datasets.

Tables 5.2 to 5.5 show how weakly-supervision with segmentation guidance is an effective strategy to train deep image registration models, further confirmed by the results shown in Figures 5.3 to 5.6. The segmentation boundaries introduce location information which leverages the model understanding of the regions occupied by each anatomical region. This observation is drawn by the improvement of the segmentation-based metrics on the fine-tuned models. However, no significant improvement was observed with the use of uncertainty weighing, compared to segmentation guidance. Uncertainty weighting was observed to dynamically change the contribution of the losses during training as show in Figure 7.1.
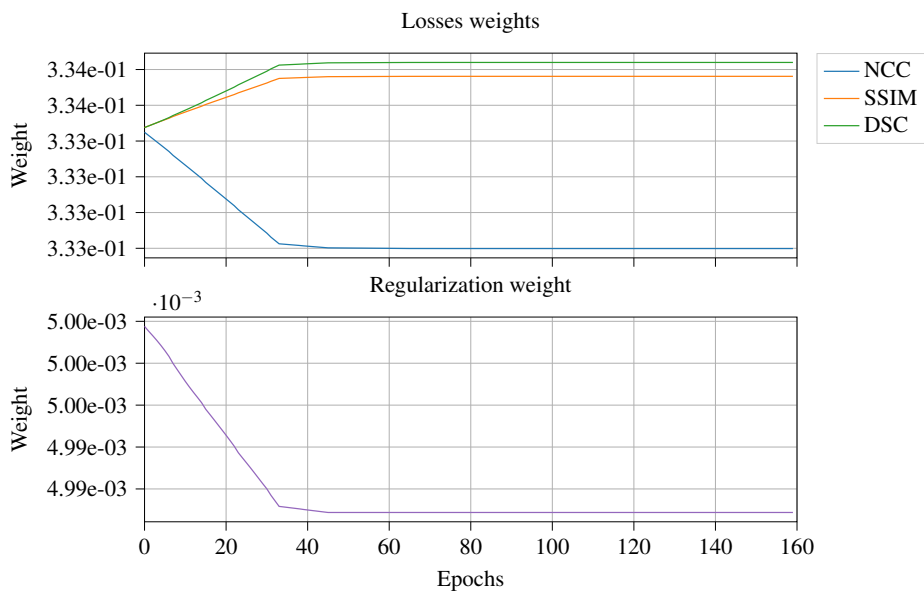
Figure 7.1: Evolution during training of the loss weights of the model trained with uncertainty-weighting, normalised cross-correlation, structural similarity index metric, and dice score coefficient (UW-NSD) trained on the Oslo-CoMet randomised controlled trial dataset (Oslo-CoMet).

Lastly, the human factor in point-based registration (PBR) methods for AR is addressed in Chapter 6. Three case scenarios were considered, from more controlled environment to closer to a real clinical scenario. First, using a machined

rigid phantom with metallic divots, which was accurately measured. Second, an anatomically realistic liver phantoms, with added metallic washers. Finally, an animal phantom was used, showing a more realistic situation. The landmarks used for registration were located in the 3D image of the model prior sampling by the surgeon. As anticipated, the results (see Figure 6.3) show a better registration in terms of fiducial registration error (FRE) and TRE for the machined rigid phantom, with easier to target landmarks. The animal model was the most challenging, as the operator had to rely on its anatomical understanding of the organ to locate the landmarks. This was particularly the case with those points situated close to the central region of the liver. These probed to be the most difficult to annotate, mostly due to the lack of geometrical cues to use as a reference. Whereas those landmarks located on the border of the parenchyma or close to significant anatomical landmarks, like the umbilical fissure, were easier to register.

It is worth mentioning the larger size and location of the landmarks of the animal model, scattered over the visible surface of the parenchyma. No landmarks were placed at the bottom of the liver, for accessibility reasons. The best results were obtained when the landmarks used for registration were spread all over the volume, i.e., not clustered, which suggest that evenly scattered targets for registration can compensate for the influence of sampling error. Yet further research is needed to draw conclusive statements, including a larger number of users, and controlled configurations of the landmark placement.

CHAPTER 8 CONCLUDING REMARKS

Image-to-patient and image-to-image registration techniques for laparoscopic surgery have been studied in the presented thesis, aiming to shed light on the research questions raised in Chapter 1. Stated the research done, we are in position of bringing an answer to these questions.

RQ 1 addressed the key components which would most significantly improve registration for surgical navigation. The intuitive notion of "improving registration" leads to think about accuracy, in terms of TRE. Paper I [17] (Chapter 3) and Paper III [19] (Chapter 6) two major key components were addressed: first, the technology used currently for tracking instruments, and second, the human factor. Favouring the use of OTS over EMTS can indeed lead to a better tracking accuracy, and hence, a better TRE when sampling the fiducials. Yet, the human factor still plays a major role, as seen in [19], determining the correct location of those predefined landmarks in the pre-operative image is bound to the judgement of the surgeon. Simplifying the procedure, or even automating the sampling of the landmarks e.g., by using surface matching, using the tracked laparoscopic ultrasound (LUS), could limit the impact of human error. As such, in Paper II [18] (Chapter 4), a simpler yet effective registration method was proposed, yielding promising results compared to the well-known point-based registration (PBR).

Deep registration, deep learning for registration, lies in the core of **RQ 2**. This concept was presented and analysed in Paper IV [20] (Chapter 5). As with most of the AI applications, the access to data is the major limiting factor to research. And as such, transfer learning was tested and shown to be effective to adapt deep registration models to different image modalities and anatomical structures. However, it is worth noting that in the presented study, this transfer was done between two 3D

and high resolution image modalities: MR and CT. Otherwise, as addressed in the publication, the resilience of trained models which prevents its direct application to other image modalities, is a factor worth addressing. Incremental learning, by introducing samples from different image modalities and structures in the training dataset, could be used to expand the applicability of the models. A final remark should be made regarding the reconstruction of the images, more advanced models should be considered in order to address this matter e.g., GANs or stable diffusion models would be worth researching.

When developing medical technology, in particular for surgical treatments, one has to consider the impact of introducing new technology in the clinical workflow, as addressed in **RQ3**. A highly disruptive procedure can risk being neglected by the surgeons, for the sake of the patient's health. In this regard, new registration procedures should be designed so the surgeon does not need to step out of its comfort zone. Even better, if it takes advantage of already established stages in the workflow e.g., using the location of the tracked LUS while the liver is being scanned to locate the lesion. This situation is well illustrated in the animal model scenario described in Paper III [19] (Chapter 6). Paper II [18] (Chapter 4) propose a new registration strategy which simplifies the procedure and can be integrated in the current clinical workflow (Chapter 5). Instead of sampling a collection of locations scattered over the patient's body, the surgeon just needs to sample one landmark and the orientation of the patient to obtain intra-operative navigation. Furthermore, while screening the organ with the LUS, a simple click on the target lesion shown on the US scan will co-register the pre-operative image with the intra-operative US, and the patient.

Lastly, **RQ 4** alludes the question most surgeons might be wondering: is deep-registration safe to be used? The results presented in Paper IV [20] (Chapter 5), although promising, are yet to be improved to meet the requirements for surgical applications. The first and most straightforward matter is, as already addressed, the reconstruction of the images. In Figures 5.4 to 5.6 it can be observed how rigid structures like the spine are deformed in the registered image. Even though these are outside the area of interest of the surgical procedure, it questions the reliability of the method. Further research is needed, with more and better-quality data showing real case samples, and exploring new architectures and training strategies.

Laparoscopic surgical navigation can, and is currently, playing a major role in the generalisation of minimally invasive surgery. And further technological developments increase the eligibility of patients to be safely treated using these techniques. As such, artificial intelligence, and deep image registration in particular, has the potential to boost the performance of navigation platforms, simplifying the registration step, and easing the burden on surgeons and radiologist. Yet, additional research is needed, driven by the increase of available clinical images and annotations to develop these powerful, yet data greedy, methods. At the risk of sounding boilerplate, the presented research is but a drop in a bast ocean of uncharted waters. The addressed topic is a multidisciplinary one where clinicians, engineers, and researchers must sit down at the same table, with their expertise, tools, and challenges. Otherwise, one risk creating technically-interesting-yet-impractical-to-use solutions, destined to spend the rest of its days on shelf. The coordination of such a picturesque group can surely yield exciting and new research lines, again, as long as collaboration leads the discussion.

> *"If I have seen further, it is by standing on the shoulders of giants."*
> —Sir Isaac Newton

BIBLIOGRAPHY

[1] Åsmund Avdem Fretland *et al.* "Open versus laparoscopic liver resection for colorectal liver metastases (the Oslo-CoMet study): Study protocol for a randomized controlled trial". In: *Trials* 16.1 (Mar. 2015), pp. 1–10. ISSN: 17456215. DOI: 10.1186/S13063-015-0577-5/FIGURES/2.

[2] Åsmund Avdem Fretland, Davit Aghayan and Bjørn Edwin. "Long-term survival after laparoscopic versus open resection for colorectal liver metastases." In: *Journal of Clinical Oncology* 37.18_suppl (June 2019), LBA3516–LBA3516. ISSN: 0732-183X. DOI: 10.1200/JCO.2019.37.18{_}suppl. LBA3516.

[3] Fernando A Alvarez, Sanchez Rodrigo Claria, Sebastian Oggero and Eduardo de Santibañes. "Parenchymal-sparing liver surgery in patients with colorectal carcinoma liver metastases". In: *World Journal of Gastrointestinal Surgery* 8.6 (2016), p. 407. ISSN: 1948-9366. DOI: 10.4240/wjgs.v8.i6.407.

[4] Joseph F. Buell *et al.* "The International Position on Laparoscopic Liver Surgery". In: *Annals of Surgery* 250.5 (2009), pp. 825–830. DOI: 10.1097/SLA.0b013e3181b3b2d8.

[5] Daniel Cherqui *et al.* "Laparoscopic Liver Resections: A Feasibility Study in 30 Patients". In: *Annals of Surgery* 232.6 (Dec. 2000), pp. 753–762. DOI: 10.1097/00000658-200012000-00004.

[6] Stavros A. Antoniou, George A. Antoniou, Athanasios I. Antoniou and Frank-Alexander Granderath. "Past, Present, and Future of Minimally Invasive Abdominal Surgery". In: *JSLS : Journal of the Society of Laparoendoscopic Surgeons* 19.3 (2015), e2015.00052. DOI: 10.4293/JSLS.2015.00052.

[7] Kaori Ito *et al.* "Laparoscopic versus open liver resection: A matched-pair case control study". In: *Journal of Gastrointestinal Surgery* 13.12 (2009), pp. 2276–2283. DOI: 10.1007/s11605-009-0993-5.

[8] Kevin Tri Nguyen, T Clark Gamblin and David A. Geller. "World Review of Laparoscopic Liver Resection—2,804 Patients". In: *Annals of Surgery* 250.5 (2009), pp. 831–841. DOI: 10.1097/SLA.0b013e3181b0c4df.

[9] Kevin Cleary and Terry M. Peters. "Image-Guided Interventions: Technology Review and Clinical Applications". In: *Annual Review of Biomedical Engineering* 12.1 (July 2010), pp. 119–142. ISSN: 1523-9829. DOI: 10.1146/annurev-bioeng-070909-105249.

[10] Fakhre Alam, Sami Ur Rahman, Sehat Ullah and Kamal Gulati. *Medical image registration in image guided surgery: Issues, challenges and research opportunities*. 2018. DOI: 10.1016/j.bbe.2017.10.001.

[11] Anna Rethy, Thomas Langø and Ronald Mårvik. *Laparoscopic ultrasound for hepatocellular carcinoma and colorectal liver metastasis: An overview*. Apr. 2013. DOI: 10.1097/SLE.0b013e31828a0b9a.

[12] Emery A. Minnard *et al.* "Laparoscopic ultrasound enhances standard laparoscopy in the staging of pancreatic cancer". In: *Annals of Surgery* 228.2 (Aug. 1998), pp. 182–187. ISSN: 00034932. DOI: 10.1097/00000658-199808000-00006.

[13] Kit Fai Lee *et al.* "Resection margin in laparoscopic hepatectomy: A comparative study between wedge resection and anatomic left lateral sectionectomy". In: *HPB* 12.9 (2010), pp. 649–653. ISSN: 14772574. DOI: 10.1111/j.1477-2574.2010.00221.x.

[14] David Martínez-Cecilia *et al.* "Impact of resection margins for colorectal liver metastases in laparoscopic and open liver resection: a propensity score analysis". In: *Surgical Endoscopy* 35.2 (Feb. 2021), pp. 809–818. ISSN: 0930-2794. DOI: 10.1007/s00464-020-07452-4.

[15] Zaed Z. R. Hamady *et al.* "One-Millimeter Cancer-Free Margin Is Curative for Colorectal Liver Metastases". In: *Annals of Surgery* 259.3 (Mar. 2014), pp. 543–548. ISSN: 0003-4932. DOI: 10.1097/SLA.0b013e3182902b6e.

[16] Sinara Vijayan *et al.* "Liver deformation in an animal model due to pneu-
moperitoneum assessed by a vessel-based deformable registration". In: *Min-
imally Invasive Therapy & Allied Technologies* 23.5 (Oct. 2014), pp. 279–
286. ISSN: 1364-5706. DOI: 10.3109/13645706.2014.914955.

[17] Andrea Teatini *et al.* "Assessment and comparison of target registration
accuracy in surgical instrument tracking technologies". In: *Proceedings
of the Annual International Conference of the IEEE Engineering in Medi-
cine and Biology Society, EMBS*. IEEE, July 2018, pp. 1845–1848. ISBN:
9781538636466. DOI: 10.1109/EMBC.2018.8512671.

[18] Javier Pérez de Frutos *et al.* "Laboratory test of Single Landmark registration
method for ultrasound-based navigation in laparoscopy using an open-source
platform". In: *International Journal of Computer Assisted Radiology and
Surgery* 13.12 (Dec. 2018), pp. 1927–1936. ISSN: 1861-6410. DOI: 10.1007/
s11548-018-1830-7.

[19] Andrea Teatini *et al.* "Influence of sampling accuracy on augmented reality
for laparoscopic image-guided surgery". In: *Minimally Invasive Therapy &
Allied Technologies* (Mar. 2020), pp. 1–10. ISSN: 1364-5706. DOI: 10.1080/
13645706.2020.1727524.

[20] Javier Pérez de Frutos *et al.* "Learning deep abdominal CT registration
through adaptive loss weighting and synthetic data generation". In: *PLOS
ONE* 18.2 (Feb. 2023). Ed. by Paolo Cazzaniga, e0282110. ISSN: 1932-6203.
DOI: 10.1371/journal.pone.0282110.

[21] Giovanni Landi and Alessandro Zampini. *Linear Algebra and Analytic Geo-
metry for Physical Sciences*. Undergraduate Lecture Notes in Physics. Cham:
Springer International Publishing, 2018. ISBN: 978-3-319-78360-4. DOI:
10.1007/978-3-319-78361-1.

[22] Rafael C. González and Richard E. Woods. *Digital Image Processing (3rd
Edition)*. Pearson Education Inc., 2006. ISBN: 013168728X.

[23] Antonio Barrientos, Luis Felipe Peñín, Carlos Balaguer and Rafael Santoja
Aracil. *Fundamentos de robótica*. McGraw-Hill, 2007. ISBN: 84-4815-636-6.

[24] Daniel Kleppner and Robert Kolenkow. *An Introduction to Mechanics*. 2nd
Editio. Cambbridge University Press, 2013. ISBN: 9780521198110.

[25]   Calvin R. Maurer Jr., Jennifer J. McCrory and J. Michael Fitzpatrick. "Estimation of accuracy in localizing externally attached markers in multimodal volume head images". In: *Medical Imaging 1993: Image Processing*. Ed. by Murray H. Loew. Vol. 1898. SPIE, Sept. 1993, pp. 43–54. DOI: 10.1117/12.154535.

[26]   C.R. Maurer *et al.* "Registration of head volume images using implantable fiducial markers". In: *IEEE Transactions on Medical Imaging* 16.4 (1997), pp. 447–462. ISSN: 02780062. DOI: 10.1109/42.611354.

[27]   J. M. Fitzpatrick and J. B. West. "The distribution of target registration error in rigid-body point-based registration". In: *IEEE Transactions on Medical Imaging* 20.9 (Sept. 2001), pp. 917–927. ISSN: 02780062. DOI: 10.1109/42.952729.

[28]   J. Michael Fitzpatrick. "Fiducial registration error and target registration error are uncorrelated". In: *Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*. Ed. by Michael I. Miga and Kenneth H. Wong. Vol. 7261. SPIE, Feb. 2009, p. 726102. DOI: 10.1117/12.813601.

[29]   J. Michael Fitzpatrick, Jay B. West and Calvin R. Maurer. "Predicting error in rigid-body point-based registration". In: *IEEE Transactions on Medical Imaging* 17.5 (1998), pp. 694–702. ISSN: 02780062. DOI: 10.1109/42.736021.

[30]   Derek L G Hill, Philipp G Batchelor, Mark Holden and David J Hawkes. "Medical image registration". In: *Physics in Medicine and Biology* 46.3 (Mar. 2001), R1–R45. ISSN: 0031-9155. DOI: 10.1088/0031-9155/46/3/201.

[31]   A Melbourne, G Ridgway and D J Hawkes. "Image similarity metrics in image registration". In: ed. by Benoit M. Dawant and David R. Haynor. Mar. 2010, p. 762335. DOI: 10.1117/12.840389.

[32]   Laurent Younes. *Shapes and Diffeomorphisms*. Vol. 171. Applied Mathematical Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019. ISBN: 978-3-662-58495-8. DOI: 10.1007/978-3-662-58496-5.

[33]   Robert M. Gray. *Entropy and Information Theory*. Boston, MA: Springer US, 2011, pp. 1–409. ISBN: 978-1-4419-7969-8. DOI: 10.1007/978-1-4419-7970-4.

[34] Monan Wang and Pengcheng Li. "A Review of Deformation Models in Medical Image Registration". In: *Journal of Medical and Biological Engineering* 39.1 (Feb. 2019), pp. 1–17. ISSN: 1609-0985. DOI: 10.1007/s40846-018-0390-1.

[35] Yabo Fu *et al. Deep Learning in Medical Image Registration: A Review*. Tech. rep. 2019.

[36] Grant Haskins, Uwe Kruger and Pingkun Yan. "Deep learning in medical image registration: a survey". In: *Machine Vision and Applications* 31 (2020). DOI: 10.1007/s00138-020-01060-x.

[37] Z. Wang, E.P. Simoncelli and A.C. Bovik. "Multiscale structural similarity for image quality assessment". In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Vol. 2. IEEE, 2003, pp. 1398–1402. ISBN: 0-7803-8104-1. DOI: 10.1109/ACSSC.2003.1292216.

[38] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity". In: *IEEE Transactions on Image Processing* 13.4 (Apr. 2004), pp. 600–612. ISSN: 1057-7149. DOI: 10.1109/TIP.2003.819861.

[39] Terry S. Yoo. *Insight into Images*. Ed. by Terry Yoo. A K Peters/CRC Press, Aug. 2004. ISBN: 978-1-56881-217-5. DOI: 10.1201/b10657.

[40] Zoran Gajic. *Linear Dynamic Systems and Signals*. Prentice Hall, 2003. ISBN: 0201618540.

[41] Liwei Wang, Yan Zhang and Jufu Feng. "On the Euclidean distance of images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (Aug. 2005), pp. 1334–1339. ISSN: 01628828. DOI: 10.1109/TPAMI.2005.165.

[42] E. Belogay, C. Cabrelli, U. Molter and R. Shonkwiler. "Calculating the Hausdorff distance between curves". In: *Information Processing Letters* 64.1 (Oct. 1997), pp. 17–22. ISSN: 00200190. DOI: 10.1016/S0020-0190(97)00140-3.

[43] Günter Rote. "Computing the minimum Hausdorff distance between two point sets on a line under translation". In: *Information Processing Letters* 38.3 (May 1991), pp. 123–127. ISSN: 00200190. DOI: 10.1016/0020-0190(91)90233-8.

[44] Patrik F. Raudaschl *et al.* "Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015". In: *Medical Physics* 44.5 (May 2017), pp. 2020–2036. ISSN: 00942405. DOI: 10.1002/mp.12197.

[45] Joon Lee *et al.* "Volumetric and Voxel-Wise Analysis of Dominant Intraprostatic Lesions on Multiparametric MRI". In: *Frontiers in Oncology* 9 (July 2019). ISSN: 2234-943X. DOI: 10.3389/fonc.2019.00616.

[46] Rita Simões *et al.* "Geometrical and dosimetric evaluation of breast target volume auto-contouring". In: *Physics and Imaging in Radiation Oncology* 12 (Oct. 2019), pp. 38–43. ISSN: 24056316. DOI: 10.1016/j.phro.2019.11.003.

[47] Zeyu Jiang, Changxing Ding, Minfeng Liu and Dacheng Tao. "Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11992 LNCS. Springer, Oct. 2020, pp. 231–241. ISBN: 9783030466398. DOI: 10.1007/978-3-030-46640-4{_}22.

[48] Davood Karimi and Septimiu E. Salcudean. "Reducing the Hausdorff Distance in Medical Image Segmentation With Convolutional Neural Networks". In: *IEEE Transactions on Medical Imaging* 39.2 (Feb. 2020), pp. 499–513. ISSN: 0278-0062. DOI: 10.1109/TMI.2019.2930068.

[49] Aston Zhang, Zachary C. Lipton, Mu Li and Alexander J. Smola. *Dive into Deep Learning*. 2020.

[50] Yann Lecun, Yoshua Bengio and Geoffrey Hinton. "Deep learning". In: *Nature 2015 521:7553* 521.7553 (May 2015), pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539.

[51] Ian Godfellow, Yoshua Bengio and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[52] Yipeng Hu *et al.* "Weakly-supervised convolutional neural networks for multimodal image registration". In: *Medical Image Analysis* 49 (Oct. 2018), pp. 1–13. ISSN: 13618415. DOI: 10.1016/j.media.2018.07.002.

[53] Alex Kendall, Yarin Gal and Roberto Cipolla. "Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, June 2018, pp. 7482–7491. ISBN: 978-1-5386-6420-9. DOI: 10.1109/CVPR.2018.00781.

[54] Alex Kendall and Yarin Gal. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" In: (Mar. 2017).

[55] Sebastian Ruder. "An Overview of Multi-Task Learning in Deep Neural Networks". In: (June 2017).

[56] Herman Alexander Jaeger *et al.* "Anser EMT: the first open-source electromagnetic tracking platform for image-guided interventions". In: *International Journal of Computer Assisted Radiology and Surgery* 12.6 (June 2017), pp. 1059–1067. ISSN: 1861-6410. DOI: 10.1007/s11548-017-1568-7.

[57] O.D. Faugeras and M. Hebert. "The Representation, Recognition, and Locating of 3-D Objects". In: *The International Journal of Robotics Research* 5.3 (Sept. 1986), pp. 27–52. ISSN: 0278-3649. DOI: 10.1177/027836498600500302.

[58] Andrew W Fitzgibbon. "Robust registration of 2D and 3D point sets". In: *Image and Vision Computing* 21.13-14 (Dec. 2003), pp. 1145–1153. ISSN: 02628856. DOI: 10.1016/j.imavis.2003.09.004.

[59] Teatini Andrea *et al.* "Validation of stereo vision based liver surface reconstruction for image guided surgery". In: *2018 Colour and Visual Computing Symposium (CVCS)*. IEEE, Sept. 2018, pp. 1–6. ISBN: 978-1-5386-5645-7. DOI: 10.1109/CVCS.2018.8496589.

[60] A. Seginer. "Rigid-body point-based registration: The distribution of the target registration error when the fiducial registration errors are given". In: *Medical Image Analysis* 15.4 (Aug. 2011), pp. 397–413. ISSN: 13618415. DOI: 10.1016/j.media.2011.01.001.

[61] Jay B. West and J. Michael Fitzpatrick. "Point-based rigid registration: clinical validation of theory". In: *Medical Imaging 2000: Image Processing*. Ed. by Kenneth M. Hanson. Vol. 3979. SPIE, June 2000, pp. 353–359. DOI: 10.1117/12.387697.

[62] P.J. Besl and Neil D. McKay. "A method for registration of 3-D shapes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (Feb. 1992), pp. 239–256. ISSN: 0162-8828. DOI: 10.1109/34.121791.

[63] Zhengyou Zhang. "Iterative point matching for registration of free-form curves and surfaces". In: *International Journal of Computer Vision* 13.2 (Oct. 1994), pp. 119–152. ISSN: 0920-5691. DOI: 10.1007/BF01427149.

[64] I. Reinertsen, M. Descoteaux, K. Siddiqi and D.L. Collins. "Validation of vessel-based registration for correction of brain shift". In: *Medical Image Analysis* 11.4 (Aug. 2007), pp. 374–388. ISSN: 13618415. DOI: 10.1016/j.media.2007.04.002.

[65] Martin A. Fischler and Robert C. Bolles. "Random sample consensus". In: *Communications of the ACM* 24.6 (June 1981), pp. 381–395. ISSN: 0001-0782. DOI: 10.1145/358669.358692.

[66] Andriy Myronenko and Xubo Song. "Point Set Registration: Coherent Point Drift". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (Dec. 2010), pp. 2262–2275. DOI: 10.1109/TPAMI.2010.46.

[67] Kyoto Kagau Co. Ltd. *Abdominal Intraoperative & Laparoscopic Ultrasound Phantom "IOUSFAN" | KYOTO KAGAKU*.

[68] Christian Askeland *et al.* "CustusX: an open-source research platform for image-guided therapy". In: *International Journal of Computer Assisted Radiology and Surgery* 11.4 (Apr. 2016), pp. 505–519. ISSN: 1861-6410. DOI: 10.1007/s11548-015-1292-0.

[69] Lars Eirik Bø, Erlend Fagertun Hofstad, Frank Lindseth and Toril A N Hernes. "Versatile robotic probe calibration for position tracking in ultrasound imaging". In: *Physics in Medicine and Biology* 60.9 (May 2015), pp. 3499–3513. ISSN: 0031-9155. DOI: 10.1088/0031-9155/60/9/3499.

[70]   W R Crum, T Hartkens and D L G Hill. "Non-rigid image registration: theory and practice". In: *The British Journal of Radiology* 77.suppl_2 (Dec. 2004), S140–S153. DOI: 10.1259/bjr/25329214.

[71]   Barbara Zitová and Jan Flusser. "Image registration methods: a survey". In: *Image and Vision Computing* 21.11 (Oct. 2003), pp. 977–1000. ISSN: 02628856. DOI: 10.1016/S0262-8856(03)00137-9.

[72]   Guorong Wu *et al.* "Unsupervised Deep Feature Learning for Deformable Registration of MR Brain Images". In: *Medical Image Computing and Computer Assisted Intervention*. Vol. 16. 2. 2013, pp. 649–656. ISBN: 3300000106. DOI: 10.1007/978-3-642-40763-5{_}80.

[73]   Max Jaderberg, Karen Simonyan, Andrew Zisserman and Koray Kavukcuoglu. "Spatial Transformer Networks". In: *Advances in Neural Information Processing Systems 28*. Ed. by C Cortes *et al.* Curran Associates, Inc., 2015, pp. 2017–2025.

[74]   Olaf Ronneberger, Philipp Fischer and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *arXiv* 9351 (May 2015), pp. 234–241. ISSN: 16113349. DOI: 10.1007/978-3-319-24574-4{_}28.

[75]   Hongming Li and Yong Fan. "Non-rigid image registration using self-supervised fully convolutional networks without training data". In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Vol. 2018-April. IEEE, Apr. 2018, pp. 1075–1078. ISBN: 978-1-5386-3636-7. DOI: 10.1109/ISBI.2018.8363757.

[76]   Guha Balakrishnan *et al.* "VoxelMorph: A Learning Framework for Deformable Medical Image Registration". In: *IEEE Transactions on Medical Imaging* 38.8 (Aug. 2019), pp. 1788–1800. ISSN: 0278-0062. DOI: 10.1109/TMI.2019.2897538.

[77]   Tony C.W. Mok and Albert C.S. Chung. "Fast Symmetric Diffeomorphic Image Registration with Convolutional Neural Networks". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020, pp. 4643–4652. ISBN: 978-1-7281-7168-5. DOI: 10.1109/CVPR42600.2020.00470.

[78] Yabo Fu *et al.* "LungRegNet: An unsupervised deformable image registration method for 4D-CT lung". In: *Medical Physics* 47.4 (Apr. 2020), pp. 1763–1774. ISSN: 00942405. DOI: 10.1002/MP.14065.

[79] Yang Lei *et al.* "4D-CT Deformable Image Registration Using an Unsupervised Deep Convolutional Neural Network". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11850 LNCS (Oct. 2019), pp. 26–33. ISSN: 16113349. DOI: 10.1007/978-3-030-32486-5{_}4.

[80] Jing Hu *et al.* "End-to-end multimodal image registration via reinforcement learning". In: *Medical Image Analysis* 68 (Feb. 2021). ISSN: 13618423. DOI: 10.1016/J.MEDIA.2020.101878.

[81] Asmund Avdem Fretland *et al.* "Laparoscopic Versus Open Resection for Colorectal Liver Metastases". In: *Annals of Surgery* 267.2 (Feb. 2018), pp. 199–207. ISSN: 15281140. DOI: 10.1097/SLA.0000000000002353.

[82] Gartner. *Understanding Gartner's Hype Cycles*. https://www.gartner.com/en/documents/3887767.

[83] Steve Aukstakalnis. *Practical augmented reality: a guide to the technologies, application, and human factors for AR and VR*. Addison-Wesley, Sept. 2016. ISBN: 9780134094236.

[84] Jeffrey H. Shuhaiber. "Augmented Reality in Surgery". In: *Archives of Surgery* 139.2 (Feb. 2004), p. 170. ISSN: 0004-0010. DOI: 10.1001/archsurg.139.2.170.

[85] Sylvain Bernhardt, Stéphane A. Nicolau, Luc Soler and Christophe Doignon. "The status of augmented reality in laparoscopic surgery as of 2016". In: *Medical Image Analysis* 37 (Apr. 2017), pp. 66–90. ISSN: 13618415. DOI: 10.1016/j.media.2017.01.007.

[86] Philip Pratt *et al.* "Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels". In: *European Radiology Experimental* 2.1 (Dec. 2018), p. 2. ISSN: 2509-9280. DOI: 10.1186/s41747-017-0033-2.

[87]   Rahul Prasanna Kumar *et al.* "Use of mixed reality for surgery planning: Assessment and development workflow". In: *Journal of Biomedical Informatics* 112 (Dec. 2020), p. 100077. ISSN: 15320464. DOI: `10.1016/j.yjbinx.2020.100077`.

[88]   Andrea Teatini, Rahul P. Kumar, Ole Jakob Elle and Ola Wiig. "Mixed reality as a novel tool for diagnostic and surgical navigation in orthopaedics". In: *International Journal of Computer Assisted Radiology and Surgery* (Feb. 2021). ISSN: 1861-6410. DOI: `10.1007/s11548-020-02302-z`.

[89]   Tobias Simpfendörfer *et al.* "Augmented Reality Visualization During Laparoscopic Radical Prostatectomy". In: *Journal of Endourology* 25.12 (Dec. 2011), pp. 1841–1845. ISSN: 0892-7790. DOI: `10.1089/end.2010.0724`.

[90]   Nazim Haouchine *et al.* "Impact of Soft Tissue Heterogeneity on Augmented Reality for Liver Surgery". In: *IEEE Transactions on Visualization and Computer Graphics* 21.5 (May 2015), pp. 584–597. ISSN: 1077-2626. DOI: `10.1109/TVCG.2014.2377772`.

[91]   Seongpung Lee *et al.* "Effective calibration of an endoscope to an optical tracking system for medical augmented reality". In: *Cogent Engineering* 4.1 (July 2017). Ed. by Zhongmin Jin. ISSN: 2331-1916. DOI: `10.1080/23311916.2017.1359955`.

[92]   Stephen Thompson *et al.* "Hand–eye calibration for rigid laparoscopes using an invariant point". In: *International Journal of Computer Assisted Radiology and Surgery* 11.6 (June 2016), pp. 1071–1080. ISSN: 1861-6410. DOI: `10.1007/s11548-016-1364-9`.

[93]   Mili Shah, Roger D. Eastman and Tsai Hong. "An overview of robot-sensor calibration methods for evaluation of perception systems". In: *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. New York, NY, USA: ACM, Mar. 2012, pp. 15–20. ISBN: 9781450311267. DOI: `10.1145/2393091.2393095`.

[94]   Z. Zhang. "A flexible new technique for camera calibration". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (Nov. 2000), pp. 1330–1334. ISSN: 01628828. DOI: `10.1109/34.888718`.

[95]   Alessia Pacioni *et al.* "Patient-specific ultrasound liver phantom: materials
       and fabrication method". In: *International Journal of Computer Assisted
       Radiology and Surgery* 10.7 (July 2015), pp. 1065–1075. ISSN: 1861-6410.
       DOI: 10.1007/s11548-014-1120-y.

PAPER I

Assessment and comparison of target registration accuracy in surgical instrument tracking technologies

A. Teatini, J. Pérez de Frutos, T. Langø, B. Edwin, O. J. Elle

Assessment and comparison of target registration accuracy in surgical instrument tracking technologies

Teatini A.[1,2], Pérez de Frutos J.[3,4], Langø T.[3], Edwin B.[1,5] , Elle O.[1,2] *Member, IEEE*

***Abstract: —* Image guided surgery systems aim to support surgeons by providing reliable pre-operative and intra-operative imaging of the patient combined with the corresponding tracked instrument location. The image guidance is based on a combination of medical images, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Ultrasonography (US), and surgical instrument tracking. For this reason, tracking systems are of great importance as they determine location and orientation of the equipment used by surgeons. Assessment of the accuracy of these tracking systems is hence of the utmost importance to determine how much error is introduced in image guided surgery only due to tracking inaccuracy. Thus, this study aimed to compare in a surgical Operating Room (OR) accuracy of the two most used tracking systems, Optical Tracking (OT) and Electromagnetic Tracking (EMT), in terms of Target Registration Error (TRE) assessment at multiple distances from the target position. Results of the experiments show that optical tracking performs more accurately in tracking the instrument tip than electromagnetic tracking in the experimental conditions. This was tested using a phantom designed for accuracy measurement in a wide variety of positions and orientations. Nevertheless, EMT remains a viable option for flexible instruments, due to its reliability in tracking without the need for line of sight.**

I. INTRODUCTION

In Image Guided Surgery (IGS), instrument tracking technologies are of paramount importance as they provide reliable information regarding position and orientation of surgical instruments and are also used to navigate the surgeon throughout the procedure. These technologies have been commonly used in neurosurgical applications [1]. However, in the past decade, application of IGS has expanded also to other surgical procedures, such as endoscopic sinus surgery [2] and laparoscopic nephrectomy [3]. Moreover, research is currently progressing to introduce image guidance in more complex procedures, such as laparoscopic liver resection or pancreatic ablation and resection [4], [5].

This paper is part of the research in image guided systems for laparoscopic liver resection surgery, as part of the H2020-MSCA-ITN EU project number 722068, High Performance Soft Tissue navigation (HiPerNav*). The liver represents one of the most frequent targets of metastases from other cancers, such as colorectal cancers, for which the estimated number of cases in a year is 550000 patients worldwide [6]. An IGS based on pre-operative CT or MRI scans would aid the surgeon towards the target metastases and spare as much healthy tissue as possible around the tumor. The most commonly used tracking systems for navigation in clinical research are OT and EMT systems. OT makes use of triangulation of light rays either reflected by a set of markers, or generated by infrared emitting diodes, to provide location and orientation of the tracked tool. However, in case of occlusion of the optical markers, i.e. loss of line of sight between cameras and markers, tracking information is interrupted. Instead, EMT makes use of magnetic fields to find fluxgate sensors or search coils, thus without relying on line of sight for tracking. This feature of EMT permits tracking of sensors inside the abdomen of the patient. Moreover, due to the small dimensions of the sensors, it is possible to position the sensor very close to the tool tip, and hence also track flexible instruments. Nonetheless, EMT sensors present a relatively small measurement volume with comparison to OT [7]. The EMT measurement volume can also be distorted and even reduced due to ferromagnetic interference of metallic objects close to the EMT field generator [8].

The purpose of this study is to assess and compare the accuracy of both OT and EMT in terms of Target Registration Error (TRE). TRE is a measure of accuracy in tracking a position on a tool at a certain distance from the sensors. For a surgical instrument, this equates to tracking the tip of the tool. TRE is defined [9] as follows:

$$\text{TRE}^2(x) = \frac{\text{FLE}^2}{N}\left(1 + \frac{1}{3}\sum_{k=1}^{3}\frac{d_k^2}{f_k^2}\right) \tag{1}$$

where x is an arbitrary target point, N the number of markers, d_k the distance x from each axis k, f_k the RMS of the distance and FLE stands for Fiducial Localization Error, which is a

[1] The Intervention Center, Oslo University Hospital, Oslo, Norway

[2] Department of Informatics, University of Oslo, Oslo, Norway

[3] Department of Health research, SINTEF AS, Trondheim, Norway

[4] Department of Computer Science, NTNU, Trondheim, Norway

[5] Oslo University Hospital, Department of Hepato-Pancreatic-Biliary surgery, Oslo, Norway

distance between a marker and its measured position [10]. TRE can be measured using a theoretical approach through (1) or via direct measurement using a phantom [11]. For this study, a direct measurement approach using a phantom was preferred over the theoretical approach to provide results closer to those of a realistic surgical scenario. Moreover, the experiment was carried out in an OR on a conventional OR table used for laparoscopic surgery in St. Olav's hospital, Trondheim.

II. MATERIALS AND METHODS

A. Phantom Design

ASTM standard F2554-10 for optical tracking accuracy measurement [12] was used as basis to design the phantom. The phantom design was improved to enable analysis of TRE also for electromagnetically tracked instruments. The phantom was built in polyoxymethylene thermoplastic (because of its hardness and low deformability), which prevents magnetic interference to EMT. Moreover, the phantom was designed to cover positions within a volume of interest (operating volume) approximately as in laparoscopic liver resection procedures. Hence, its dimensions are: 150 mm height, 200 mm width and 300 mm depth. A total of 71 target positions were machined onto the phantom, divided into 28 titanium target divot pins for TRE and 43 vertically drilled target holes. The titanium pins present a conical shape at 1 mm depth and 60° angle. The divot holes are designed to be complementary to a pointer adapter. The 43 vertical drilled holes are 2 mm in diameter and 10 mm deep, designed to hold an Northern Digital Inc.® (NDI) [13] EMT catheter sensor probe based on previous experiences from other studies [8], [14] but were not used throughout this study. The target positions, according to standard ASTM F2554-10, are placed at five different orientations, over a total of seven different planes.

Optical and electromagnetic reference sensors were rigidly connected through a bracket to the phantom to track the position of the phantom on the surgical table (NDI® 6DOF Reference Disk Part Number 610066 and Polaris Rigid Body Part Number 8700339). The reference support bracket allows for stable and repeatable repositioning of the reference sensors. The aforementioned pointer adapter was manufactured to allow testing of differently sized surgical instruments. This is also built in polyoxymethylene. It is designed as a cylindrical plastic holder with threaded holes located 120° apart. These are meant for tightening screws. A titanium tip, 1 mm in diameter and 3 mm in length, is found at the opposite side of the adapter, which fits perfectly in the titanium target pins of the phantom, as mentioned before. The phantom and tool adapter can be seen in Fig. 1. A MICROCUT® 837-I milling machine, equipped with Mitutoyo® AT115 Linear Scales and a metal tip to reach inside the divots were used to accurately measure the positions of the target divots on the phantom. The system comprises two 600 mm linear scales for displacements on the horizontal plane, and one 1100 mm linear scale on the vertical axis. The measuring accuracy is 8 μm for horizontal displacements and of 10 μm for vertical movements [15].

B. Experimental Protocol

The phantom was placed on a MAQUET® surgical table in the OR (Fig. 1). Both tracking systems were positioned in the configuration they will have during navigation in laparoscopic abdominal surgery, according to discussions with the surgeons. A Planar Field NDI Aurora® V2 system electromagnetic generator was rigidly connected to the surgical table using the field generator mounting arm of NDI® on the assistant side of the surgical table i.e., left patient supine side. This position of the EMT generator avoids occlusion line of sight of the interventional staff in visualizing anesthesia data and would only represent an insignificant burden to the assisting surgeon [8]. The optical tracking system NDI Polaris Spectra® was positioned at the head of the surgical table. The Phantom was positioned on the right side of the surgical table to replicate the position of the liver of the patient. The positions of the tracking systems were slightly adjusted in order to allow the measurement volumes of both tracking systems to maintain instrument tracking throughout all target positions of the phantom.

Accuracy of the tracking systems in terms of TRE was evaluated through two experiments. Experiment 1 examined the accuracy when the sensors, for EMT NDI 6DOF flex Tube 1.3 mm® Part Number 610060 (accuracy specifications for NDI® sensors can be found in reference material provided by NDI®) and for OT a custom built optical marker frame with NDI® passive infrared markers, were positioned on a rigid wooden instrument at four distances from the tool tip, classified as: A, B, C and D with increasing distances from the tool tip, as visible in Fig. 1. The positions of the sensors were extracted using the recordings from the NDI Toolbox® software. The tool tip was calibrated from the sensor positions through the pivoting functionality provided in the software. This reflects a real-life scenario in which tools are sterilized and calibrated through pivoting in the OR once clamped to the surgical instrument. The tracked tool tip was then shifted into
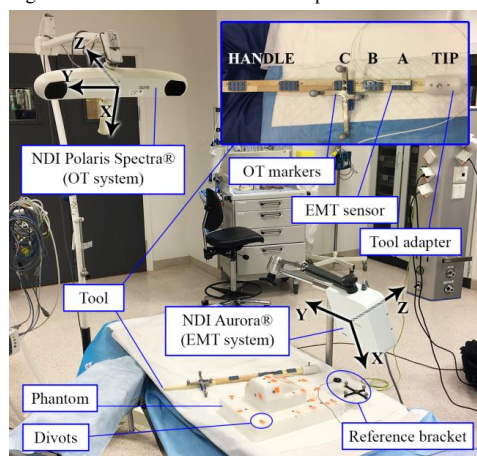

Fig. 1 Experimental Setup, with optical and electromagnetic tracking systems facing the TRE phantom.

1846

Table 1. Results of the study, errors in *x,y,z* axes are with respects to the axes of the Polaris cameras and Aurora generator, visible in Fig. 1.

| Tracking System | | Accuracy (mm) | | | | | | |
	Position	Pivot (RMS)	Pivot Error (RMS)	Error x	Error y	Error z	Mean ± SD	Max error
OT	A	99.98	0.31	1.27	1.64	2.90	1.94±0.57	2.90
	B	114.51	0.22	0.89	1.87	4.12	2.29±0.48	3.47
	C	162.82	0.33	1.55	1.39	3.54	2.16±0.54	3.06
	D	211.58	0.41	1.70	1.66	3.13	2.17±0.76	4.25
	HANDLE	410.88	0.57	1.57	1.70	3.97	2.41±0.63	3.50
	Average	*199.95*	*0.37*	*1.39*	*1.65*	*3.53*	**2.19±0.59**	*3.44*
EM	A	94.07	0.79	3.69	11.02	3.49	6.07±2.69	11.89
	B	110.32	0.73	4.06	9.99	4.22	6.09±2.78	12.14
	C	160.63	0.49	4.86	10.68	3.21	6.25±2.81	12.25
	D	210.90	0.38	4.72	11.00	2.86	6.25±2.77	12.29
	TIP	8.10	1.00	4.67	9.01	2.76	5.48±2.49	11.82
	Average	*116.80*	*0.68*	*4.40*	*10.34*	*3.31*	**6.03±2.71**	*12.08*

the 28 divot positions, and the trajectories were recorded for a total of 25 seconds for each position, of which 400 samples were used as data for the analysis to average the jitter errors of the tracking signal. TRE was calculated as the RMS difference of the measured divot positions and the tracked tip positions. The inaccuracy of the instrument is defined as the RMS of the errors of all three axes of the systems. Experiment 2 was performed to compare tracking accuracy in a similar scenario with laparoscopic tools. The optical marker plate was located at a distance of approximately 400 mm, which approximately represents the average distance from the tip to the handle of a laparoscopic instrument. Whereas, the EMT sensor was attached close to the tip of the tool.

The aim of this study was to test the hypothesis that there is a difference between TRE for EMT and OT tracked tools at multiple distances from the target tip. The two separate experimental results were analyzed using IBM SPSS software® [16].

III. RESULTS

For Experiment 1, a factorial ANOVA i.e. two-way analysis of variance, was conducted to compare the main effects of navigation technology and distance from target position and the effect of interaction between them on the accuracy. Levene's test of equality of variances deemed significant results, therefore a p-value of 0.001 was used to test significance of the results.

The main effect for tracking technology yielded significant results, with a F ratio of $F_{(1, 204)}=201.596$ and $p=2.8377E-32$ ($p<0.05$), indicating a significant difference between accuracy in millimeters of OT ($\mu=2.14$, $\sigma2=0.61$, $n=104$) and EMT ($\mu=6.15$, $\sigma2=2.78$, $n=108$). This is also visible in the results provided in Table 1. The main effect for position did not deem significant results, $F_{(3, 204)}=0.113$ with $p=0.952$ ($p > 0.05$) while comparing the 4 different distances tested for TRE. No significant interaction effects were found between the two factors.

With respects to Experiment 2, an independent samples T-test analysis of the data was performed to evaluate the TRE in both EMT and OT sensors with the EMT sensor at the tip and the OT at the length of a laparoscopic instrument. Levene's test for equality of variance revealed unequal variance between the two conditions $F_{(2.161)}=4.6E-5$, hence equal variances were not assumed for this experiment. Significant differences were found in the scores for EMT ($\mu=2.41$, $\sigma2=0.64$, $n=27$) and OT ($\mu=5.48$, $\sigma2=2.54$, $n=27$); $t(29.439)=-6.081$ with $p=1.2E-6$ ($p<0.05$). This reflects the results found for Experiment 1.

IV. DISCUSSION

Based on the study conducted in this paper, it can be inferred that OT performs more accurately than EMT, in terms of tracking the tip of the instrument. This was tested in several target positions on different orientation planes, throughout a precisely measured verification phantom. OT was found to be more accurate in all positions and orientations with respect to EMT. However, during the experimental tests, it was found that one of the positions of the phantom was not possible to track using optical tracking due to lack of line of sight between that position and the Polaris cameras, whereas this was not a problem for tracking using the EMT sensor.

OT showed reliable tracking information throughout all positions of the phantom, with an average TRE over the axes and distances from tip of 2.19 mm. This means that if the markers are placed approximately at a distance of 200 mm from the tip, the accuracy of optical tracking at the tip will be approximately around 2 mm, as reported in Table 1. The error is larger in the *z* axis of the Polaris® camera. This is due to error in triangulation of the light rays in depth of the camera and has been verified in multiple studies on OT accuracy [10].

The inaccuracy of EMT, with an average TRE of 6.03 mm, is dependent on several factors, which include: inaccuracy of the pivoting calibration to the tip of the instrument, influence of ferromagnetic equipment such as the surgical table and the tools present in the OR and inhomogeneity of the measurement volume of EMT generator. Pivot calibration was considered to

1847

be the largest contribution to the increase of error in EMT, as this is dependent both on the type of algorithm used and the sampling frequency of the tracking system. For this research, pivot calibrations were performed through collection in a 20 seconds interval for both systems. Thus, the lower sampling frequency of the Aurora® EMT system (40 Hz), compared to that of Polaris® (60 Hz), contributed to a less accurate calibration. The effect of interference from metallic instrumentation, including the surgical table, has been researched throughout several studies [7], [8], [14], [17] and calibration of the distortion is an ongoing topic to reduce this error [11].

Overall, in order to decrease the TRE for EMT tracked instrument tip positions, the suggestion is to position the EMT sensor as close as possible to the tracked position of interest. However pivot calibration for EMT tracked instruments should also be avoided to determine the offset relative to the tip since it can cause a larger error in OR settings: at 8 mm distance offset from the tool (position TIP in Table 1), the mean error was found to be 5.48 mm. Hence, the best solution could be to make use of surgical equipment with manufactured incorporated EMT sensors and which therefore provide a correct geometry of the tool. From Experiment 2, it is also deducible that OT represents a more accurate and reliable tracking technology, even when, for an EMT system, the sensor is positioned close to the tip (TIP) and the OT markers located at a distance equivalent to that of the handle of a laparoscopic instrument (position HANDLE in Table 1). Further studies should be conducted using TRE assessment for EMT tracked instruments without introduction of inaccuracy of pivoting calibration e.g. using tools with fixed geometries.

V. CONCLUSIONS

The presented study aimed to assess the accuracy of OT and EMT systems at multiple distances on surgical tools. This definition of accuracy is known as target registration error. TRE was analyzed through empirical testing using a precisely measured phantom in an OR to study navigation systems for surgical conditions. Results show the outperformance of OT systems compared to EMT. This was found to be valid also in conditions in which an EMT sensor is placed close to the tip of the surgical tool and the OT marker is clamped to the handle of the instrument. The experiments also showed the advantage of EMT of robustness to occlusions, which allowed this tracking system to sample a point which instead could not be tracked by the OT system.

Therefore, EMT technology is of great interest in laparoscopic environment to track flexible instruments. However, for rigid instruments such as a laparoscope camera or a grasper, OT systems provide greater accuracy at all distances along the instrument. The reported accuracy for EMT sensors is approximately 1 mm. However, this is only a measure of how accurate the sensor itself is, similarly to FLE [9], and not how accurately other parts of the instrument can be tracked based on the position of a sensor attached to the instrument. Therefore, further studies should be conducted to better assess TRE for EMT tools under multiple ferromagnetic influences e.g., C-arms, but also with multiple experimental conditions,

e.g. using NDI Tabletop Field Generator®. These results should also be compared to those of FLE for EMT sensors, similarly to what had been studied in [7], [8], [14], [17].

REFERENCES

[1] C. F. Freyschlag *et al.*, "Optical Neuronavigation without Rigid Head Fixation During Awake Surgery," *World Neurosurg.*, vol. 97, pp. 669–673, 2017.

[2] C. Winne, M. Khan, F. Stopp, E. Jank, and E. Keeve, "Overlay visualization in endoscopic ENT surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 6, no. 3, pp. 401–406, 2011.

[3] P. Mountney, J. Fallert, S. Nicolau, L. Soler, and P. W. Mewes, "An Augmented Reality Framework for Soft Tissue Surgery," *Med. Image Comput. Comput. Interv. -- MICCAI 2014*, vol. 8673, pp. 423–431, 2014.

[4] L. Bond, B. Schulz, T. VanMeter, and R. C. G. Martin, "Intra-operative navigation of a 3-dimensional needle localization system for precision of irreversible electroporation needles in locally advanced pancreatic cancer," *Eur. J. Surg. Oncol.*, vol. 43, no. 2, pp. 337–343, 2017.

[5] M. Feuerstein, T. Mussack, S. M. Heining, and N. Navab, "Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection," *IEEE Trans. Med. Imaging*, vol. 27, no. 3, pp. 355–369, 2008.

[6] F. X. Bosch, J. Ribes, M. Díaz, and R. Cléries, "Primary liver cancer: Worldwide incidence and trends," *Gastroenterology*, vol. 127, no. SUPPL., pp. 5–16, 2004.

[7] L. Maier-Hein *et al.*, "Standardized assessment of new electromagnetic field generators in an interventional radiology setting," *Med. Phys.*, vol. 39, no. 6Part1, pp. 3424–3434, 2012.

[8] L. E. Bø, H. O. Leira, G. A. Tangen, E. F. Hofstad, T. Amundsen, and T. Langø, "Accuracy of electromagnetic tracking with a prototype field generator in an interventional OR setting," *Med. Phys.*, vol. 39, no. 1, pp. 399–406, 2011.

[9] J. M. Fitzpatrick, "Fiducial registration error and target registration error are uncorrelated," vol. 7261, p. 726102, 2009.

[10] R. Elfring, M. De La Fuente, and K. Radermacher, "Assessment of optical localizer accuracy for computer aided surgery systems," *Comput. Aided Surg.*, vol. 15, no. 1–3, pp. 1–12, 2010.

[11] A. M. Franz, T. Haidegger, W. Birkfellner, K. Cleary, T. M. Peters, and L. Maier-Hein, "Electromagnetic tracking in medicine -A review of technology, validation, and applications," *IEEE Trans. Med. Imaging*, vol. 33, no. 8, pp. 1702–1725, 2014.

[12] ASTM, "Measurement of Positional Accuracy of Computer Assisted Surgical Systems 1," *Astm*, vol. i, pp. 1–11, 2015.

[13] © 2018 Northern Digital Inc., "NDI." [Online]. Available: https://www.ndigital.com/.

[14] Z. Yaniv, E. Wilson, D. Lindisch, and K. Cleary, "Electromagnetic tracking in the clinical environment," *Med. Phys.*, vol. 36, no. 3, pp. 876–892, 2009.

[15] Mitutoyo Corporation, "AT115 Linear Scales." p. 1.

[16] "SPSS IMB Statistics 25." [Online]. Available: https://www.ibm.com/products/spss-statistics.

[17] E. Shen, G. Shechter, J. Kruecker, and D. Stanton, "Quantification of AC electromagnetic tracking system accuracy in a CT scanner environment," vol. 6918, no. 914, p. 65090L, 2007.

PAPER II

Laboratory test of Single Landmark registration method for ultrasound-based navigation in laparoscopy using an open-source platform

J. Pérez de Frutos, E. F. Hofstad, O. V. Solberg, G. A. Tangen, F. Lindseth , T. Langø, O. J. Elle, R. Mårvik

ORIGINAL ARTICLE

CrossMark

Laboratory test of Single Landmark registration method for ultrasound-based navigation in laparoscopy using an open-source platform

Javier Pérez de Frutos[1] · Erlend F. Hofstad[1] · Ole Vegard Solberg[1] · Geir Arne Tangen[1] · Frank Lindseth[1,2] · Thomas Langø[1] · Ole Jakob Elle[3] · Ronald Mårvik[4]

Abstract

Purpose Test the feasibility of the novel Single Landmark image-to-patient registration method for use in the operating room for future clinical trials. The algorithm is implemented in the open-source platform CustusX, a computer-aided intervention research platform dedicated to intraoperative navigation and ultrasound, with an interface for laparoscopic ultrasound probes.

Methods The Single Landmark method is compared to fiducial landmark on an IOUSFAN (Kyoto Kagaku Co., Ltd., Japan) soft tissue abdominal phantom and T2 magnetic resonance scans of it.

Results The experiments show that the accuracy of the Single Landmark registration is good close to the registered point, increasing with the distance from this point (12.4 mm error at 60 mm away from the registered point). In this point, the registration accuracy is mainly dominated by the accuracy of the user when clicking on the ultrasound image. In the presented set-up, the time required to perform the Single Landmark registration is 40% less than for the FLRM.

Conclusion The Single Landmark registration is suitable for being integrated in a laparoscopic workflow. The statistical analysis shows robustness against translational displacements of the patient and improvements in terms of time. The proposed method allows the clinician to accurately register lesions intraoperatively by clicking on these in the ultrasound image provided by the ultrasound transducer. The Single Landmark registration method can be further combined with other more accurate registration approaches improving the registration at relevant points defined by the clinicians.

Keywords Registration · Laparoscopy · Computed-assisted surgery · Ultrasound · Multimodal visualization

Introduction and background

With the improvements in minimally invasive surgery techniques and instruments in recent years, there is a trend towards more use of the laparoscopic approach, although open surgery remains the gold standard for abdominal surgeries. Advantages of laparoscopic surgery include a less traumatizing intervention and a better post-operative phase for the patient, also decreased morbidity, quicker recovery,

less blood loss and improved long-term outcomes when compared to open surgery [1–4]. Nonetheless, there are concerns like risk of gas embolism due to pneumoperitoneum [2] or the limited space and field of view. To overcome the reduced field of view, the surgeons make use of a laparoscopic video camera for instrument guidance and other image modalities like ultrasound (US) for inspection and assessment of the lesion.

Laparoscopic ultrasound (LUS) was introduced originally by Yamakawa and co-workers in 1958 [5], providing real-time information of the inside of the organs. Jakimowicz and Reuers introduced LUS scanning for examination of the biliary tree during laparoscopic cholecystectomy in 1991 [6]. Since then, the use of LUS has expanded with the increase in minimally invasive procedures. Today, LUS is applied in a large number of procedures, such as screening for lymph nodes identification and tumour scanning; diagnostic detec-

Javier Pérez de Frutos
javier.perezdefrutos@sintef.no

1   Department of Health, SINTEF A.S., Trondheim, Norway

2   Computer Science Department, NTNU, Trondheim, Norway

3   The Intervention Centre, Oslo Rikhospital, Oslo, Norway

4   Department of Gastrointestinal Surgery, St. Olavs Hospital, Trondheim, Norway

Fig. 1 Navigation in laparoscopic surgery based on preoperative CT images

tion, localization and assessment of the extent of a tumour; and in therapeutic procedures as a guidance tool [7, 8].

With the technical improvements in image processing, computers and tracking systems, Image-Guided Navigation Platforms (IGNPs) emerged as an assisting tool for laparoscopic surgery. This software platform allows the surgeon to plan the operation beforehand [9] and also to have accurate and relevant information about the anatomy of the patient during surgery, with three-dimensional (3D) models of the anatomy and the used tools in the same view [10]. The combination of navigation and LUS will enable more soft tissue surgery in the abdomen to be performed with the laparoscopic technique. Tracked LUS together with registered preoperative data, e.g. computed tomography (CT), magnetic resonance imaging (MRI) or positron emission tomography (PET), provides a real-time US view matched to segmented models from preoperative data. This gives the surgeon an updated map of the target anatomy and structures during the procedure [11]. Navigated LUS also makes easier to relate the oblique two-dimensional (2D) US images to relevant anatomy.

Surgical margins are a major concern in hepatectomy interventions like hepatocellular carcinoma (HCC) and colorectal liver metastases (CLM) resections. The recommended surgical margin is of 2 cm for HCC and 1 cm for CLM [12]. IGNP can possibly enable surgeons to perform successful interventions with smaller resection margins, through the combination of medical images and intraoperative registration.

Image-to-patient registration is the first requirement to perform navigated LUS. This is spatially locating the preoperative images and the patient with respect to a common coordinate reference frame. For this purpose, tracking systems detect and compute the position and orientation of the tools and the patient in the operating room (OR), creating a virtual environment with a common coordinate reference frame. After completing the registration, the image information can be overlaid and shown together with the real-time position of the tools and the patient, allowing further navigation. Currently, there are four spatial tracking technologies being used in the OR: mechanical, optical, electromagnetic and acoustic [8, 13].

In this study, an optical tracking system was used to locate the tools and the liver phantom. Optical tracking systems typically consist of highly reflective markers or infrared emitting diodes attached to the patient (or OR table) and tools, infrared (IR) light sources to illuminate the reflective markers, IR cameras to detect the markers or diodes, and software that computes the position and orientation, i.e. tracking six spatial degrees of freedom, of the objects based on the spatial location of the markers.

Figure 1 shows the setting from a laparoscopic adrenalectomy using preoperative 3D CT images for the initial in-the-OR planning of the procedure, just before inserting the trocars. The view direction of the volume was then set by the view direction of the laparoscope as it was introduced. The LUS image could be displayed in the same scene, with an indication of the probe position using the open-source CustusX [9] platform.

This article presents the Single Landmark registration method (SLRM), as part of the open-source platform CustusX [9] for US-navigated laparoscopic surgery. This software allows the surgeon to integrate and fuse real-time LUS images with preoperative data, segmented models from med-

ical images, and models of tracked tools. The main novelty of this study is a fast and simple to use method for image-to-patient registration in the OR, validated on a soft tissue abdominal US compatible phantom. This is an incremental step to reach navigated LUS integrated in the clinical workflow. The ultimate goal is to efficiently combine all information sources to provide a real-time visualization of the anatomy of the patient combined with the tracked instruments.

Materials and methods

CustusX image-guided intervention platform

CustusX is an open-source IGNP developed by the research group at the Norwegian Centre for Innovative Ultrasound Solutions in Trondheim, Norway [9]. This platform integrates medical image visualization and real-time tracking of the surgical instruments, providing complete navigation for surgery in minimal invasive procedures. It also includes an interface to acquire real-time US images, which can be overlaid onto the virtual model of the patient.

The software is based on C++ and uses the Qt framework [14]. CustusX uses several external open-source libraries like VTK [15] for image visualization and processing, ITK [16] for segmentation and registration algorithms, and CTK [17] for processing DICOM files.

Single Landmark registration

The SLRM is a rigid image-to-patient registration algorithm that uses the orientation of a tracked tool and an anatomical reference point or landmark, for aligning the image data to the reference frame of the patient. The registration involves two phases: an initial registration using the orientation of the tool and a landmark that enables navigation, and the re-registration using the target lesion(s) intraoperatively. Although a surgical pointer is used here, the orientation and reference points can be acquired with any tracked instrument, as suggested in [11].

The algorithm assumes the tracked tool is oriented along the longitudinal axis of the patient and lying parallel to the coronal plane as suggested in Fig. 2. Because, in the prone orientation, the pointer will face downwards and might occlude the reflective markers, SLRM allows to specify whether the patient is in supine or prone position, so the pointer can be oriented upwards in both situations. Incorrect orientation of the tracked tool would result in misalignments between the virtual model and the patient anatomy reference frame. Therefore, the user is allowed to sample the orientation several times.

Fig. 2 Suggested location and orientation of the tracked tool (arrow), in this case, over the sternum of the patient
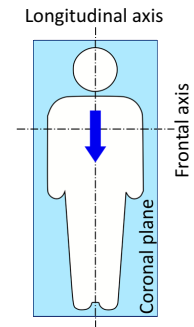
Reference points must be first manually marked in the preoperative data images, to perform the registration. By registering a reference point in both the patient and the 3D volume, the virtual model is rigidly translated (see Fig. 3a–d). Thus, there is an accurate match between the virtual model and anatomy in that point.

For the second phase of the method, new landmarks can be sampled during surgery, like tumours or anatomical structures. These new points can be used to re-register the patient model, improving the accuracy of the initial registration in a close neighbourhood of the point. Whenever a new point is registered, the transformation offset is updated to match such point but keeping the orientation constant. Therefore, the full potential of the SLRM can then be exploited using a LUS transducer, as it is the main tool used by surgeons to confirm the location of the lesions intraoperatively. Once the tumour is visible in the US image, the user can register the virtual volume including the lesion by clicking on the centre of the tumour shown in the US image. The platform allows to zoom in the US slice, improving the point sampling of the user and minimizing the effect of the screen resolution.

Figure 4a, b shows the procedure to register a lesion using the US image. The same virtual model as the one displayed in Fig. 3 is rendered translucent so the tumour (green point) can be seen. In Fig. 4a, the tumour shown corresponds to the US image on the right side. After clicking on the centre of the lesion in any of the US images (green arrow), the SLRM registers the selected tumour with the point clicked by the user.

Optical tracking has been used in this study. However, in a real situation surgical tools like the LUS would be electromagnetically tracked, as the optical tracking systems require line of sight and are not able to track the movements of the articulated tip once within the abdomen.

As aforementioned, the SLRM applies a rigid transformation on the virtual model. Therefore, anatomical movements like respiratory motion, or pneumoperitoneum, are not taken into account. These factors could result in deformations on the liver of several centimetres [18]. However, due to the local
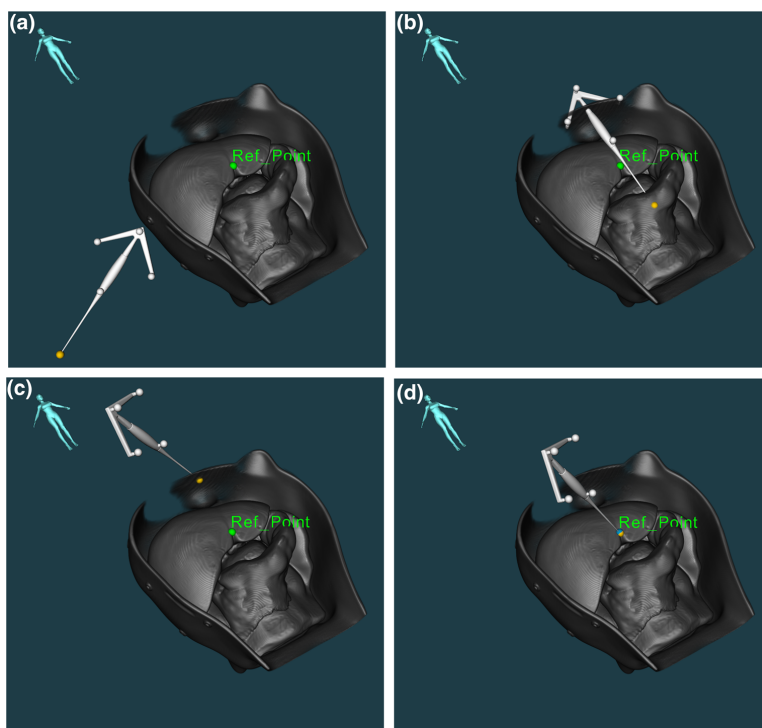
Fig. 3 SLRM image-to-patient registration steps: **a** initial location of tool as to sample the orientation; **b** the virtual model is oriented accordingly to the acquired orientation; **c** the reference point is marked with the pointer on the phantom; and **d** complete registration of the virtual model after manually sampling the reference point

registration using the LUS, the effect of these deformations can be reduced locally on the registered lesion.

Abdominal Intraoperative and Laparoscopic Ultrasound Phantom IOUSFAN

For this experiment, an Abdominal Intraoperative and Laparoscopic Ultrasound Phantom IOUSFAN (Kyoto Kagaku Co., Ltd., Japan) [19] was used (see Fig. 5). The phantom contains the most relevant abdominal structures and includes different types of lesions within each of them. The whole phantom is contained in a rigid case, where fiducial reference markers were attached before acquiring MR and CT scans.

Preoperative data

The MR T2 DICOM data were imported in the navigation system, though the same results could be obtained using other image modalities. A 3D reconstruction of the IOUS-FAN phantom is shown in Figs. 3 and 4.

Navigation equipment

US images were acquired with a SonixMDP US scanner (Ultrasonix Medical Corp., Richmond, Canada) and a L14-5/38 linear transducer (Prosonic Gyeongbuk, South Korea), as seen in Fig. 6a. A tracking frame with reflective markers was attached to the US probe and the table, becoming the latter the OR reference frame. A grid was fixed to the table with a resolution of centimetres, to measure the displacements of the phantom. A surgical pointer with reflective markers was used to register the landmarks. The POLARIS Spectra® (Northern Digital® Inc., Canada) and NDI® spherical passive retro-reflective markers were used for optical tracking [20]. Figure 6a and b shows the experiment set-up.

The US probe spatial calibration was verified using the evaluation wire phantom used in [21] and the Wire Widget
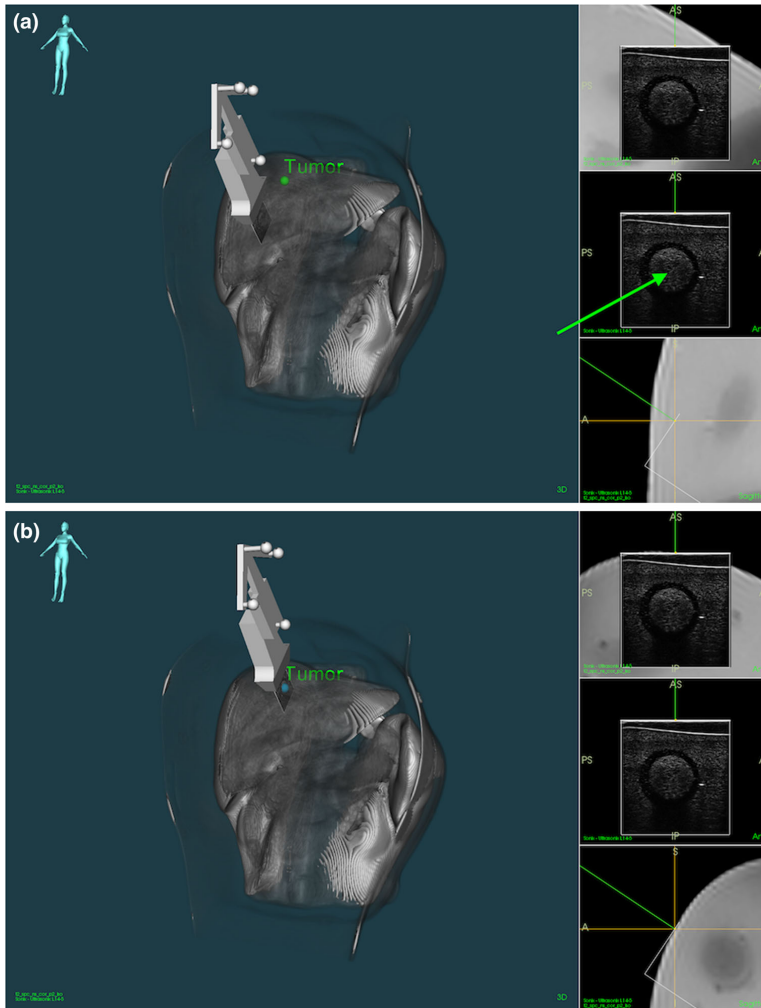
Fig. 4 SLRM registration of a lesion (green point) using the US image. **a** Before and **b** after the registration

of CustusX [9]. The verification process compares the centre-line of 3D US reconstructions of the wires with the calibrated virtual model, where the crossing point of the wires is used as the calibration point. Different trajectories were followed when scanning, acquiring between 100 and 200 US images per scan. The surgical pointer was calibrated using the pivot calibration option available in the Tool Tracker® application of NDI ToolBox® (20 s scan, 60 Hz), resulting in an average of 604 valid samples per scan. The average calibration errors are shown in Table 1. Both calibrations were done using POLARIS Spectra® for tracking.

Set-up

The IOUSFAN phantom liver is placed on the table, and the US probe is attached to the case so the same lesion is used for each displacement and sample (see Fig. 7b) and oriented to obtain a clear image of the lesion. Using the POLARIS Spectra® optical tracking system and the reference frame attached to the table, the US probe and the pointer are tracked and spatially located (see Fig. 6). The Ultrasonix scanner streams US images to CustusX through an OpenIGTLink [22] network.

Fig. 5 Abdominal Intraoperative and Laparoscopic Ultrasound Phantom IOUSFAN [19]

Table 1 Calibration errors in millimetres of the US probe and the surgical pointer

Instrument	US probe	Surgical pointer
Average	0.21	0.44
Standard deviation	0.49	0.05

Experiment protocol

The aim of this experiment was to test the feasibility of the SLRM to be used in future clinical trials. For this, SLRM was compared to fiducial-based rigid landmark registration method (FLRM). For both methods, the phantom is moved ten times by 10, 50 and 100 mm in the frontal and longitudinal axes, independently. After each displacement, the virtual model is re-registered to correct the displacement using the same reference points as in the first image-to-patient registration. The tumour location, tracked with the US, is then used for verification. The centre of the tumour is manually marked using the US image and compared to the position of the same tumour manually marked in the virtual model. The chosen lesion is located 60 mm away from the registered point.

The initial image-to-patient SLRM is performed by leaving the pointer over the liver and oriented as in Fig. 2 (see Fig. 7a). Then, the reference point is registered using the pointer. Because of its easy access, the reference point chosen for this experiment is where the round and the falciform

Fiducial points are manually marked in the virtual model using CustusX. The image-to-patient registration is accomplished using the surgical pointer to sample the landmarks and the orientation of the phantom.
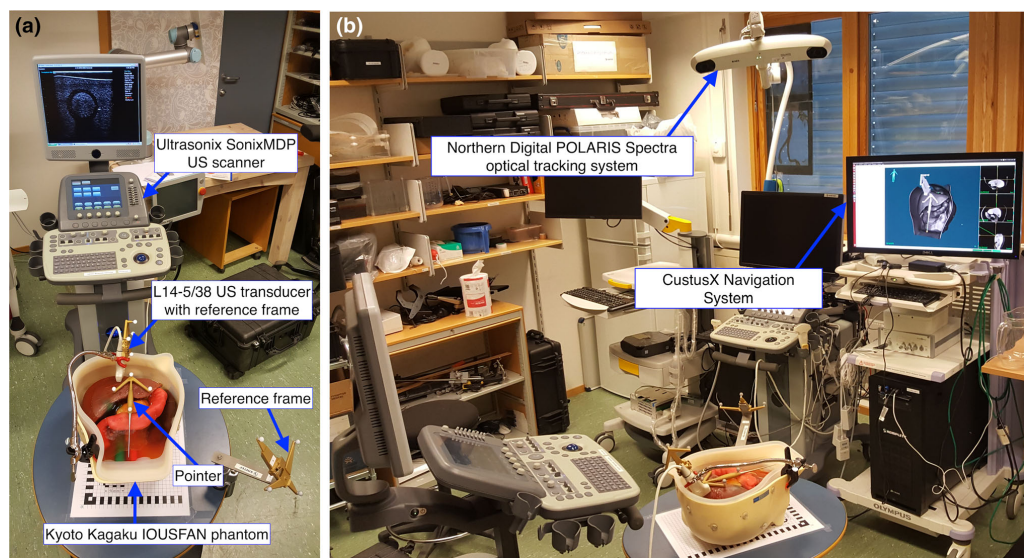

Fig. 6 **a** SonixMDP US scanner, IOUSFAN and tools and frames with optical markers; **b** POLARIS optical tracking system and CustusX navigation system
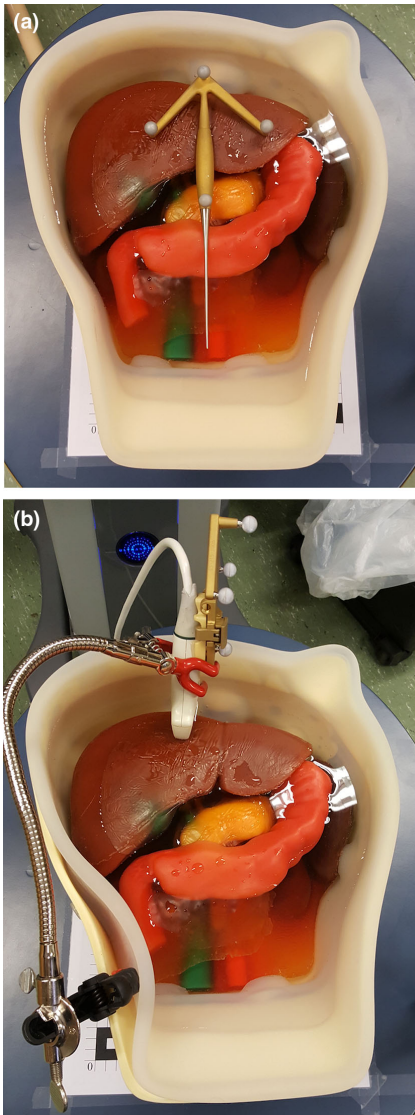
For comparison, a FLRM is done using five fiducial markers distributed over the case of the phantom. The same procedure is followed as with the SLRM, performing a complete registration after each displacement and tracking the location of the tumour using the US transducer.

User time, i.e., time required by the user to perform an image-to-patient registration, is measured for each registration performed using a chronometer. For the SLRM, the time measured corresponds to that between the moment the orientation of the tool is recorded and when the reference point is registered. In the case of the FLRM, the user time is measured between the sampling of the first and the fifth fiducial points.

Experimental results

A total of 60 target registration error (TRE) [23] samples were computed for each registration method. The TRE is measured as the Euclidean distance between the centre of the tumour, found using the US probe, and the location of this same lesion in the virtual model, for each displacement in the frontal and longitudinal axes (see Experiment protocol). Table 2 shows the average TRE results. These same values are plotted in Fig. 8a and b where it can be seen that the average TRE does not vary greatly with the direction the distance displaced. The repeatability of each group is computed as the standard deviation of the mean.

The data were analysed using IBM® SPSS® Statistics Version 25 software [24]. A p value below 0.05 was considered statistically significant. SLRM TRE data were analysed using a one-way ANOVA. Variables were categorized in six groups according to the displacement and the axis, i.e. X10, X50, X100, Y10, Y50 and Y100. Levene's test showed no difference between the variances of the groups [$F(5, 54) = 2.226, p = 0.065$]. The ANOVA deemed statistical difference between the means [$F(5) = 2.437, p = 0.046$]. However, multiple composition analysis using Tukey's honest significant difference and Scheffe's method showed no statistical difference between pairs of groups, with a significance level of 95%. Therefore, the means of the groups are statistically similar, and thus, SLM is robust against displacements of the patient.

A one-way ANOVA was used to compare the TRE for the two registration methods. The samples were grouped according to the method, i.e. SLRM or FLRM. Variances of the groups were found to be statistically different [$F(1, 118) = 4.34, p = 0.039$]. Therefore, a Welch test of equal means was performed, resulting in statistically dissimilar means [$F(1, 113.397) = 5004.32, p = 0.000$]. So, the TRE performance of SLRM and FLRM is statistically different, FLRM showing the best results.

Fig. 7 **a** Optical pointer placed on the phantom to perform the image-to-patient SLRM registration. **b** US probe attached to the phantom

ligament meet on the liver (*Ref_Point* in Fig. 3a, d). This same reference point is used for the re-registration after each displacement. Nonetheless, as long as the fiducial point can be located in both the virtual model and the patient anatomy, users are free to choose any more accessible fiducial point.

Table 2 TRE between the tumour visualized the US image and in the MRI scan, using SLRM and FLRM

Displacement	10 mm		50 mm		100 mm	
Displacement axis	Frontal	Longitudinal	Frontal	Longitudinal	Frontal	Longitudinal
SLRM						
Average	11.3	11.1	11.3	11.1	10.7	11.3
Standard deviation	0.4	0.7	0.5	0.4	0.4	0.5
Minimum	10.5	10.1	10.3	10.6	10.1	10.4
Maximum	11.8	12.4	11.9	12.0	11.3	12.0
Repeatability	0.11	0.23	0.17	0.12	0.13	0.17
FLRM						
Average	4.6	4.7	4.6	5.1	4.4	5.2
Standard deviation	0.2	0.4	0.3	0.4	0.3	0.3
Minimum	4.4	4.3	4.2	4.7	3.7	4.7
Maximum	5.0	5.5	5.1	5.8	4.9	5.6
Repeatability	0.05	0.09	0.11	0.11	0.13	0.08
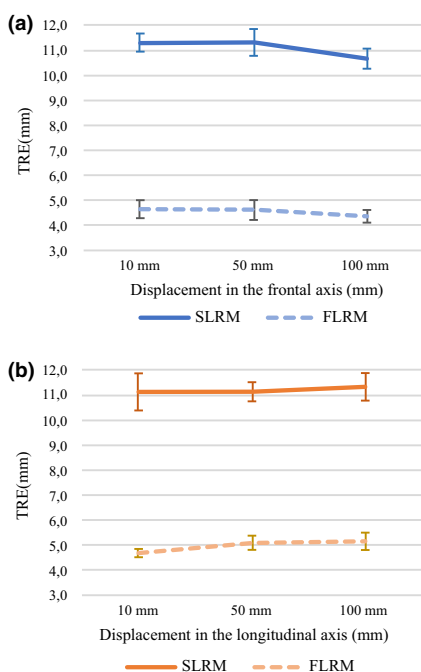
Fig. 8 TRE values shown in Table 2 for the SLRM and FLRM in the **a** frontal axis and the **b** longitudinal axis

Table 3 User time in seconds for the SLRM and FLRM

Registration method	FLRM	SLRM
Average	19.63	7.62
Standard deviation	1.68	0.63
Repeatability	0.22	0.08

the test proved the means between the groups to be different $[t(75.292) = 51.627, p = 0.000]$. Therefore, there is a significant difference in time required between the two methods, SLRM being faster than FLRM (40% for the current experiment).

Discussion

In this study, the novel SLRM image-to-patient intraoperative registration method is introduced and validated for clinical use. It is currently implemented in CustusX [9] IGNP. The experiments tested the capability of the algorithm to perform intraoperative image-to-patient registration, with special focus on the complexity of the steps and the required user time. Experimental set-up comprised of an L14-5/38 linear transducer connected to an UltraSonix SonixMDP scanner, a surgical pointer, a POLARIS Spectra® optical tracker, and an IOUSFAN soft tissue abdominal phantom. Tool calibration was conducted as described in navigation equipment section. Because the US probe is fixed to the phantom case, only the spatial calibration was verified.

During the experiment, the surgical pointer is used for setting the orientation, speeding the initial registration process. However, it is possible to perform a conventional patient registration, by using several fiducials or contour registration, and then keep the resulting orientation of the virtual model

A total of 60 samples of time were measured for each method. Statistics are shown in Table 3. The time samples were classified in SLRM and FLRM and analysed using a T-test analysis. Statistical differences were found between the variances of the groups $[F(1, 118) = 29.994, p = 0.000]$, and

instead of using a pointer. This may result in a more correct orientation, improving accuracy during the re-registration phase in a larger area around the registered point.

The multiple composition statistical analysis of the TRE shows no correlation between the SLRM TRE performance and the displacement of the phantom. The reported repeatability values are considered adequate for the presented study. It is also shown that the TRE is highly correlated to the registration method, i.e. SLRM or FLRM, FLRM showing the best results. The user time shows a statistical difference between methods, being the SLRM faster than the FLRM (see Table 3).

Therefore, it is concluded that the SLRM image-to-patient registration is suitable to be integrated in a laparoscopic intervention workflow in combination with a tracked LUS. The use of SLRM will result in a major improvement in terms of time consumption without compromising the TRE close to the registered point. The strategy is to use this simple and efficient method as a starting point for an intraoperative fine-tuning registration method based on 3D US data acquired by the tracked 2D LUS probe, while staging the liver in the initial phase before resection.

However, it must be considered that the algorithm assumes that the user provides the correct orientation of the patient and the correct location of the reference point. Therefore, the SLRM TRE is highly dependent on the precision with which the user samples these parameters. Discrepancies between the preoperative scans and the position and location of the patient anatomy during the operation, e.g. anatomical shift involving rotation, when the liver is mobilized, will be an additional source of error to be considered. Also, the point-sampling accuracy of the user when clicking on the US slice would affect the outcome of the registration, though this effect is mitigated with the possibility of zoom in the US image to better aim for the desired point. Further experimentation would be required to quantify the effect of these error sources; however, this is out of the scope of the current study. Human accuracy with laparoscopic tools and computer input devices have been studied in [25–27].

To improve the accuracy at different regions, the surgeon is able to re-register new points of such areas that can be visualized in the LUS images. The method matches the virtual model with the updated location of the intraoperatively registered point. Therefore, the proposed method allows the surgeon to accurately register lesions, especially in situations where FLRM cannot be performed, e.g. due to large differences between the preoperative data and the situation in the OR.

As such, future work will focus on combining the SLRM algorithm with more robust or accurate registration methods, e.g. matching corresponding vessel structures in the virtual model and US, where the SLRM could be beneficial to locally improve the TRE within a close area around the lesion. Fur-

thermore, deformable registration methods could be used to further improve the TRE around the registered point using SLRM.

Conclusion

This study introduces the novel SLRM using an open-source platform for US-based navigation in laparoscopy. SLRM has been shown suitable to be integrated in the normal workflow of a laparoscopic surgical procedure. Furthermore, the accuracy of the tracking system, the calibration and the registration are well within the recommended surgical margin limits for hepatectomy interventions [12, 21].

Reduced user time and simple steps are two of the principal advantages of the proposed method, together with the possibility of performing the image-to-patient registration during the preparation of the patient and intraoperatively using the LUS. Thus, providing real-time and accurate information of the anatomy of the organ to the surgeon in the neighbourhood of the registered point, where other registration methods may not be applicable.

In the current implementation, the SLRM performs rigid transformations over the image data. Thus, the algorithm is sensitive to factors such as the point-sampling accuracy of the user when clicking on the centre of the lesion.

Therefore, future work will focus on integrating SLRM with more robust registration techniques. Furthermore, SLRM can be combined with deformable registration algorithms to improve accuracy in real time in the region around the lesion.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Velanovich V (2000) Laparoscopic vs open surgery: a preliminary comparison of quality-of-life outcomes. Surg Endosc 14(1):16–21
2. Zhang Y, Chen X-M, Sun D-L (2015) Comparison of laparoscopic versus open left lateral segmentectomy. Int J Clin Exp Med 8(1):904–909
3. Kazaryan AM, Pavlik Marangos I, Rosseland AR, Røsok BI, Mala T, Villanger O, Mathisen Ø, Giercksky K-E, Edwin B (2010) Laparoscopic liver resection for malignant and benign lesions: ten-year Norwegian single-center experience. Arch Surg 145(1):34–40
4. Buell JF, Thomas MT, Rudich S, Marvin M, Nagubandi R, Ravindra K, Brock G, McMasters KM (2008) Experience with more than 500 minimally invasive hepatic procedures. Ann Surg 248(3):475–486
5. Yamakawa K, Naito S, Azuma K, Wagai T (1958) Laparoscopic diagnosis of the intra-abdominal organs. Jpn J Gastroenterol Surg 55(3):741
6. Jakimowicz JJ, Ruers T (1991) Ultrasound-assisted laparoscopic cholecystectomy: preliminary experience. Dig Surg 8(2):114–117
7. Jakimowicz JJ (2006) Intraoperative ultrasonography in open and laparoscopic abdominal surgery: an overview. Surg Endosc Interv Tech 20(2):S425–S435
8. Langø T, Vijayan S, Rethy A, Våpenstad C, Solberg OV, Mårvik R, Johnsen G, Hernes TN (2012) Navigated laparoscopic ultrasound in abdominal soft tissue surgery: technological overview and perspectives. Int J Comput Assist Radiol Surg 7(4):585–599
9. Askeland C, Solberg OV, Bakeng JBL, Reinertsen I, Tangen GA, Hofstad EF, Iversen DH, Våpenstad C, Selbekk T, Langø T, Hernes TAN, Olav Leira H, Unsgård G, Lindseth F (2016) CustusX: an open-source research platform for image-guided therapy. Int J Comput Assist Radiol Surg 11(4):505–519
10. Galloway R, Peters T (2008) Overview and history of image-guided interventions. In: Peters T (ed) Image-guided interventions: technology and applications. Springer, Boston, pp 1–21
11. Mårvik R, Langø T, Tangen GA, Andersen JO, Kaspersen JH, Ystgaard B, Sjølie E, Fougner R, Fjøsne HE, Hernes TAN (2004) Laparoscopic navigation pointer for three-dimensional image-guided surgery. Surg Endosc Interv Tech 18(8):1242–1248
12. Rethy A, Langø T, Ronald M (2013) Laparoscopic ultrasound for hepatocellular carcinoma and colorectal liver metastasis: an overview. Surg Laparosc Endosc Percutaneous Tech 23(2):135–144
13. Simon D (1997) Intra-operative position sensing and tracking devices. In: Proceedings of the first joint CVRMed/MRCAS conference
14. The Qt Company (2013) Qt, Oct 2013. [Online]. www.qt.io
15. Kitware Inc. (2002) VTK, July 2002. [Online]. www.vtk.org
16. Kitware Inc. (2000) ITK, Mar 2000. [Online]. www.itk.org
17. Nolden M, Zelzer S, Seitel A, Wald D, Müller M, Franz AM, Maleike D, Fangerau M, Baumhauer M, Maier-Hein L, Maier-Hein KH, Meinzer H-P, Wolf I (2013) The medical imaging interaction toolkit: challenges and advances. Int J Comput Assist Radiol Surg 8(4):602–620
18. Vijayan S, Reinertsen I, Hofstad EF, Rethy A, Hernes TAN, Langø T (2014) Liver deformation in an animal model due to pneumoperitoneum assessed by a vessel-based deformable registration. Minim Invasive Ther Allied Technol 23(5):279–286
19. L. Kyoto Kagaku Co., Abdominal Intraoperative & Laparoscopic Ultrasound Phantom "IOUSFAN", [Online]. www.kyotokagaku.com/products/detail03/us-3.html. Accessed Dec 2017
20. Northern Digital Inc. (2017) Polaris Spectra and Vicra, Northen Digital Inc., Mar 2017. [Online]. www.ndigital.com/medical/products/polaris-family/
21. Bø LE (2015) Versatile robotic probe calibration for position tracking in ultrasound imaging. Phys Med Biol 60:3499–3513
22. Tokuda J, Fischer GS, Papademetris X, Yaniv Z, Ibanez L, Cheng P, Liu H, Blevins J, Arata J, Golby AJ, Kapur T, Pieper S, Burdette EC, Fichtinger G, Tempany C (2010) OpenIGTLink: an open network protocol for image-guided therapy environment. Int J Med Robot 5(4):423–434
23. Fitzpatrick JM, West JB (2001) The distribution of target registration error in rigid-body point-based registration. IEEE Trans Med Imaging 20(9):917–926
24. IBM (2017) IBM SPSS statistics, IBM, [Online]. www.ibm.com/products/spss-statistics. Accessed Apr 2017
25. MacKenzie IS, Kauppinen T, Solfverberg M (2001) Accuracy measures for evaluating computer pointing devices. In: Proceedings of the ACM conference on human factors in computing systems, New York
26. International Organization for Standardization (2007) Part 400: Principles and requirements for physical input devices. In: ISO 9241 (ed) Ergonomics of human-system interaction. International Organization for Standardization, Geneza, p 35
27. Trejo A, Jung M-C, Oleynikov D, Hallbeck MS (2007) Effect of handle design and target location on insertion and aim with a laparoscopic surgical tool. Appl Ergon 38(6):745–753

PAPER III

Influence of sampling accuracy on augmented reality for laparoscopic image-guided surgery

A. Teatini, J. Pérez de Frutos, B. Eigl, E. Pelanis, D. L. Aghayan, M. Lai, R. P Kumar, R. Palomar, B. Edwin, O. J. Elle

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

🔓 OPEN ACCESS   Check for updates

Influence of sampling accuracy on augmented reality for laparoscopic image-guided surgery

Andrea Teatini[a,b] 🔟, Javier Pérez de Frutos[c,d], Benjamin Eigl[e], Egidijus Pelanis[a,f], Davit L. Aghayan[a,f,g] 🔟, Marco Lai[h], Rahul Prasanna Kumar[a] 🔟, Rafael Palomar[a,d] 🔟, Bjørn Edwin[a,f,i] 🔟 and Ole Jakob Elle[a,c] 🔟

[a]The Intervention Centre, Oslo University Hospital Rikshospitalet, Oslo, Norway; [b]Department of Informatics, University of Oslo, Oslo, Norway; [c]SINTEF Digital, SINTEF A.S, Trondheim, Norway; [d]Department of Computer Science, NTNU, Trondheim, Norway; [e]CAScination, Bern, Switzerland; [f]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [g]Department of Surgery N1, Yerevan State Medical University, Yerevan, Armenia; [h]Philips Research, High Tech, Eindhoven, The Netherlands; [i]Hepato-Pancreatic-Biliary Surgery, Oslo University Hospital, Oslo, Norway

ABSTRACT

Purpose: This study aims to evaluate the accuracy of point-based registration (PBR) when used for augmented reality (AR) in laparoscopic liver resection surgery.

Material and methods: The study was conducted in three different scenarios in which the accuracy of sampling targets for PBR decreases: using an assessment phantom with machined divot holes, a patient-specific liver phantom with markers visible in computed tomography (CT) scans and *in vivo*, relying on the surgeon's anatomical understanding to perform annotations. Target registration error (TRE) and fiducial registration error (FRE) were computed using five randomly selected positions for image-to-patient registration.

Results: AR with intra-operative CT scanning showed a mean TRE of 6.9 mm for the machined phantom, 7.9 mm for the patient-specific phantom and 13.4 mm in the *in vivo* study.

Conclusions: AR showed an increase in both TRE and FRE throughout the experimental studies, proving that AR is not robust to the sampling accuracy of the targets used to compute image-to-patient registration. Moreover, an influence of the size of the volume to be register was observed. Hence, it is advisable to reduce both errors due to annotations and the size of registration volumes, which can cause large errors in AR systems.

Introduction

Image-guided surgery (IGS) systems aim to provide navigation to surgeons in order to improve the accuracy and safety of the procedures. IGS utilizes computer-based systems to provide virtual image overlays and target the surgical sites [1]. In the past 30 years, with the technological advances in computer science and medical imaging, IGS has greatly expanded [1–3]. IGS combines medical images, such as magnetic resonance imaging (MRI) or computed tomography (CT), with the intra-operative images shown during the surgery to the surgeon through a laparoscope camera, endoscope camera or ultrasound (US) image. This information is displayed to surgeons either through 3D models on separate monitors or overlaid as augmented reality (AR). Instrument tracking technologies are used in IGS to provide reliable

information regarding position and orientation of surgical instruments. Moreover, tracking technologies are also used in registration tasks [4].

This study's field of application of IGS is Liver Laparoscopic Resection Surgery (LLRS). Conventionally, before LLRS, the patient undergoes volumetric scans, such as CT or MRI, to diagnose and to plan optimal treatment. These scans, known as pre-operative scans, are used by the surgeons to plan the removal of tumours from the liver [5]. However, when the patient is on the surgical table, his/her position and orientation are different from the scanning position. For this reason, surgeons use their understanding of the anatomy of the liver and intra-operative imaging modalities (such as laparoscopic video and US) to spatially correlate the anatomy of the liver to the

diagnostic CT or MRI scans. This approach may not only introduce inaccuracies, but also makes the surgery more dependent on the experience of the surgeon. Moreover, in laparoscopic surgery, pneumoperitoneum (inflation of the patient's abdomen) is always performed to have sufficient space to introduce laparoscopic instruments and the laparoscopic camera into the abdomen. This is problematic because pneumoperitoneum also deforms the shape of the organs [6], making the anatomical correlation with the CT scan even more complicated for the surgeon.

IGS can help the surgeon avoid an intra-operative unfavourable incidence by providing surgical navigation. This is becoming more important especially because parenchyma-sparing (PS) liver resection has become a standard surgical treatment for colorectal liver metastases (CRLM) [7] because it facilitates repeated liver resections, which can increase survival [8]. During these procedures, lesions need to be located intra-operatively, resection performed with planned margin, and vessels cut at optimal locations. IGS can provide surgeons with an overview of the vascular structures, position of the lesion, etc. directly in the operating field. For example, for resection in posterosuperior liver segments, which in some cases are more complicated than formal resections due to accessibility and visibility [9], IGS could be used to focus attention in the correct direction and on the structure where the lesion is located.

Within IGS, AR is a computer vision technique in which computer-generated images are superimposed onto video frames to enhance the visualization and improve the spatial understanding of the scene. In IGS, AR is achieved *via* superposition of 3D reconstructions of segmented medical volumes, such as the CT scan, of structures of interest (e.g. tumours or blood vessels as shown in Figure 1), on the laparoscope camera [10].

This study aims to understand the influence of human-induced error for AR on the laparoscope camera during LLRS. This experimental work studies AR accuracy in three experiments: an accuracy verification phantom, a patient-specific liver phantom and an *in vivo* porcine model with intra-operative CT scan.

Material and methods

Background methodology

In LLRS, the surgeon conventionally performs surgery visualizing the organs through the video of a laparoscopic camera, together with assistance of medical images and, when available, 3D reconstructed surface models through segmentation. All this information can be displayed on a separate screen within the OR (as shown in Figure 1) or can be combined to the laparoscope perspective into AR (as shown in Figure 2). As aforementioned, this study focuses on AR. The following chapter describes the main algorithms used to achieve AR in this study: hand-eye camera calibration, point-based registration (PBR) and AR re-projection.

Hand-eye camera calibration

'Hand-eye' camera calibration was coined in the robotics field [11]. In IGS for LLRS, hand-eye camera

Figure 1. Use of segmentation models and CT scans as a navigation map for laparoscopic liver resection surgery.
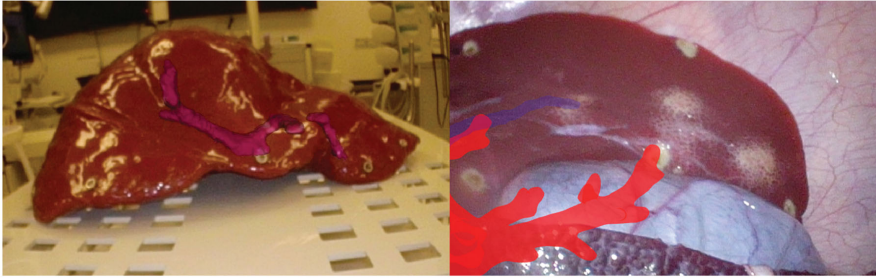
Figure 2. Frames of AR showing re-projected blood vessel structures in the phantom (left) and the *in vivo* study (right).
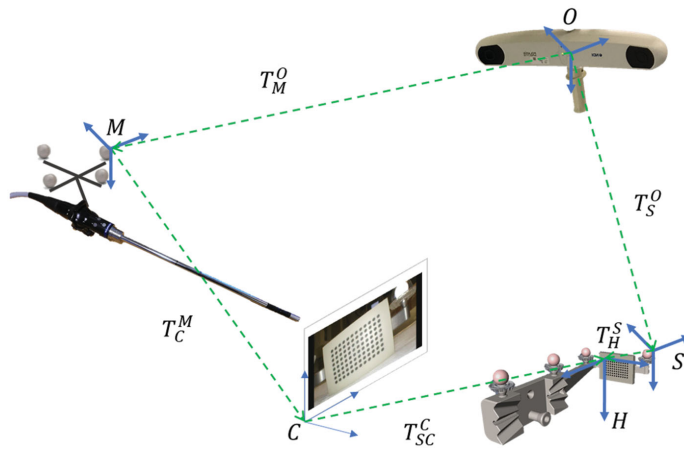

Figure 3. Hand-eye camera calibration transformation diagram, from the laparoscope camera (*C*) to the optical markers attached to it (*M*).

calibration aims to compute the transformation between a stereo laparoscope camera (the 'eye'), ENDOEYE flex3D (Olympus, Tokyo, Japan), and the instrument markers rigidly attached to the camera (the 'hand'), as shown in Figure 3. Instrument tracking was achieved in this study using the optical tracking system Polaris Spectra (NDI, Waterloo, Canada). Optical markers were rigidly attached to the laparoscope camera using an NDI Polaris Rigid Body (Part Number 8700449). The six degrees of freedom of the laparoscope camera were then tracked at a sampling rate of 60 Hz. To perform camera calibration and hand-eye camera calibration, a calibration plate with a 96-dot pattern, with four optical markers at accurately machined locations, was manufactured by Cascination (Cascination, Bern, Switzerland). In this study, the left channel of the stereo laparoscope camera was calibrated and used for AR evaluation. The calibration plate is white with laser-printed black circles,

OpenCV algorithms were used for detection of the ellipsoidal centroids and for camera calibration [12]. To compute hand-eye calibration, Equation (1) could have been used in order to compute the hand-eye calibration transform from a single pose (i.e. T_C^M, according to Figure 3)

$$T_C^M = T_O^M \cdot T_S^O \cdot T_{SC}^S \cdot T_C^{SC} \tag{1}$$

where O is the coordinate system for the optical tracking system, M for the optical markers attached to the laparoscope camera, S for the optical markers on the calibration plate, SC for the calibration plate and C is the coordinate system origin for the camera pose. The notation used in this paper indicates as superscript the coordinate system *with respect to which* the transformation is applied, and subscripted is the *towards which* coordinate system. Moreover, all transformations described in this study are 4×4 matrices in homogeneous coordinates. T_O^M and T_O^S are

provided by the optical tracking system, whereas T_{SC}^S is obtained though pose estimation of the calibration plate according to Zhang [12]. The calibration plate was manufactured so that the axes and origins of S and SC coincide, therefore, for this study, T_{SC}^S is a 4×4 identity matrix.

In order to improve both accuracy and reliability of the hand-eye camera calibration, instead of using Equation (1), the authors of this study implemented a multiple posed (N-posed) hand-eye camera calibration, based on Lee et al. [13]. In order to extend the set of equation for multiple (N) poses, we can rewrite Equation (1) as follows:

$$
\begin{bmatrix}
\begin{bmatrix}
I \otimes \left(R_O^M\right)^{-1} & Z_{9,3} \\
Z_{3,9} & \left(R_O^M\right)^{-1}
\end{bmatrix}_1 \\
\vdots \\
\begin{bmatrix}
I \otimes \left(R_O^M\right)^{-1} & Z_{9,3} \\
Z_{3,9} & \left(R_O^M\right)^{-1}
\end{bmatrix}_N
\end{bmatrix}
\times
\begin{bmatrix}
vec\left(R_C^M\right) \\
t_C^M
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
\begin{bmatrix}
vec\left(R_C^O\right) \\
t_C^O + \left(R_O^M\right)^{-1} \cdot t_O^M
\end{bmatrix}_1 \\
\vdots \\
\begin{bmatrix}
vec\left(R_C^O\right) \\
t_C^O + \left(R_O^M\right)^{-1} \cdot t_O^M
\end{bmatrix}_N
\end{bmatrix}
\tag{2}
$$

Minimization of this linear system of matrices through least squares estimation results in matrix T_C^M. More information regarding the accuracy and theory of the algorithms is available from studies by Lee et al. [13] and Lai and Shan [14].

Point-based registration

In the medical field, image registration aims to establish spatial correspondences between volumetric datasets [15]. The alignment between volumetric CT/MRI images (and segmented models) to the liver configuration when the patient is on the operating table is a field of image registration known as image-to-patient or image-to-physical registration. Two solutions are most commonly in use in the literature to perform image-to-patient registration: PBR [16] and the single landmark registration method [17]. Both algorithms work by sampling and matching a set of correspondent positions between coordinate systems. Within this study, PBR was implemented and tested to connect the image space I (the CT/MRI coordinates) and the patient's position on the surgical table P. Consequently, image-to-patient registration is represented by transform T_I^P. To achieve AR on the camera perspective, we use image-to-patient registration to transform the image coordinates I into camera coordinates C, as shown in Figure 4 following the equation:

$$
T_I^C = \left(T_C^M\right)^{-1} \cdot T_O^M \cdot T_P^O \cdot T_I^P
\tag{3}
$$

However, if the points used for registration are already with respect to the optical tracking system, transformation T_P^O will be an identity matrix. This can happen if we are registering to coordinate system O by sampling positions using an optically tracked instrument.

Re-projection of volumes in augmented reality

To complete the AR after Equation (3), re-projection of 3D volumes to 2D images was performed using an additional transformation, commonly referred to as
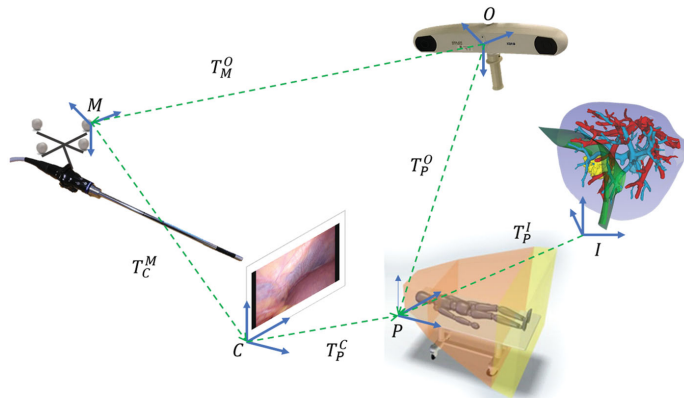

Figure 4. AR transformation diagram which combines hand-eye calibration and image-to-patient registration.

the perspective projection matrix or camera intrinsic matrix [12]. Re-projection of models in the I (Image) coordinate system as AR on the camera view can be performed through Equation (4):

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \cdot \left(T_C^M \right)^{-1} \cdot T_O^M \cdot T_P^O \cdot T_I^P \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (4)$$

where u and v are the 2D positions on the image plane in pixels, x, y, z are the 3D positions in the P coordinate system and matrix K is the intrinsic parameters matrix, computed through camera calibration. t is a nonhomogeneous transformed vector 3D positions in C coordinates (3×1 vector of the resulting 4×1 transformed points). Rectification parameters are also used during projection of the volumes.

Experimental protocol

The experiment protocol of this study aims to examine the influence of human errors (registration-related errors) on AR re-projection errors. The algorithms to generate the AR were kept consistent throughout each experiment. Three experiments were conducted with decreasing accuracy of sampling positions during

image-to-patient registration. The first experiment makes use of a precisely machined, custom built optical validation phantom which follows ASTM standard F2554-10 for optical tracking accuracy measurement, described by Teatini et al. [4]. The second experiment evaluates AR accuracy on a patient-specific liver phantom with markers visible in the CT. Finally, in order to test the AR in a fully clinical scenario, the algorithms described above were tested through a porcine experiment (*in vivo* model).

For each experiment, image-to-patient registration was performed multiple times, always using five registration landmarks and the rest of the landmarks to compute the inaccuracy of the AR. Re-projection error in AR was computed as distance between manually annotated positions (ground truth positions) and their correspondent re-projected positions on the AR frames, as done also by Teatini et al. [18]. An example of the re-projected points, for all three experiments, is visible in Figure 5. Two parameters were evaluated in this study: fiducial registration error (FRE) and target registration error (TRE). FRE represents the accuracy of re-projected markers used to compute image-to-patient registration, and TRE the accuracy for all other points across the rest of the volume [16]. Both errors are computed as re-projection errors. These TRE and FRE fully represent, in our
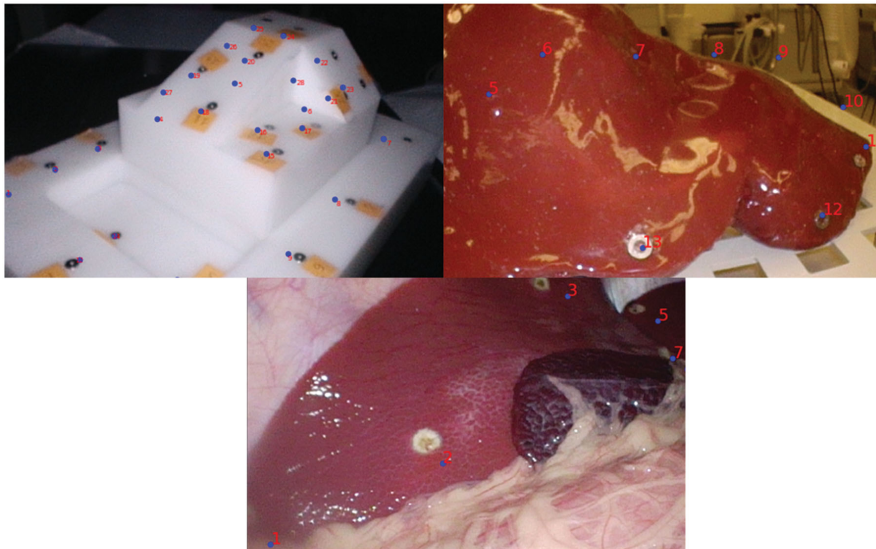

Figure 5. AR frames for each experiment procedure, validation phantom (left top), liver phantom (right top) and *in vivo* study (bottom). The registration targets were manually annotated from the frames, TRE and FRE were computed as the distances to the re-projected corresponding dots.

opinion, the accuracy of AR in terms of how well re-projected volumes are in comparison to ground truth positions. Moreover, to provide the reader with FRE and TRE of the AR in millimetres, and not only in pixels, the authors make use of the inverse of matrix K in Equation (4), as described by Thompson et al. [19].

Validation phantom

The optical validation phantom was custom produced with 28 titanium target divot pins designed for TRE calculation on various planes and orientations, more details are available in [4]. Based on the algorithms previously described, the divot pins were registered and then re-projected onto the laparoscope frame as AR (shown in Figure 5). In this scenario, the accuracy of sampling registration targets is very good because they are precisely machined targets at measured locations.

Liver phantom

The patient-specific liver phantom is a phantom designed based on the CT scans of a patient. The liver phantom used throughout the experiments was produced by the ARTORG research centre (ARTORG, Bern, Switzerland) according to Pacioni et al. [20]. Fourteen stainless steel metallic M6 washers were glued around the whole surface of the liver phantom and served as landmarks. An intra-operative SOMATON Definition Edge CT scan (SIEMENS, Munich, Germany) was used to obtain a CT scan in the OR with the liver phantom positioned on the surgical table. The washers were segmented from the CT scan through intensity-based thresholding and clustered into positions through fuzzy means classification.

In comparison to the previous experiment on the validation phantom, the liver phantom introduces more inaccuracy in determining the correct positions to sample. This is due to the deformability of the phantom as well as the fact that the surgeon has to aim to the centre of the washers with a tracked pointer, instead of using precisely machined divot holes. Moreover, the 3D spatial locations of the markers are not precisely measured (clustered from the CT scan, as aforementioned), which further increases the sampling error.

In vivo model

The pre-clinical trial was necessary to calculate the AR accuracy in a more realistic clinical scenario: the positions for PBR are points sampled directly on the liver surface. For this reason, they are neither visible in the CT scan (like the metallic washers in the liver phantom), nor precisely machined divot positions (as in the validation phantom). Hence, the correspondence between the positions in the CT scan and the laparoscopic camera perspective is based on the surgeon's anatomical understanding of the liver when required to annotate the locations sampled laparoscopically on the CT scan. An in vivo model of 59.5 kg was positioned on the surgical table in an OR equipped with the intra-operative CT scanner. After establishing the pneumoperitoneum to 13 mmHg through a Veress needle, an intra-operative CT scan was performed. Intra-operative imaging was performed rather than pre-operative imaging to minimize the inaccuracy due to soft-tissue deformation due to pneumoperitoneum. Successively, through an optically tracked monopolar cauterizer (Aesculap, Tuttlingen, Germany) 15 cauterization marks (ablation marks) were performed by the surgeon on the liver surface across the whole visible surface (similarly to [15]).

Reproducing the previous experiments, five of the targets were used to perform image-to-patient registration, whereas the other 10 were used to compute the accuracy. The cauterization marks performed on the liver surface were matched with the annotations made by the surgeon on a segmentation model of the liver parenchyma from the intra-operative CT scan.

This experiment relied on the surgeon's anatomical understanding to calculate the positions of the landmarks on the CT scan. These annotation errors are computed through fiducial localization error (FLE) [21]. FLE was evaluated by laparoscopic insertion of needles at the centre location of the ablation marks, as to provide approximate ground truth positions for the annotations (approximate because insertion of the fiducials can cause some deformation to the liver tissue; moreover, segmentation and reconstruction errors may also present).

Results

For each experiment, a total of 100 AR frames were manually annotated to evaluate TRE and FRE. Five markers were used for registration and the rest for accuracy evaluations. Table 1 summarizes the average TRE for all three experiments for each registration procedures in millimetres.

Validation phantom

A total of 748 re-projected divot positions were used to evaluate TRE and 164 for FRE. The average TRE across registrations was found to be

Table 1. Results of TRE and FRE in (mm) for each experiment, separated into registration procedure, with standard deviation.

Registration	Validation phantom TRE (mm)	Validation phantom FRE (mm)	Liver phantom TRE (mm)	Liver phantom FRE (mm)	*In vivo* study TRE (mm)	*In vivo* study FRE (mm)
1	8.57 ± 2.83	9.02 ± 1.2	7.32 ± 5.53	7.72 ± 8.03	13.07 ± 5.44	12 ± 5.37
2	8.44 ± 1.84	7.39 ± 0.96	7.51 ± 6.06	6.69 ± 6.05	11.72 ± 6.3	10.6 ± 6.24
3	4.95 ± 1.26	5.46 ± 0.93	8.02 ± 6.85	6.68 ± 5.66	11.21 ± 5.53	10.63 ± 5.86
4	6.08 ± 1.41	6.94 ± 0.77	9.15 ± 7.17	7.25 ± 5.49	12.85 ± 4.89	11.89 ± 6.62
5	8.72 ± 6.08	7.92 ± 4.73	7.57 ± 5.92	6.45 ± 5.77	14.08 ± 6.86	12.71 ± 6.1
6	5.7 ± 1.15	7.37 ± 1.87	8.25 ± 6.14	8.56 ± 5.01	12.16 ± 6.17	12.21 ± 5.87
7	5.9 ± 1.57	4.93 ± 0.54	7.12 ± 5.76	7.88 ± 6.47	14.25 ± 6.97	12.67 ± 8.07
8	8.41 ± 1.52	7.53 ± 0.86	8.09 ± 6.4	5.18 ± 3.94	14.99 ± 5.52	12.58 ± 6.11
9	4.82 ± 0.48	5.5 ± 0.24	6.99 ± 5.69	9.16 ± 7.06	13.89 ± 6.69	12.99 ± 6.83
10	7.08 ± 1.35	7.3 ± 0.85	8.44 ± 6.42	5.7 ± 3.3	15.48 ± 8.07	10.09 ± 7.33
Mean	**6.87 ± 1.95**	**6.93 ± 1.29**	**7.85 ± 6.19**	**7.13 ± 5.68**	**13.37 ± 6.25**	**11.84 ± 6.44**

Note that the registration targets are not related between experiments.

$\mu = 6.87$ mm, with standard deviation of $\sigma = 1.95$ mm. FRE resulted in $\mu = 6.93$ mm, with standard deviation of $\sigma = 1.29$ mm.

Liver phantom

A total of 1403 re-projected metallic marker centroid positions were used to evaluate TRE and 450 for FRE. The average TRE across registrations was found to be $\mu = 7.85$ mm, with standard deviation of $\sigma = 6.19$ mm. FRE resulted in $\mu = 7.13$ mm, with standard deviation of $\sigma = 5.68$ mm.

In vivo *model*

A total of 3074 re-projected cauterization points were used in order to compute TRE and 1137 for FRE. The average TRE across registrations was found to be $\mu = 13.37$ mm, with standard deviation of $\sigma = 6.25$ mm. FRE resulted in $\mu = 11.84$ mm, with standard deviation of $\sigma = 6.44$ mm. FLE, computed as the *rms* between the inserted fiducials and the cauterization points annotated was, on average, 16.40 mm.

Because of non-normality of parts of the data, six Kruskal-Wallis tests were conducted in SPSS (IBM, Armonk, NY) to compare the TRE and FRE across registration procedures for each experiment. Significant differences ($p < .05$) between registration procedure accuracies were found for the validation phantom, for both TRE and FRE, $\chi^2(9) = 290.06$, $p = 3.34E^{-57}$ and $\chi^2(9) = 78.69$, $p = 2.93E^{-13}$, respectively. No significant differences were found for the liver phantom TRE but for the FRE $\chi^2(9) = 21.28$, $p = .011$. Significant differences between registration procedures were found for TRE and FRE the *in vivo* experiment, $\chi^2(9) = 152.04$, $p = 3.33E^{-28}$ and $\chi^2(9) = 29.96$, $p = .00045$, respectively.

Successively, an additional Kruskal-Wallis test (because of non-normality of the data) showed that there was a statistically significant difference between the different experiments in terms of TRE, $\chi^2(2) = 1223.61$, $p = 1.976E^{-266}$, with a mean rank score of 1666.21 for the validation phantom, 1781.41 for the liver phantom and 3222.93 for the *in vivo* study. FRE also revealed significant differences, $\chi^2(2) = 254.72$, $p = 4.88E^{-56}$, with a mean rank score of 668.46 for the validation phantom, 595.34 for the liver phantom and 1017.01 for the *in vivo* study.

Discussion

Based on the results obtained from the three experiments, the accuracy with which a position is annotated on a CT scan volume affects both the TRE and FRE in AR. This is inferred from the comparison between the results in the validation phantom and liver phantom, where points are measured or automatically clustered, and the results on the *in vivo* model, which depends on human interaction through annotation.

It is noteworthy to mention that, although the average TREs and FREs of the validation and the liver phantoms are very similar, TRE varies significantly between registration procedure in the validation phantom but not for the liver phantom. This can be explained by the differences in sizes between the volumes to be registered. The validation phantom presents a volume of 4320 cm^3, whereas the patient-specific phantom is 1882 cm^3 (the *in vivo* liver was 2393 cm^3). Moreover, the validation phantom targets are partially symmetrical, and some positions are almost collinear (as can be seen in Figure 5). Sampling of five positions across the volume was performed randomly in each of the experiments. Therefore, depending on the positions of the targets, a larger volume will be more affected than a smaller volume. Based on these results, for registration of large volumes, it is preferable to use a larger number

of targets to compute the image-to-patient registration and possibly better spatial disposition of these targets with respect to the volume (such as ensuring non-collinearity between registration targets or closeness to areas of interest such as tumours).

The statistical differences between TRE and FRE in the *in vivo* study may depend on the FLE for the cauterization markers used to compute image-to-patient registration. If, e.g. the accuracy of annotating a position were to be inexact, this could greatly affect the image-to-patient registration matrix, causing an increase in both TRE and FRE (meaning, a decrease in AR accuracy). In order to mitigate the effect of sampling error in PBR, the authors propose using intra-operative fiducials on the liver surface, which could be removed post-surgery (or made of biocompatible/biodegradable material). These fiducials could be detected in the intra-operative CT/MRI scan and used to perform PBR. This would probably reduce the sampling error to the accuracy evaluated in the liver phantom testing (it would greatly reduce FLE) and may greatly improve the accuracy of the AR reprojection. However, this may prolong and complicate the surgical procedure.

Within this study, targets located on the centre of the liver surface were the most complicated to be annotated correctly and should therefore be avoided. Alternatively, using targets on the edges of the liver could also reduce FLE (if the parenchyma were rigid enough during sampling). Positions that are also stable in the liver include intersections of the parenchyma with major blood vessels (such as the portal vein) or bifurcations of blood vessel structures. These structures can serve as solid registration targets; however, they can only be sampled from within the liver after resection.

Some limitations to this study include the fact that the scan used to perform AR was intra-operative, which is currently not within the surgical workflow for most hospitals. Pre-operative imaging does not account for intra-operative deformations, such as pneumoperitoneum (static deformations); hence, inaccuracies due to the non-rigidity of the liver would be present in the current surgical workflow, as mentioned by Thompson et al. [19,22]. However, intra-operative CT and MRI scanners are under production for Hybrid OR suites, and show very large increases in AR accuracy and alternative methods of acquiring intra-operative data (such as stereo surface reconstruction [23] or laparoscopic US [24]) could be used. Alternatives to intra-operative imaging, such as the use of biomechanical modelling are a valid solution,

as described in [25–28], but may be affected by the need of estimating of viscoelastic properties of soft tissue and boundary conditions. If elastography were to be used to characterize the viscoelasticity of the tissue, and the boundary conditions were known for each patient, biomechanical modelling could be used to account for deformations.

It would be interesting to study the effect of spatial disposition of registration targets across volumes of variable size, to validate the assumption that the volume changes caused the differences in accuracy between the validation phantom and the other experiments. Moreover, another limitation to the study is that the cauterization marks were performed on the liver surface and might not represent fully the error in the depth axis. However, the marks were performed across the liver as deep as possible towards the diaphragm, which allowed us to calculate TRE and FRE for positions at various depths, though not as much as a blood vessel or a tumour deep within the tissue. Another limitation is the fact that both camera calibrations and hand-eye calibrations are performed as they would be in a surgical environment, without thorough refinement of the calibration procedures. Furthermore, the annotations for PBR could have been performed by multiple surgeons to further validate the hypothesis that annotations (FLE) could cause significant differences in terms of registration accuracy for AR.

The investigated inaccuracy in terms of TRE and FRE is larger than that accepted by surgeons. However, the use of AR, complemented with intra-operative US, could still be useful for visualization of the structures in the resection field. Thus, it might help surgeons better understand spatial distribution of anatomical structures and lead to safer surgery. Even with current quality, where resection lines cannot be followed blindly due to the system's inaccuracy, there is still clinical value in the use of this AR system, especially for spatial understanding.

Conclusions

This study aims to show that accuracy in sampling registration targets can contribute to decreases/increases in the accuracy of the AR through PBR. The laparoscope camera, CT scanner, optical markers and algorithms used were consistent throughout all experiments, the only difference was the volumes to be registered. Results show that the accuracy through PBR can change based on the accuracy in sampling the positions to compute image-to-patient registration, possibly also the size of the volume to be

registered and spatial disposition of registration targets used to perform image-to-patient registration.

The overall accuracy for the AR in terms of TRE for the *in vivo* model was around 13 mm, and 11 mm in terms of FRE. However, the results also say that the TRE AR accuracy can worsen greatly based on the registration procedure, if the targets are based on the surgeon's annotations (TRE can result to be larger than a centimeter, as shown in Table 1 and validated in other studies [22,18]).

The main indication for liver resection is CRLM in the western world. If the proposed solution were to be used for clinical use in CRLM, the error for the AR should not exceed 6 mm of inaccuracy according to the authors. The reason is that 6 mm are acceptable for surgeons because the safety margin for CRLM (resection margin) is 1–3 mm [29], and normally, surgery is planned with a 1 cm margin. Therefore, a planned resection line with 1 cm of margin using an AR with 6 mm of error, will allow 3 mm of space in addition to the 1 mm safety margin.

Based on the results of this study, it is necessary to improve the image-to-patient registration, possibly with the use of user-independent fiducials for registration and a smaller volume to be registered. Overall, improvements of this AR system are necessary; however, we have proven that better sampling accuracy can lead to much better accuracies, which will allow AR to be of use for surgery.

Acknowledgements

Ethical approvals

All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted. This article does not contain any studies with human participants performed by any of the authors.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Funding

ORCID

Andrea Teatini 🔟 http://orcid.org/0000-0003-4764-5685
Davit L. Aghayan 🔟 http://orcid.org/0000-0001-7051-3512
Rahul Prasanna Kumar 🔟 http://orcid.org/0000-0002-5951-7317
Rafael Palomar 🔟 http://orcid.org/0000-0002-9136-4154
Bjørn Edwin 🔟 http://orcid.org/0000-0002-3137-6225
Ole Jakob Elle 🔟 http://orcid.org/0000-0003-2359-1272

References

[1] Cleary K, Peters TM. Image-guided interventions: technology review and clinical applications. Annu Rev Biomed Eng. 2010;12(1):119–142.

[2] Hallet J, Soler L, Diana M, et al. Trans-thoracic minimally invasive liver resection guided by augmented reality. J Am Coll Surg. 2015;220(5):e55–e60.

[3] Bernhardt S, Nicolau SA, Soler L, et al. The status of augmented reality in laparoscopic surgery as of 2016. Med Image Anal. 2017;37:66–90.

[4] Teatini A, Frutos JD, Langø T, et al. Assessment and comparison of target registration accuracy in surgical instrument tracking technologies. Conf Proc IEEE Eng Med Biol Soc. 2018;2018:1845–1848.

[5] Palomar R, Cheikh FA, Edwin B, et al. Surface reconstruction for planning and navigation of liver resections. Comput Med Imaging Graph. 2016;53:30–42.

[6] Heiselman JS, Clements LW, Collins JA, et al. Characterization and correction of intraoperative soft tissue deformation in image-guided laparoscopic liver surgery. J Med Imag. 2018;5(2):021203.

[7] Aghayan DL, Pelanis E, Fretland ÅA, et al. Laparoscopic parenchyma-sparing liver resection for colorectal metastases. Radio Oncol. 2017;52(1):36–41.

[8] Fretland AA, Dagenborg VJ, Bjørnelv GMW, et al. Laparoscopic versus open resection for colorectal liver metastases. Ann Surg. 2018;267(2):199–207.

[9] Aghayan DL, Fretland ÅA, Kazaryan AM, et al. Laparoscopic versus open liver resection in the posterosuperior segments: a sub-group analysis from the OSLO-COMET randomized controlled trial. HPB (Oxford). 2019;21(11):1485–1490.

[10] Mountney P, Fallert J, Nicolau S, et al. An augmented reality framework for soft tissue surgery. Lecture Notes Comput Sci. 2014;17:423–431.

[11] Thompson S, Stoyanov D, Schneider C, et al. Hand-eye calibration for rigid laparoscopes using an invariant point. Int J Comput Assist Radiol Surg. 2016;11(6):1071–1080.

[12] Zhang Z. A flexible new technique for camera calibration. IEEE Trans Pattern Anal Machine Intell. 2000;22(11):1330–1334.

[13] Lee S, Lee H, Choi H, et al. Effective calibration of an endoscope to an optical tracking system for medical augmented reality. Cogent Eng. 2017;4:1–11.

[14] Lai M, Shan C. Hand-eye camera calibration with an optical tracking system. Proceedings of the 12th International Conference on Distributed Smart Cameras. New York (NY): ACM; 2018. p. 18.

[15] Modat M, Ridgway GR, Taylor ZA, et al. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed. 2010;98(3):278–284.

[16] Fitzpatrick JM. Fiducial registration error and target registration error are uncorrelated. 2009;7261:726102. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.813601.

[17] Pérez de Frutos J, Hofstad EF, Solberg OV, et al. Laboratory test of Single Landmark registration method for ultrasound-based navigation in laparoscopy using an open-source platform. Int J Comput Assist Radiol Surg. 2018;13(12):1927–1936.

[18] Thompson S, Totz J, Song Y, et al. Accuracy validation of an image guided laparoscopy system for liver resection. Proc SPIE. 2015;9415:941509.

[19] Thompson S, Schneider C, Bosi M, et al. In vivo estimation of target registration errors during augmented reality laparoscopic surgery. Int J Comput Assist Radiol Surg. 2018;13(6):865–874.

[20] Pacioni A, Carbone M, Freschi C, et al. Patient-specific ultrasound liver phantom: materials and fabrication method. Int J Comput Assist Radial Surg. 2015;10(7):1065–1075.

[21] Liu W, Ding H, Han H, et al. The study of fiducial localization error of image in point-based registration. Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society Engineering the Future of Biomedice; EMBC; 2009. p. 5088–5091.

[22] Teatini A, Pelanis E, Aghayan DL, et al. The effect of intraoperative imaging on surgical navigation for laparoscopic liver resection surgery. Sci Rep. 2019;9(1):18687 doi:10.1038/s41598-019-54915-3. PMC:31822701

[23] Andrea T, Congcong W, Rafael P, et al. Validation of stereo vision based liver surface reconstruction for image guided surgery. Colour Vis Comput Symp. 2018;2018:1–6.

[24] Fusaglia M, Tinguely P, Banz V, et al. A novel ultrasound-based registration for image-guided laparoscopic liver ablation. Surg Innov. 2016;23(4):397–406.

[25] Faure F, Duriez C, Delingette H, et al. SOFA: a multi-model framework for interactive physical simulation. 2012.

[26] Nikolaev S, Peterlik I, Cotin S, et al. Stochastic correction of boundary conditions during liver surgery to cite this version: HAL Id: hal-01823810 Stochastic Correction of Boundary Conditions during Liver Surgery. 2018.

[27] Özgür E, Koo B, Le Roy B, et al. Preoperative liver registration for augmented monocular laparoscopy using backward–forward biomechanical simulation. Int J Comput Assist Radiol Surg. 2018;13(10):1629–1640.

[28] Peterlik I, Courtecuisse H, Rohling R, et al. Fast elastic registration of soft tissues under large deformations to cite this version: HAL Id: hal-01613757. 2018.

[29] Postriganova N, Kazaryan AM, Røsok BI, et al. Margin status after laparoscopic resection of colorectal liver metastases: does a narrow resection margin have an influence on survival and local recurrence? HPB. 2014;16(9):822–829.

PAPER IV

Learning deep abdominal CT registration through adaptive loss weighting and synthetic data generation

J. Pérez de Frutos, A. Pedersen, E. Pelanis, D. Bouget S. Survarachakan, T. Langø, O. J. Elle, and F. Lindseth.
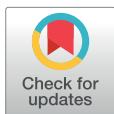
Learning deep abdominal CT registration through adaptive loss weighting and synthetic data generation

Javier Pérez de Frutos [1]*, **André Pedersen**[1,2,3], **Egidijus Pelanis**[4], **David Bouget**[1], **Shanmugapriya Survarachakan**[5], **Thomas Langø**[1,6], **Ole-Jakob Elle**[4], **Frank Lindseth**[5]

1 Department of Health Research, SINTEF, Trondheim, Norway, **2** Department of Clinical and Molecular Medicine, Norwegian University of Technology (NTNU), Trondheim, Norway, **3** Clinic of Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway, **4** The Intervention Centre, Oslo University Hospital, Oslo, Norway, **5** Department of Computer Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, **6** Research Department, Future Operating Room, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway

* javier.perezdefrutos@sintef.no

Abstract

Purpose

This study aims to explore training strategies to improve convolutional neural network-based image-to-image deformable registration for abdominal imaging.

Methods

Different training strategies, loss functions, and transfer learning schemes were considered. Furthermore, an augmentation layer which generates artificial training image pairs on-the-fly was proposed, in addition to a loss layer that enables dynamic loss weighting.

Results

Guiding registration using segmentations in the training step proved beneficial for deep-learning-based image registration. Finetuning the pretrained model from the brain MRI dataset to the abdominal CT dataset further improved performance on the latter application, removing the need for a large dataset to yield satisfactory performance. Dynamic loss weighting also marginally improved performance, all without impacting inference runtime.

Conclusion

Using simple concepts, we improved the performance of a commonly used deep image registration architecture, VoxelMorph. In future work, our framework, DDMR, should be validated on different datasets to further assess its value.

Introduction

For liver surgery, minimally invasive techniques such as laparoscopy have become as relevant as open surgery [1]. Among other benefits, laparoscopy has shown to yield higher quality of life, shorten recovery time, lessen patient trauma, and reduce blood loss with comparable long-term oncological outcomes [1]. Overcoming challenges from limited field of view to manoeuvrability, and a small work space are the foundations of laparoscopy success. Image-guided navigation platforms aim to ease the burden off the surgeon, by bringing better visualisation techniques to the operating room [2, 3]. Image-to-patient and image-to-image registration techniques (hereafter image registration) are at the core of such platforms to provide clinically valuable visualisation tools. The concept of image registration refers to the alignment of at least two images, matching the location of corresponding features across images in order to express them into a common space. Both rigid and non-rigid registration are the two main strategies to define the alignment between the images. Rigid registration uses affine transformations, which are quicker to compute but less accurate as these are applied globally. Non-rigid registration, also known as deformable registration, defines a diffeomorphism, i.e., a point-to-point correspondence, between the images. However, non-rigid registration comes at the expense of higher computational needs and thus hardware constraints might hinder the development and deployment of such algorithms. In medicine, image registration is mandatory for fusing clinically relevant information across images; groundwork for enabling image-guided navigation during laparoscopic interventions [4, 5]. Additionally, laparoscopic preoperative surgical planning benefits from abdominal computed tomography (CT) to magnetic resonance imaging (MRI) registration to better identify risk areas in a patient's anatomy [6].

During laparoscopic liver surgeries, intraoperative imaging (e.g., video and ultrasound) is routinely used to assist the surgeon in navigating the liver while identifying the location of landmarks. In parenchyma-sparing liver resection (i.e., wedge resection) for colorectal liver metastasis, a minimal safety margin around the lesions is defined to ensure no recurrence and spare healthy tissue [7]. When dealing with narrow margins and close proximity to critical structures, a high accuracy in the registration method employed is paramount to ensure the best patient outcome. Patient-specific cross-modality registration between images of different nature (e.g., CT to MRI) is practised [8], yet being a more complex process compared to mono-modal registration.

The alignment of images can be evaluated through different metrics based either on intensity information from the voxels, shape information from segmentation masks, or spatial information from landmarks' location or relative distances. The most common intensity-based similarity metrics are the normalised cross-correlation (NCC), structural similarity index measure (SSIM), or related variations [9, 10]. For segmentation-based metrics, the most notorious are the Dice similarity coefficient (DSC) and Hausdorff distance (HD) [11]. However, target registration error (TRE) is the gold standard metric for practitioners, conferring a quantitative error measure based on the target lesion location across images [12].

Research on the use of convolutional neural networks (CNNs) for image registration has gained momentum in recent years, motivated by the improvements in hardware and software. One early application of deep learning-based image registration (hereafter deep image registration) was performed by Wu *et al.* [13]. They proposed a network built with two convolutional layers, coupled with principal component analysis as a dimensionality reduction step, to align brain MR scans. Expanding upon the concept, Jaderberg *et al.* [14] introduced the spatial transformer network, including a sampling step for data interpolation, allowing for gradients to be backpropagated. Hence, further enabling neural network deformable image-to-image registration applications. Publications on CNNs for image-registration show a preference for

encoder-decoder architectures like U-Net [15], followed by a spatial transformer network, as can be seen in Quicksilver [16], VoxelMorph [9], and other studies [17]. Mok *et al.* [18] proposed a Laplacian pyramid network for multi-resolution-based MRI registration, enforcing the non-rigid transformation to be diffeomorphic.

The development of weakly-supervised training strategies [19, 20] enabled model training by combining intensity information with other data types (e.g., segmentation masks). Intensity-based unsupervised training for non-rigid registration was explored for abdominal and lung CT [21, 22]. Building cross-modality image registration models through reinforcement learning has also been explored [23]. However, semi-supervised training of convolutional encoder-decoder architectures has been favoured for training registration models and producing the displacement map [24].

In our study, the focus is brought towards improving the training scheme of deep neural networks for deformable image registration to cater more easily to use-cases with limited data. We narrowed the scope to mono-modal registration, and the investigation of transfer learning across image modalities and anatomies. Our proposed main contributions are:

- an augmentation layer for on-the-fly data augmentation (compatible with TensorFlow GPU computational graphs), which includes generation of ground truth samples for non-rigid image registration, based on thin plate splines (TPS), removing the need for pre-computation and storage of augmented copies on disk,

- an uncertainty weighting loss layer to enable adaptive multi-task learning in a weakly-supervised learning approach,

- and the validation of a cross-anatomy and cross-modality transfer learning approach for image registration with scarce data.

Materials and methods

Dataset

In this study, two datasets were selected for conducting the experiments: the Information eXtraction from Images (IXI) dataset and Laparoscopic Versus Open Resection for Colorectal Liver Metastases: The Oslo-CoMet Randomized Controlled Trial dataset [1, 25].

The IXI dataset contains 578 T1-weighted head MR scans from healthy subjects collected from three different hospitals in London. This dataset is made available under the Creative Commons CC BY-SA 3.0 license. Only T1-weighted MRIs were used in this study, but other MRI sequences such as T2 and proton density are also available. Using the advanced normalization tools (ANTs) [26], the T1 images were registered to the symmetric Montreal Neurological Institute ICBM2009a atlas, to subsequently obtain the segmentation masks of 29 different regions of the brain. Ultimately, left and right parcels were merged together resulting in a collection of 17 labels (see the online resource Table A in S1 Appendix). The data was then stratified into three cohorts: training (n = 407), validation (n = 87), and test (n = 88) sets.

The Oslo-CoMet trial dataset, compiled by the Intervention Centre, Oslo University Hospital (Norway), contains 60 contrast-enhanced CTs. The trial protocol for this study was approved by the Regional Ethical Committee of South Eastern Norway (REK Sør-Øst B 2011/1285) and the Data Protection Officer of Oslo University Hospital (Clinicaltrials.org identifier NCT01516710). Informed written consent was obtained from all participants included in the study. Manual delineations of the liver parenchyma, i.e., liver segmentation masks, were available as part of the Oslo-CoMet dataset [4]. Additionally, an approximate segmentation of the vascular structures was obtained using the segmentation model available in the public

livermask tool [27]. The data was then stratified into three cohorts: training (n = 41), validation (n = 8), and test (n = 11) sets.

Preprocessing

Before the training phase, both CT and MR images, as well as the segmentation masks, were resampled to an isotropic resolution of 1 mm^3 and resized to $128 \times 128 \times 128$ voxels. Additionally, the CT images were cropped around the liver mask before the resampling step. Cubic spline interpolation was used for resampling the intensity images, whereas segmentations were interpolated using nearest neighbour. The segmentation masks were stored as categorical 8-bit unsigned integer single-channel images, to enable rapid batch generation during training.

To overcome the scarcity of image registration datasets for algorithm development, we propose an augmentation layer, implemented in TensorFlow [28], to generate artificial moving images during training. The augmentation layer allows for data augmentation and preprocessing. The layer features gamma (0.5 to 2) and brightness augmentation (±20%), rotation, and rigid and non-rigid transformations, to generate the moving images. Data preprocessing includes resizing and intensity normalisation to the range [0, 1]. The maximum displacements, rigid and non-rigid, can be constrained to mimic real-case scenarios. In our case, 30 mm and 6 mm respectively. Rotation was limited to 10°, for any of the three coordinate axes.

The non-rigid deformation was achieved using TPS applied on an $8 \times 8 \times 8$ grid, with a configurable maximum displacement. Rigid transformations include rotation and translation.

Model architecture

The baseline architecture consists of a modified VoxelMorph model [9], based on a U-Net [29] variant. The model was used to predict the displacement map, as depicted in Fig 1. After the augmentation step, the fixed (I_f) and the generated moving (I_m) images were concatenated into a two-channel volumetric image and fed to the VoxelMorph model. The model returns the displacement map (Φ) i.e., a volume image with three channels, which describes the relative displacement of each voxel along each of the three coordinate axes. Finally, the predicted fixed image (I_p) is reconstructed by interpolating voxels on the moving image at the locations defined by the displacement map. This way, the model can be trained by comparing the predicted image with the original fixed image.

When provided, the segmentations (S_m) are likewise updated using the same displacement map. The symmetric U-Net architecture was designed with six contraction blocks featuring 32, 64, 128, 256, 512, and 1024 convolution filters respectively. Each contracting block consisted of a convolution with kernel size $3 \times 3 \times 3$ and a LeakyReLU activation function, followed by max pooling with stride 2. The decoder blocks consisted of a convolution and a LeakyReLU activation function, followed by a nearest neighbour interpolation upsampling layer. The output convolutional layer, named Head in Fig 1, was set to two consecutive convolutions of 16 filters with LeakyReLU activation function. A convolution layer with three filters was used as the output layer. This produces a displacement map with the same size as the input images and three channels, one for each displacement dimension.

Model training

The registration model was trained in a weakly-supervised manner, as proposed by Hu *et al.* [19]. Instead of evaluating the displacement map directly as in traditional supervised training, only the final registration results were assessed during training.

Due to the complexity of the task at hand, a single loss function would provide limited insight of the registration result, therefore a combination of well-known loss functions was
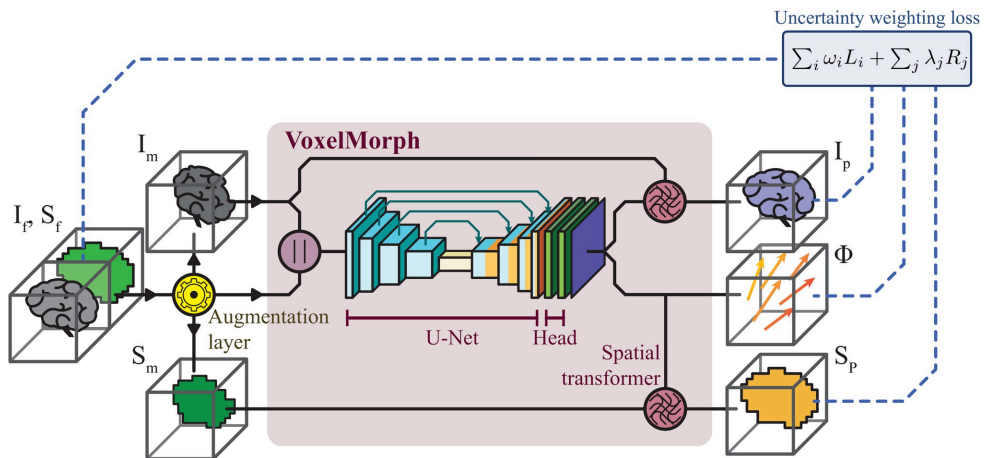
Fig 1. Proposed deep image registration pipeline. Generation of artificial moving images on-the-fly, prediction of the displacement map using a modified U-Net, and finding the optimal loss weighting automatically using uncertainty weighting.

https://doi.org/10.1371/journal.pone.0282110.g001

deemed necessary. Balancing the contribution of these operators can be challenging, time consuming, and prone to errors. We therefore used uncertainty weighting (UW) [30], which combines losses as a weighted sum and enables the loss weights to be tuned dynamically during backpropagation. Our loss function $\mathcal{L}$ was implemented as a custom layer, and consists of a weighted sum of N loss functions L and M regularisers $\mathcal{R}$:

$$\mathcal{L}(\mathbf{y}_t, \mathbf{y}_p) = \sum_{i=1}^{N} \omega_i L_i(\mathbf{y}_t, \mathbf{y}_p) + \sum_{j=1}^{M} \lambda_j \mathcal{R}_j \qquad (1)$$

such that $\sum \omega_i = \sum \lambda_i = 1$. By default, the weights ω_i and λ_i were initialised to equally contribute in the weighted sum, but can be set manually from a priori knowledge on initial loss and regularisation values. In our experiments, the default initialisation for the loss weights was used, except for the regularisation term, which was initialised to 5×10^{-3}.

For training the neural networks we used the Adam optimiser. Gradient accumulation was performed to overcome memory constraints and enable larger batch training. The batch size was set to one, but artificially increased by accumulating eight mini-batch gradients. The learning rate was set to 10^{-3} with a scheduler to decrease by 10 whenever the validation loss plateaued with a patience of 10. The models were trained using a custom training pipeline. Training curves can be found in Figs A-D in S3 Appendix. The training was limited to 10^5 epochs, and manually stopped if the model stopped converging. The model with the lowest validation loss was saved.

Experiments

The experiments were conducted on an Ubuntu 18.04 Linux desktop computer with an Intel®Xeon®Silver 4110 CPU with 16 cores, 64 GB of RAM, an NVIDIA Quadro P5000 (16 GB VRAM) dedicated GPU, and SSD hard-drive. Our framework, DDMR, used to conduct the experiments was implemented in Python 3.6 using TensorFlow v1.14. To accelerate the

Table 1. Configurations trained on both the IXI and Oslo-CoMet datasets.

Design	Model	Loss function
BL-N	Baseline	NCC
BL-NS	Baseline	NCC, SSIM
SG-ND	Segmentation-guided	NCC, DSC
SG-NSD	Segmentation-guided	NCC, SSIM, DSC
UW-NSD	Uncertainty weighting	NCC, SSIM, DSC
UW-NSDH	Uncertainty weighting	NCC, SSIM, DSC, HD

BL: baseline, SG: segmentation-guided, UW: uncertainty weighting, N: normalized cross correlation, S: structural similarity index measure, D: Dice similarity coefficient, H: Hausdorff distance.

research within the field, the source code is made openly available on GitHub (https://github.com/jpdefrutos/DDMR).

As aforesaid, our aim was to improve the training phase of image registration for CNN models. To that extent, four experiments were carried out:

(i) Ablation study Different training strategies and loss function combinations were evaluated to identify the key components in deep image registration. Three different training strategies were considered, all using weakly-supervised learning: 1) the baseline (BL) using only intensity information, 2) adding segmentation guidance (SG) to the baseline, and 3) adding uncertainty weighting (UW) to the segmentation-guided approach. For all experiments, the input size and CNN backbone were tuned prior and kept fixed. All designs are described in Table 1 and evaluated on both the IXI and Oslo-CoMet datasets. Six loss weighting schemes were tested, using different combinations of loss functions, including both intensity and segmentation-based loss functions. For the second experiment, the entire model was finetuned directly or in two steps, i.e., by first finetuning the decoder, keeping the encoder frozen, and then finetuning the full model. A learning rate of 10^{-4} was used when performing transfer learning.

(ii) Transfer learning To assess the benefit finetuning for deep image registration to applications with a small number of samples available, e.g., abdominal CT registration.

(iii) Baseline comparison The trained models were evaluated against a traditional registration framework (ANTs), to better understand the potential of deep image registration. This experiment was performed only on the Oslo-CoMet dataset, as ANTs was used to generate the segmentations on the IXI dataset. Two different configuration were tested: symmetric normalisation (SyN), with mutual information as optimisation metric, and SyN with cross-correlation as metric (SyNCC).

(iv) Training runtime The last experiment was conducted to assess the impact of the augmentation layer (see Fig A in S2 Appendix). The GPU resources were monitored during training. Only the second epoch was considered, as the first one served as warm-up.

Evaluation metrics

The evaluation was done on the test sets of the IXI and Oslo-CoMet datasets, for which the fixed-moving image pairs were generated in advance, such that the same image pairs were used across methods during evaluation. After inference, the displacement maps were resampled back to isotropic resolution using piecewise 3D linear interpolation. The final predictions

were then evaluated using four sets of metrics to cover all angles. Image similarity was assessed under computation of NCC and SSIM metrics. Segmentations were converted into one-hot encoding and evaluated using DSC, HD, and HD95 (95th percentile of HD) measured in milli- metres. The background class was excluded in the segmentation metrics computation. For image registration, TRE was estimated using the centroids of the segmentation masks of the fixed image and the predicted image, also measured in millimetres. In addition, the methods were compared in terms of inference runtime, only measuring the prediction and application of the displacement map, as all other operations were the same between the methods.

Five sets of statistical tests were conducted to further assess: 1) performance contrasts between designs, 2) benefit of transfer learning, 3) benefit of segmentation-guiding, 4) benefit of uncertainty weighting, and 5) performance contrasts between the baseline and segmenta- tion-guided models, and the traditional methods in ANTs (SyN and SyNCC). For the tests, the TRE metric was used, as it is considered the gold standard for surgical practitioners. Test 1) was conducted on the evaluations of the IXI test set, whereas tests 2) to 5) were performed using the results on the Oslo-CoMet test dataset only. Furthermore, for the tests only involving the Oslo-CoMet dataset, the two-step transfer learning approach was used as reference, as these models showed the best results.

For the statistical test 1), multiple pairwise Tukey's range tests were conducted on the IXI experiment comparing all the designs described in Table 1. For 2), benefit of transfer learning was assessed using a one-sided, non-parametric test (Mann-Whitney U test), com- paring the differences between the BL-N, SG-NSD, and UW-NSD designs on the Oslo- CoMet experience. The three generated p-values were corrected for multiple comparison using the Benjamini-Hochberg method. For 3)-5), Mann-Whitney U tests were conducted comparing BL-N and SG-NSD to assess benefit of segmentation-guiding, SG-NSD and UW-NSD to assess benefit for uncertainty weighting, and BL-N and SG-NSD against SyN and SyNCC to assess the difference in performance between deep image registration and traditional image registration solutions, respectively. The two p-values were corrected using the Benjamini-Hochberg method. The results for all five sets of tests can be found in S4 Appendix.

The Python libraries statsmodels (v0.12.2) [31] and SciPy (v1.5.4) [32] were used for the sta- tistic computations. A significance level of 0.05 was used to determine statistical significance.

Results

In Tables 2 to 5, the best performing methods in terms of individual performance metrics, i.e., most optimal mean and lowest standard deviation, were highlighted in bold. See the online resources for additional tables and figures not presented in this manuscript.

On the IXI dataset, fusing NCC and SSIM improved performance in terms of intensity- based metrics for the baseline model, whereas segmentation metrics were degraded (see Table 2). Adding segmentation-guiding drastically increased performance across all metrics compared to the baseline. Minor improvement was observed using uncertainty weighting, whereas adding the Hausdorff loss was not beneficial. In terms of TRE, multiple pairwise Tukey's range tests confirmed the benefit of segmentation-guiding ($p < 0.001$), however, no significant improvement was observed in introducing uncertainty-weighing ($p = 0.9$). The complete pairwise comparison can be found in Table A in S1 Appendix.

On the Oslo-CoMet dataset, a similar trend as for the IXI dataset was observed (see Table 3). However, in this case, the baseline model was more competitive, especially in terms of intensity-based metrics. Nonetheless, segmentation-guiding was still better overall ($p < 0.001$ in terms of TRE), as well as uncertainty weighting ($p = 0.0093$ in terms of TRE).

Table 2. Evaluation of the models trained on the IXI dataset.

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime
BL-N	0.23±0.16	0.52±0.12	0.03±0.01	109.71±26.19	100.26±27.91	29.47±8.46	0.83±0.77
BL-NS	0.25±0.16	**0.53±0.12**	0.02±0.01	145.36±22.41	138.19±23.48	30.06±9.07	0.73±0.56
SG-ND	0.45±0.24	0.46±0.10	0.61±0.08	4.64±1.37	2.15±0.54	1.08±0.39	0.82±0.58
SG-NSD	0.46±0.24	0.46±0.11	0.61±0.07	4.54±1.42	2.10±0.49	1.07±0.37	0.74±0.64
UW-NSD	**0.47±0.24**	0.46±0.11	**0.63±0.08**	**4.44±1.40**	**2.03±0.51**	**0.97±0.36**	**0.72±0.59**
UW-NSDH	**0.47±0.24**	0.46±0.11	0.61±0.07	4.63±1.49	2.14±0.52	1.06±0.36	0.75±0.59
Unregistered	0.45±0.21	0.24±0.07	0.07±0.06	21.77±5.15	18.73±4.88	11.53±3.01	-

The best performing methods for each metric are highlighted in bold.

Finetuning the entire model trained on the IXI dataset to the Oslo-CoMet dataset (see Table 4 transfer nonfrozen), yielded similar intensity-based metrics overall, but drastically improved the segmentation-guided and uncertainty weighted models in terms of segmentation metrics. The best performing models overall used uncertainty weighting. When finetuning the model in two steps, the uncertainty weighted designs were further improved to some extent (see Table 5 frozen encoder). The statistical analysis 2) shows significance improvement of the TRE when performing transfer learning and finetuning in two steps, for the UW-NSD model ($p = 0.0014$) and SG-NSD ($p = 0.0021$) models. No statistical significance was observed for the BL-N ($p = 0.8608$).

The traditional methods, SyN and SyNCC, performed well on the Oslo-CoMet test set. However, the segmentation masks were distorted, in particular the vascular segmentations mask (see Fig C in S5 Appendix). Both methods performed similarly, but the SyNCC was considerably slower. Segmentation guidance was deemed critical in obtaining better performance in terms of TRE ($p < 0.001$) compared to SyN and SyNCC. Yet no significant difference was observed on the baseline models ($p = 0.5845$). All deep learning models had similar inference runtimes of less than one second, which was expected as the final inference model architectures were identical. On average, the CNN-based methods were $\sim 13\times$ and $\sim 421\times$ faster than SyN and SyNCC, respectively. The deep learning models struggled with image reconstruction, unlike ANTs (see Fig C in S5 Appendix). For instance, anatomical structures outside the segmentation masks were poorly reconstructed in the predicted image, e.g., the spine of the patient.

Table 3. Evaluation of the models trained on the Oslo-CoMet dataset.

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime
BL-N	0.52±0.10	**0.20±0.07**	0.23±0.09	54.10±7.22	30.36±3.58	18.11±7.62	0.78±1.50
BL-NS	**0.62±0.13**	0.17±0.07	0.29±0.07	37.69±8.04	22.06±5.17	13.95±4.78	**0.76±1.44**
SG-ND	0.55±0.15	0.16±0.06	0.38±0.14	**22.03±8.27**	**12.74±6.12**	**7.60±3.96**	0.76±1.46
SG-NSD	0.58±0.13	0.12±0.07	**0.35±0.07**	25.22±7.92	14.49±4.22	8.91±3.08	0.77±1.49
UW-NSD	0.54±0.13	0.11±0.06	0.26±0.07	25.08±6.67	18.47±5.34	11.52±3.32	0.77±1.49
UW-NSDH	0.59±0.14	0.14±0.06	0.35±0.11	24.49±8.67	14.57±5.93	8.34±4.31	0.78±1.50
Unregistered	0.60±0.13	0.09±0.05	0.24±0.08	24.60±5.56	19.06±4.89	11.86±2.75	-

The best performing methods for each metric are highlighted in bold.

Table 4. Evaluation of models trained on the Oslo-CoMet dataset from finetuning the entire architecture.

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime
BL-N	0.52±0.08	**0.17±0.07**	0.23±0.07	57.98±5.36	33.00±5.14	24.09±5.92	0.77±1.45
BL-NS	**0.61±0.09**	0.16±0.07	0.14±0.03	82.91±6.96	59.94±6.41	34.41±13.03	0.77±1.46
SG-ND	0.56±0.13	0.14±0.07	0.43±0.09	15.81±5.56	9.05±3.18	5.89±3.10	0.79±1.56
SG-NSD	0.58±0.13	0.14±0.07	0.42±0.10	16.26±6.37	9.50±3.51	5.84±3.01	0.76±1.48
UW-NSD	0.58±0.12	0.14±0.06	**0.48±0.11**	15.53±5.80	**7.84±3.17**	4.05±2.41	**0.76±1.47**
UW-NSDH	0.59±0.12	0.14±0.06	0.47±0.10	**15.29±5.65**	7.91±2.82	**3.95±2.09**	0.78±1.51
Unregistered	0.60±0.13	0.09±0.05	0.24±0.08	24.60±5.56	19.06±4.89	11.86±2.75	-

The best performing methods for each metric are highlighted in bold.

https://doi.org/10.1371/journal.pone.0282110.t004

The use of the augmentation layer resulted in a negligible increase in training runtime of 7.7% per epoch and 0.47% ($\sim$ 74 MB of 16 GB) increase in GPU memory usage (see Fig A in S2 Appendix).

Discussion

Development of CNNs for image registration is challenging, especially when data is scarce. We therefore developed a framework called DDMR to train deep registration models, which we have evaluated through an ablation study. By pretraining a model on a larger dataset, we found that performance can be greatly improved using transfer learning, even if the source domain is from a different image modality or anatomic origin. Through the development of novel augmentation and loss weighting layers, training was simplified by generating artificial moving images on-the-fly, removing the need to store augmented samples on disk, while simultaneously learning to weigh losses in a dynamic fashion. Furthermore, by guiding registration using automatically generated segmentations and adaptive loss weighting, registration performance was enhanced. In addition, negligible increase in inference runtime and GPU memory usage was observed. The added-value of our method lies in the use of generic concepts, which can therefore leverage most deep learning-based registration designs.

From Tables 2 to 5, segmentation guidance boosts the performance of the image registration both on the SG and UW models, further confirmed by the results of the performance contrast analysis shown in Table A in S4 Appendix ($p < 0.001$), and Figs A-C in S5 Appendix,

Table 5. Evaluation of the models trained on the Oslo-CoMet dataset from finetuning in two steps.

Model	SSIM	NCC	DSC	HD	HD95	TRE	Runtime
BL-N	0.52±0.07	**0.19±0.07**	0.24±0.06	60.92±26.06	39.96±30.25	22.34±8.60	0.79±1.52
BL-NS	**0.62±0.10**	0.17±0.07	0.14±0.04	85.71±6.40	60.93±4.15	32.84±11.90	0.76±1.45
SG-ND	0.56±0.12	0.14±0.07	0.44±0.09	16.12±5.29	8.87±2.94	5.12±2.52	**0.77±1.48**
SG-NSD	0.58±0.12	0.15±0.07	0.43±0.08	16.93±6.50	9.17±3.02	5.21±2.40	0.77±1.49
UW-NSD	0.60±0.11	0.15±0.06	**0.53±0.13**	15.13±5.68	**6.97±2.83**	3.40±1.91	**0.77±1.48**
UW-NSDH	0.60±0.12	0.15±0.06	0.50±0.12	**14.79±5.79**	7.37±2.99	3.55±2.14	**0.77±1.48**
SyN	0.61±0.13	0.20±0.07	0.49±0.01	17.93±3.44	9.62±1.57	22.34±4.96	10.01±3.69
SyNCC	0.63±0.13	0.20±0.07	0.49±0.01	18.59±2.99	9.64±1.61	22.31±5.04	323.81±87.13
Unregistered	0.60±0.13	0.09±0.05	0.24±0.08	24.60±5.56	19.06±4.89	11.86±2.75	-

The best performing methods for each metric are highlighted in bold.

https://doi.org/10.1371/journal.pone.0282110.t005

found in the online resources. The introduction of landmarks to guide the training, in the form of boundaries of the different segmentation masks, allows for a better understanding of the regions occupied by each anatomical structure. This observation is drawn by the improvement of the segmentation-based metrics on the finetuned models (see Table 5 frozen encoder). And further confirmed by the statistical tests 2) and 3), in which the Mann-Whitney U test showed significant difference for the segmentation guidance ($p < 0.001$) and uncertainty weighting models ($p = 0.0093$) (see Table C in S4 Appendix). No statistical difference was observed for the baseline models (see Table 5 frozen encoder). A larger dataset is required to fully assess the significance of the transfer learning, as only eleven test samples were available for this study.

Surprisingly, adding HD to the set of losses had limited effect on the performance. We believe this is due to HD being sensitive to outliers and minor annotation errors, which is likely to happen as the annotations used in this study were automatically generated. Furthermore, NCC proved to be a well-suited intensity-based loss function, with no real benefit of adding an additional intensity-based loss function such as SSIM.

From studying the adaptive loss weights evolution during training (see Figs E-L in S3 Appendix), it is possible to deduce an interpretation regarding influence and benefit from each loss component over the network. Evidently, SSIM was favoured over NCC during training, even though SSIM was deemed less useful for image registration compared to NCC. A rationale can be hypothesised from SSIM being easier to optimise, being a perception-based loss. Interestingly, the loss weight curves all seemed to follow the same pattern. Upweighted losses are linearly increased until a plateau is reached and the opposite behaviour happens for the downweighted losses. This may indicate that uncertainty weighting lacks the capacity of task prioritisation, which could have been helpful at a later stage in training. Such an approach has been proposed in the literature [33], simply not for similar image registration tasks. Hence, a comparison of other multi-task learning designs might be worth investigating in future work.

From Tables 4 and 5 it can be observed the benefit of using segmentation guidance for training deep registration models. Furthermore, when compared to the traditional method ANTs, using SyN and SyNCC, a significant improvement on TRE is observed ($p < 0.001$) (Table D in S4 Appendix), with differences close to 17 mm on the Oslo-CoMet test set. Further improved using uncertainty weighting. No significant value was observed between the baseline model and ANTs ($p = 0.5845$), which shows that naive image-only training is not enough for the model to understand the registration task. Not surprisingly, runtimes of the deep registration models are dramatically better than those of ANTs, taking the latter up to five minutes on average using the SyNCC configuration.

A sizeable downside in training CNNs for image registration remains the long training runtime. Having access to pretrained models in order to perform transfer learning alleviates this issue, but the substantial amount of training data required, and in our use case annotated data, persists as another tremendous drawback.

Once deployed, such registration models often fail to generalise to other anatomies, imaging modalities, and data shifts in general, resulting in ad hoc solutions. As part of future work investigations, developing more generic deep image registration models would be of interest, tackling both training and deployment shortcomings.

In this study, only synthetic moving images and mostly algorithm-based annotations were used for evaluation. To verify the clinical relevance of the proposed models, a dataset with manual delineations of structures both for the fixed and moving images, and with clinically relevant movements, is required. To illustrate this situation, Table E in S4 Appendix shows a comparison between manual and automatic segmentations of the parenchyma and

the vascular structures on the Oslo-CoMet test set images. Both DSC and HD95 are reported. A good concordance between the automatic and manual parenchyma segmentations can be observed. However, vascular segmentation poses a more challenging problem for automatic methods to tackle. In future work, assessment of the impact of vascular segmentations of diverse quality could be considered. This investigation would require the delineation of the entire training set, which itself is extremely challenging and was thus deemed outside the scope of this study. Nevertheless, such investigation is of definite value and should be part of future works, additionally including human qualitative evaluation of the clinical relevance.

The sole focus on mono-modal registration can be considered as a limitation from our work. Especially when selecting the loss functions. For instance, in multi-modal registration it is common to use mutual information. Hence, investigating the translation between mono and multi-modal designs is of value to assess applicability over various registration tasks. The recent introduction of the new Learn2Reg challenge dataset [24] represents an adequate alley for further investigation over this aspect. While the U-Net architecture, used in this study, is not recent, a substantial number of publications have favoured it for image registration, as shown to outperform vision transformers on smaller datasets [34]. Alternatively, generative adversarial models should be tested, as these networks have shown to produce more realistic looking images [35]. Self-attention [36] for encoding anatomical information, or graph-based neural networks [37] for improved vascular segmentation-guided registration, are concepts that also should be considered in future work.

Conclusion

In the presented study, we demonstrated that registration models can be improved through transfer learning and adaptive loss weighting even with minimal data without manual annotations. The proposed framework DDMR also enables on-the-fly generation of artificial moving images, without the need to store copies on disk. In future work, DDMR should be validated on data of other anatomies and imaging modalities to further assess its benefit.

Supporting information

S1 Appendix. Additional data details. Additional details about the segmentations produced for the IXI dataset.
(PDF)

S2 Appendix. Resources impact of the augmentation layer. Details on the GPU resources usage by the proposed augmentation layer.
(PDF)

S3 Appendix. Training curves. Figures of the training curves, as well as the adaptive loss weighting.
(PDF)

S4 Appendix. Statistical analysis. Results of the statistical tests described in the manuscript, and additional comparison between manually and automatically generated segmentations.
(PDF)

S5 Appendix. Qualitative results. Examples of predictions on the IXI and Oslo-CoMet test datasets.
(PDF)

S1 File. Evaluation metrics. Collection of the evaluation metrics of the proposed models when tested on both the IXI and the Oslo-CoMet test datasets.
(CSV)

Author Contributions

Conceptualization: Javier Pérez de Frutos.

Data curation: Egidijus Pelanis, David Bouget.

Formal analysis: Javier Pérez de Frutos, André Pedersen.

Funding acquisition: Thomas Langø, Ole-Jakob Elle, Frank Lindseth.

Investigation: Javier Pérez de Frutos, André Pedersen.

Methodology: Javier Pérez de Frutos, André Pedersen.

Software: Javier Pérez de Frutos, André Pedersen, Shanmugapriya Survarachakan.

Supervision: Thomas Langø, Ole-Jakob Elle, Frank Lindseth.

Validation: Javier Pérez de Frutos, André Pedersen.

Writing – original draft: Javier Pérez de Frutos, André Pedersen, Egidijus Pelanis, David Bouget.

Writing – review & editing: Shanmugapriya Survarachakan, Thomas Langø, Ole-Jakob Elle, Frank Lindseth.

References

1. Fretland ÅA, Dagenborg VJ, Bjørnelv GMW, Kazaryan AM, Kristiansen R, Fagerland MW, et al. Laparoscopic Versus Open Resection for Colorectal Liver Metastases. Annals of Surgery. 2018; 267:199–207. https://doi.org/10.1097/SLA.0000000000002353

2. Alam F, Rahman SU, Ullah S, Gulati K. Medical image registration in image guided surgery: Issues, challenges and research opportunities. Biocybernetics and Biomedical Engineering. 2018; 38(1):71–89. https://doi.org/10.1016/j.bbe.2017.10.001

3. Cash DM, Miga MI, Glasgow SC, Dawant BM, Clements LW, Cao Z, et al. Concepts and Preliminary Data Toward the Realization of Image-guided Liver Surgery. Journal of Gastrointestinal Surgery. 2007; 11:844–859. https://doi.org/10.1007/s11605-007-0090-6 PMID: 17458587

4. Pelanis E, Teatini A, Eigl B, Regensburger A, Alzaga A, Kumar RP, et al. Evaluation of a novel navigation platform for laparoscopic liver surgery with organ deformation compensation using injected fiducials. Medical Image Analysis. 2021; 69:101946. https://doi.org/10.1016/j.media.2020.101946 PMID: 33454603

5. Prevost GA, Eigl B, Paolucci I, Rudolph T, Peterhans M, Weber S, et al. Efficiency, Accuracy and Clinical Applicability of a New Image-Guided Surgery System in 3D Laparoscopic Liver Surgery. Journal of Gastrointestinal Surgery. 2020; 24:2251–2258. https://doi.org/10.1007/s11605-019-04395-7 PMID: 31621024

6. Baisa NL, Bricq S, Lalande A. MRI-PET Registration with Automated Algorithm in Pre-clinical Studies. arXiv. 2017;.

7. Martínez-Cecilia D, Wicherts DA, Cipriani F, Berardi G, Barkhatov L, Lainas P, et al. Impact of resection margins for colorectal liver metastases in laparoscopic and open liver resection: a propensity score analysis. Surgical Endoscopy. 2021; 35(2):809–818. https://doi.org/10.1007/s00464-020-07452-4 PMID: 32107633

8. Fusaglia M, Tinguely P, Banz V, Weber S, Lu H. A Novel Ultrasound-Based Registration for Image-Guided Laparoscopic Liver Ablation. Surgical Innovation. 2016; 23:397–406. https://doi.org/10.1177/1553350616637691 PMID: 26969718

9. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging. 2019; 38(8):1788–1800. https://doi.org/10.1109/TMI.2019.2897538 PMID: 30716034

10. Dosovitskiy A, Brox T. Generating Images with Perceptual Similarity Metrics based on Deep Networks. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, editors. Advances in Neural Information Processing Systems. vol. 29; 2016. Available from: https://proceedings.neurips.cc/paper/2016/file/371bce7dc83817b7893bcdeed13799b5-Paper.pdf.

11. Survarachakan S, Prasad PJR, Naseem R, Pérez de Frutos J, Kumar RP, Langø T, et al. Deep learning for image-based liver analysis—A comprehensive review focusing on malignant lesions. Artificial Intelligence in Medicine. 2022; 130:102331. https://doi.org/10.1016/j.artmed.2022.102331 PMID: 35809970

12. Maurer CR, Fitzpatrick JM, Wang MY, Galloway RL, Maciunas RJ, Allen GS. Registration of head volume images using implantable fiducial markers. IEEE Transactions on Medical Imaging. 1997; 16 (4):447–462. https://doi.org/10.1109/42.611354 PMID: 9263002

13. Wu G, Kim M, Wang Q, Gao Y, Liao S, Shen D. Unsupervised Deep Feature Learning for Deformable Registration of MR Brain Images. In: Medical Image Computing and Computer Assisted Intervention. vol. 16; 2013. p. 649–656.

14. Jaderberg M, Simonyan K, Zisserman A, kavukcuoglu k. Spatial Transformer Networks. In: Advances in Neural Information Processing Systems. vol. 28; 2015. Available from: https://proceedings.neurips.cc/paper/2015/file/33ceb07bf4eeb3da587e268d663aba1a-Paper.pdf.

15. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234–241.

16. Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: Fast predictive image registration—A deep learning approach. NeuroImage. 2017; 158:378–396. https://doi.org/10.1016/j.neuroimage.2017.07.008 PMID: 28705497

17. Rohé MM, Datar M, Heimann T, Sermesant M, Pennec X. SVF-Net: Learning Deformable Image Registration Using Shape Matching. vol. 2878; 2017. p. 266–274.

18. Mok TCW, Chung ACS. Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 12263; 2020. p. 211–221.

19. Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, et al. Weakly-supervised convolutional neural networks for multimodal image registration. Medical Image Analysis. 2018; 49:1–13. https://doi.org/10.1016/j.media.2018.07.002 PMID: 30007253

20. Li H, Fan Y. Non-rigid image registration using self-supervised fully convolutional networks without training data. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018. p. 1075–1078.

21. Fu Y, Lei Y, Wang T, Higgins K, Bradley JD, Curran WJ, et al. LungRegNet: An unsupervised deformable image registration method for 4D-CT lung. Medical Physics. 2020; 47(4):1763–1774. https://doi.org/10.1002/mp.14065 PMID: 32017141

22. Lei Y, Fu Y, Harms J, Wang T, Curran WJ, Liu T, et al. 4D-CT Deformable Image Registration Using an Unsupervised Deep Convolutional Neural Network. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2019; 11850:26–33.

23. Hu J, Luo Z, Wang X, Sun S, Yin Y, Cao K, et al. End-to-end multimodal image registration via reinforcement learning. Medical Image Analysis. 2021; 68. https://doi.org/10.1016/j.media.2020.101878 PMID: 33197714

24. Hering A, Hansen L, Mok TCW, Chung ACS, Siebert H, Häger S, et al. Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Transactions on Medical Imaging. 2022; p. 1–1. https://doi.org/10.1109/TMI.2022.3213983 PMID: 36264729

25. Fretland ÅA, Kazaryan AM, Bjørnbeth BA, Flatmark K, Andersen MH, Tønnessen TI, et al. Open versus laparoscopic liver resection for colorectal liver metastases (the Oslo-CoMet study): Study protocol for a randomized controlled trial. Trials. 2015; 16(1):73. https://doi.org/10.1186/s13063-015-0577-5 PMID: 25872027

26. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage. 2011; 54(3):2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025 PMID: 20851191

27. Pedersen A. andreped/livermask: v1.3.1; 2021.

28. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al.. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems; 20115. Available from: https://www.tensorflow.org.

29. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Springer, Cham; 2016. p. 424–432.

30. Cipolla R, Gal Y, Kendall A. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018. p. 7482–7491.

31. Seabold S, Perktold J. statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference; 2010.

32. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020; 17:261–272. https://doi.org/10.1038/s41592-019-0686-2 PMID: 32015543

33. Guo M, Haque A, Huang DA, Yeung S, Fei-Fei L. Dynamic Task Prioritization for Multitask Learning. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018.

34. Jia X, Bartlett J, Zhang T, Lu W, Qiu Z, Duan J. U-Net vs Transformer: Is U-Net Outdated in Medical Image Registration? arXiv. 2022. https://doi.org/10.48550/arXiv.2208.04939

35. Bhadra S, Zhou W, Anastasio MA. Medical image reconstruction with image-adaptive priors learned by use of generative adversarial networks. 2020; 11312:206–213. https://doi.org/10.1117/12.2549750

36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. In: Advances in Neural Information Processing Systems. vol. 30; 2017. Available from: https://papers.nips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

37. Montaña-Brown N, Ramalhinho Ja, Allam M, Davidson B, Hu Y, Clarkson MJ. Vessel segmentation for automatic registration of untracked laparoscopic ultrasound to CT of the liver. International Journal of Computer Assisted Radiology and Surgery. 2021; 16(7):1151–1160. https://doi.org/10.1007/s11548-021-02400-6 PMID: 34046826

PAPER IV

Online supplementary material

S1: Additional data details

Javier Pérez de Frutos[1*], André Pedersen[1,2,3], Egidijus Pelanis[4], David Bouget[1], Shanmugapriya Survarachakan[5], Thomas Langø[1,6], Ole-Jakob Elle[4], and Frank Lindseth[5]

[1]Department of Health Research, SINTEF, Trondheim, Norway
[2]Department of Clinical and Molecular Medicine, Norwegian University of Science and University (NTNU), Trondheim, Norway
[3]Clinic of Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[4]Intervention Centre, Oslo University Hospital, Oslo, Norway
[5]Department of Computer Science, Norwegian University of Science and University (NTNU), Trondheim, Norway
[6]Research Department, Future Operating Room, St. Olavs hopsital, Trondheim University Hospital, Trondheim, Norway
[*]Corresponding author: Javier Pérez de Frutos, javier.perezdefrutos@sintef.no

Document description

This document contains additional details about the segmentations produced for the IXI dataset. A total of 28 regions where segmented on the IXI dataset using the MNI 152 space atlas. From these, 24 classes distinguished between left and right sections, which were merged together resulting in a collection of 16 labels. A total of 17 labels, including the background, were used. For more details, see Table A.

Table A. Labels used from the MNI 152 space atlas.

Parietal grey matter (left and right)	Fornix (left and right)
Lateral ventricle (left and right)	Caudate (left and right)
Occipital grey matter (left and right)	Cerebellum (left and right)
Globus pallidus (left and right)	Thalamus (left and right)
Putamen (left and right)	Frontal grey matter (left and right)
Subthalamic nucleus (left and right)	Temporal grey matter (left and right)
3rd ventricle	Brain stem
4th ventricle	Cerebrospinal fluid
	Background

S2: Resources impact of the augmentation layer

Javier Pérez de Frutos[1*], André Pedersen[1,2,3], Egidijus Pelanis[4], David Bouget[1], Shanmugapriya Survarachakan[5], Thomas Langø[1,6], Ole-Jakob Elle[4], and Frank Lindseth[5]

[1]Department of Health Research, SINTEF, Trondheim, Norway
[2]Department of Clinical and Molecular Medicine, Norwegian University of Science and University (NTNU), Trondheim, Norway
[3]Clinic of Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[4]Intervention Centre, Oslo University Hospital, Oslo, Norway
[5]Department of Computer Science, Norwegian University of Science and University (NTNU), Trondheim, Norway
[6]Research Department, Future Operating Room, St. Olavs hopsital, Trondheim University Hospital, Trondheim, Norway
[*]Corresponding author: Javier Pérez de Frutos, javier.perezdefrutos@sintef.no

Document description

This document contains details on the GPU resources usage by the proposed augmentation layer. Figs. Aa and Ab which show the GPU usage during one training epoch. The relative increase in time is of 7.7% (ca. 10.5 s). Fig. Ac shows the memory usage on the GPU during one training epoch with and without using the augmentation layer. The increase in memory allocation is of 0.47% (ca. 74 MB / 16 GB).
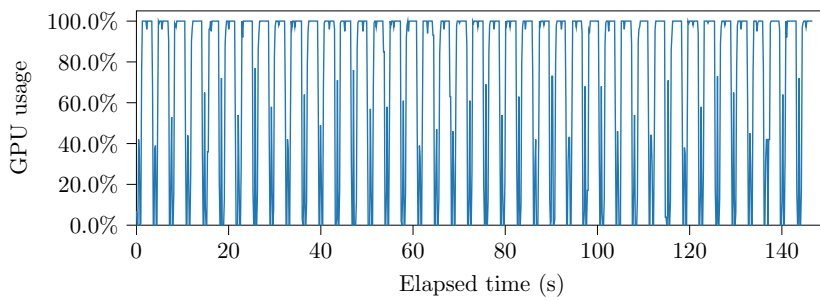
(a) GPU usage with augmentation layer
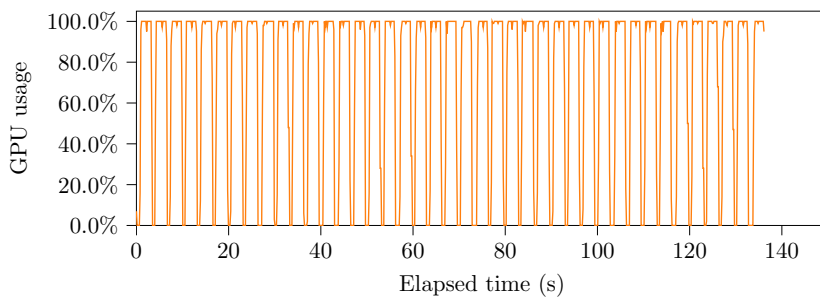
(b) GPU usage without augmentation layer
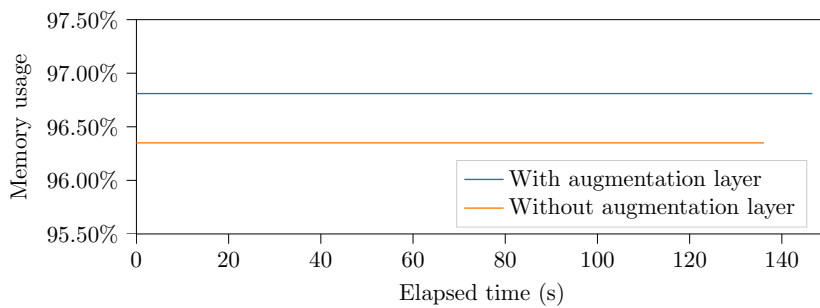
(c) Memory usage with (blue) and without using the augmentation layer
(orange)

Figure A. Impact of the augmentation layer on the GPU resources during one training epoch.

S3: Training curves

Javier Pérez de Frutos[1*], André Pedersen[1,2,3], Egidijus Pelanis[4], David Bouget[1], Shanmugapriya Survarachakan[5], Thomas Langø[1,6], Ole-Jakob Elle[4], and Frank Lindseth[5]

[1]Department of Health Research, SINTEF, Trondheim, Norway
[2]Department of Clinical and Molecular Medicine, Norwegian University of Science and University (NTNU), Trondheim, Norway
[3]Clinic of Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[4]Intervention Centre, Oslo University Hospital, Oslo, Norway
[5]Department of Computer Science, Norwegian University of Science and University (NTNU), Trondheim, Norway
[6]Research Department, Future Operating Room, St. Olavs hopsital, Trondheim University Hospital, Trondheim, Norway
[*]Corresponding author: Javier Pérez de Frutos, javier.perezdefrutos@sintef.no

Document description

This document contains figures showing the training curves, as well as figures related to the adaptive loss weighting.

Training curves

Figs. A - D show training and validation losses for the six different configurations described in the main manuscript, for the IXI and Oslo-CoMet datasets (see Table 1, Section 2.5). Furthermore, training runtime of 160 epochs is shown as part of the legends of both figures for reference.
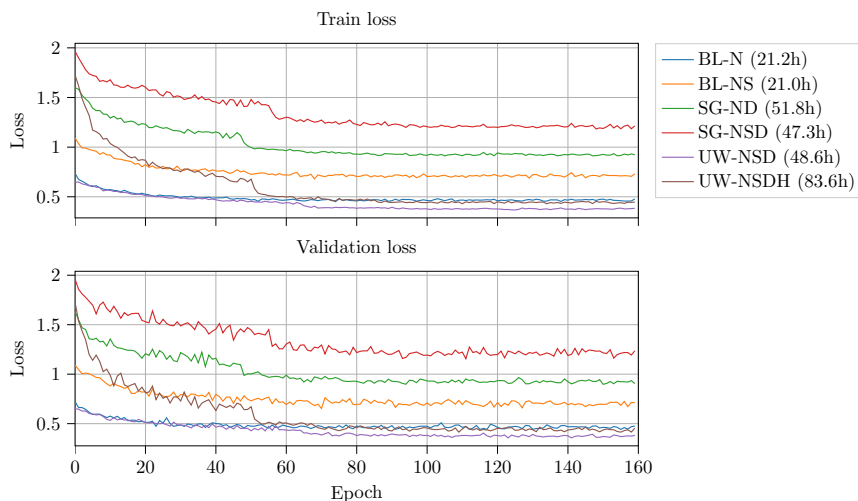
Figure A. Train and validation loss of the models trained on the IXI dataset.
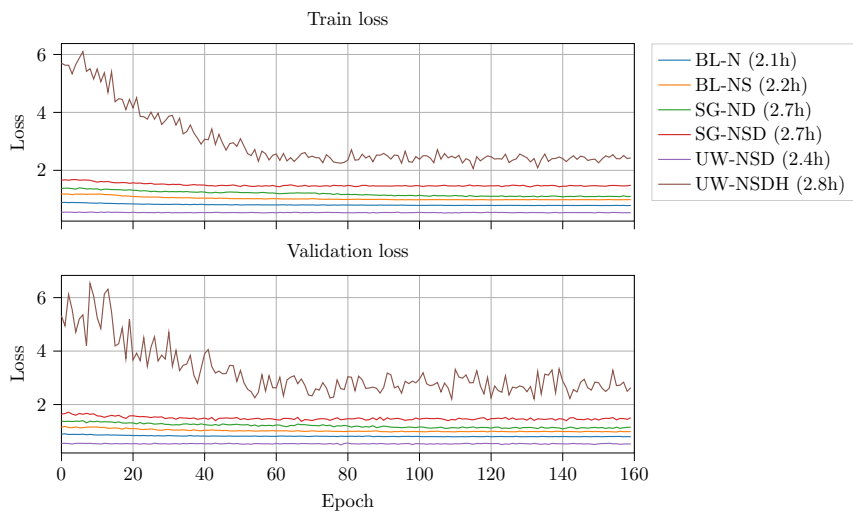
Figure B. Train and validation loss of the models trained on the Oslo-CoMet dataset.
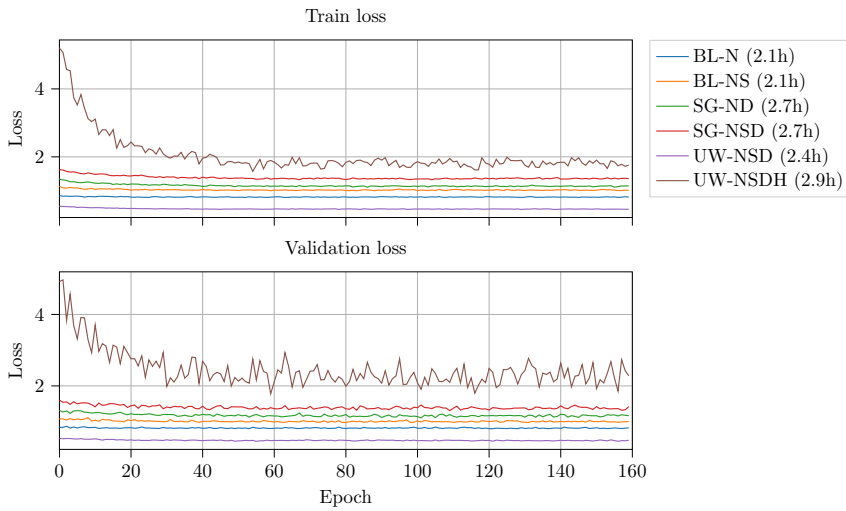
Figure C. Train and validation loss of the models trained on the Oslo-CoMet dataset from finetuning the entire architecture.
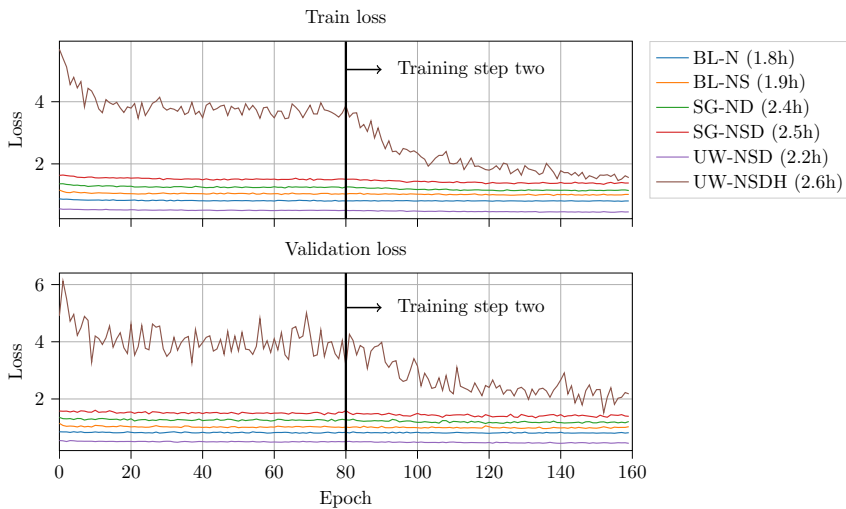

Figure D. Train and validation loss of the models trained on the Oslo-CoMet dataset from finetuning in two steps.

Adaptive loss weighting results

Figs. E - L show the evolution of the dynamic weight in the UW models for the studied cases.
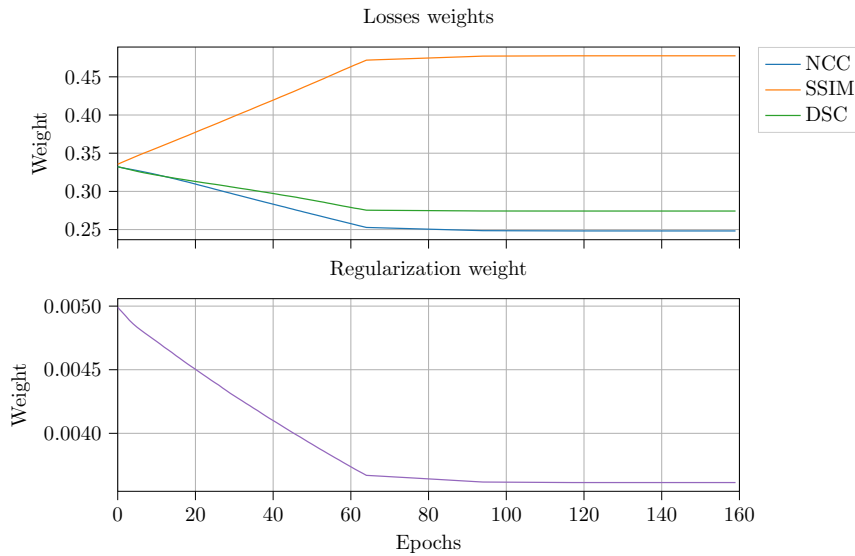
Figure E. Loss weights of the model UW-NSD trained on the IXI dataset.
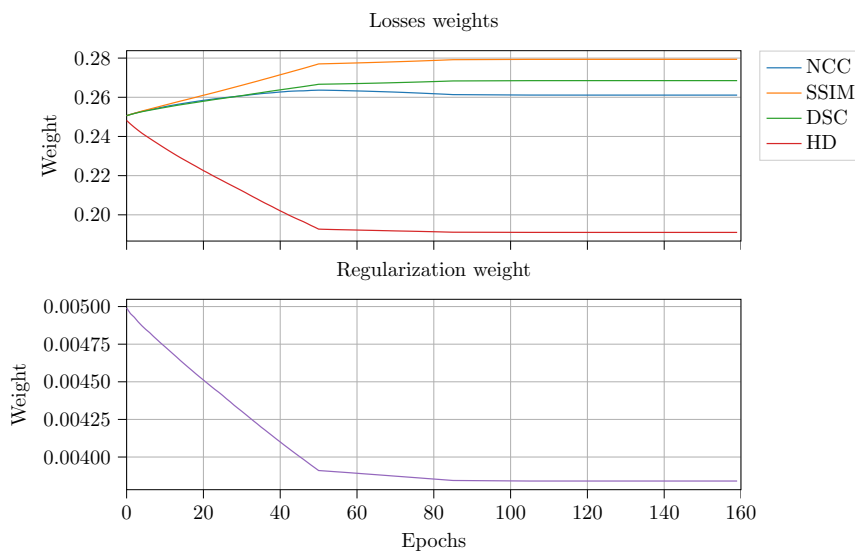
Figure F. Loss weights of the model UW-NSDH trained on the IXI dataset.
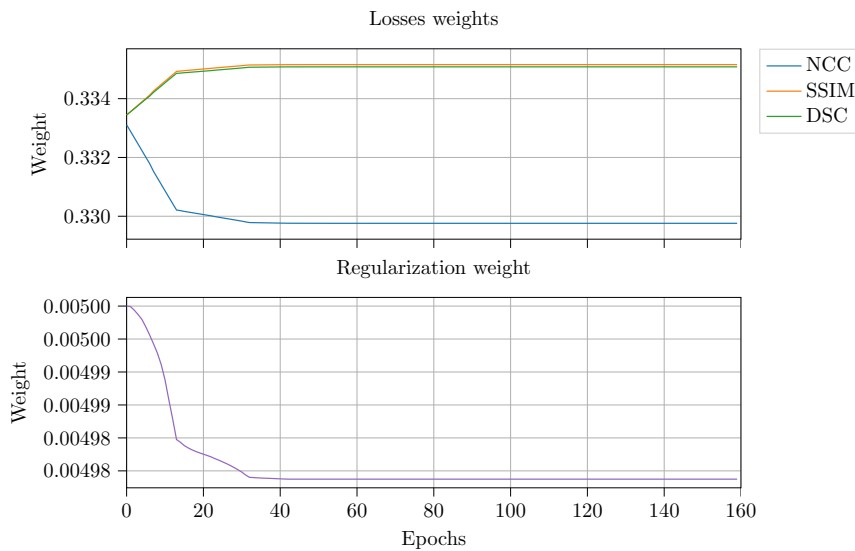
Figure G. Loss weights of the model UW-NSD trained on the Oslo-CoMet dataset.
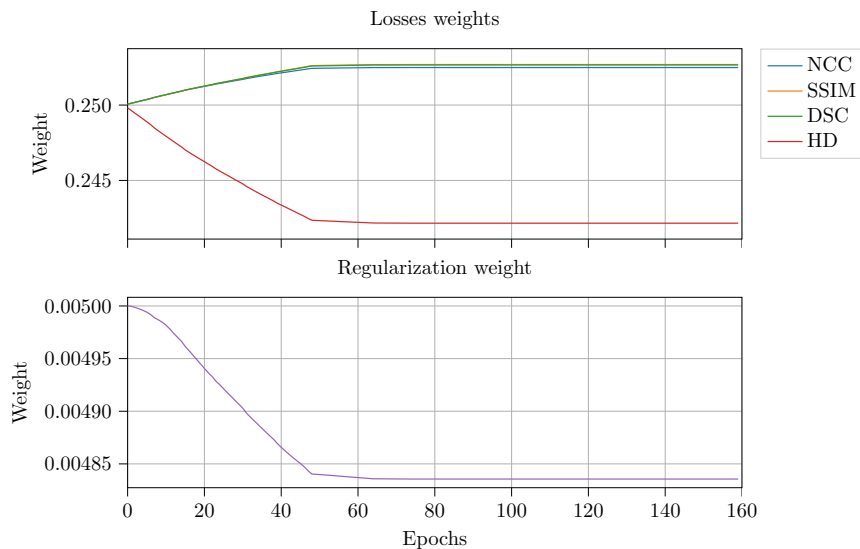
Figure H. Loss weights of the model UW-NSDH trained on the Oslo-CoMet dataset.
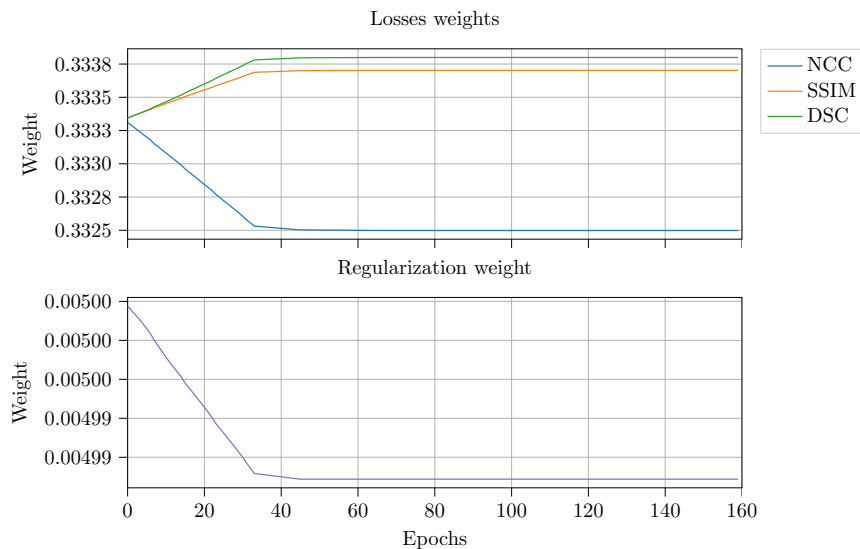
Figure I. Loss weights of the model UW-NSD trained on the Oslo-CoMet dataset from finetuning the entire architecture.
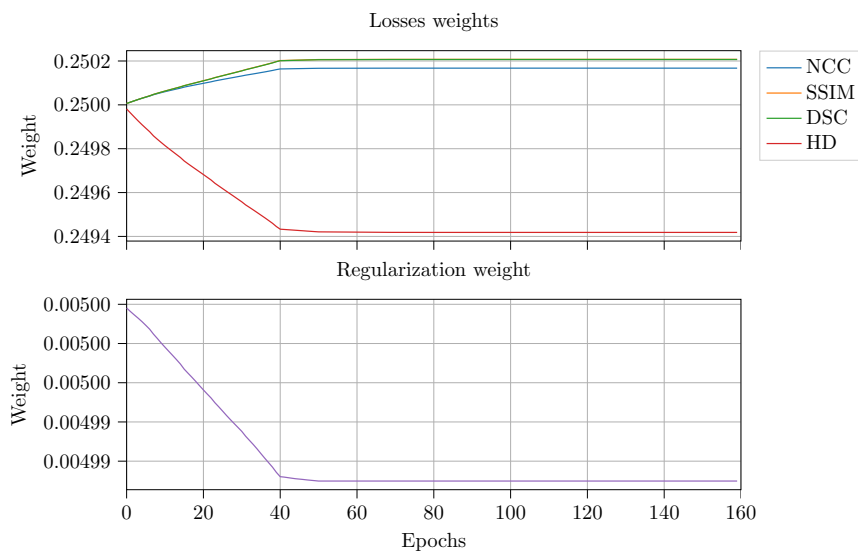
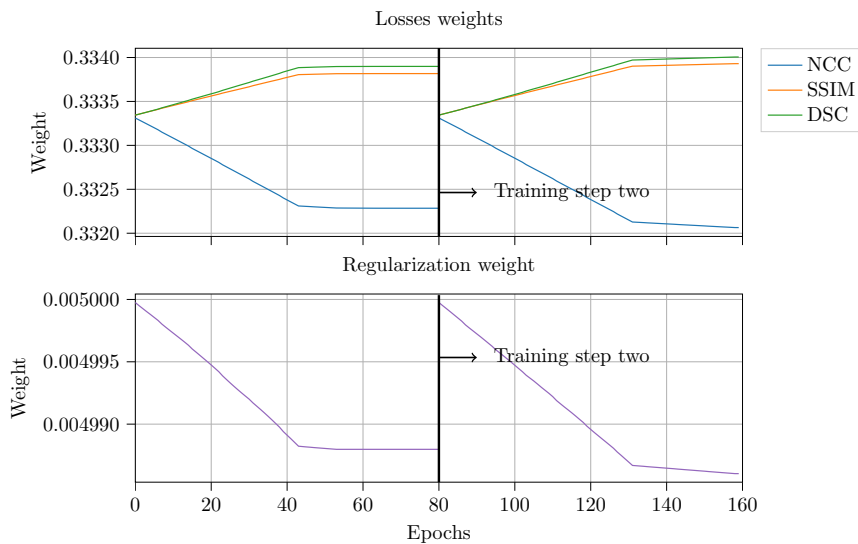**Figure J.** Loss weights of the model UW-NSDH trained on the Oslo-CoMet dataset from finetuning the entire architecture.

Figure K. Loss weights of the model UW-NSD trained on the Oslo-CoMet dataset from finetuning in two steps.
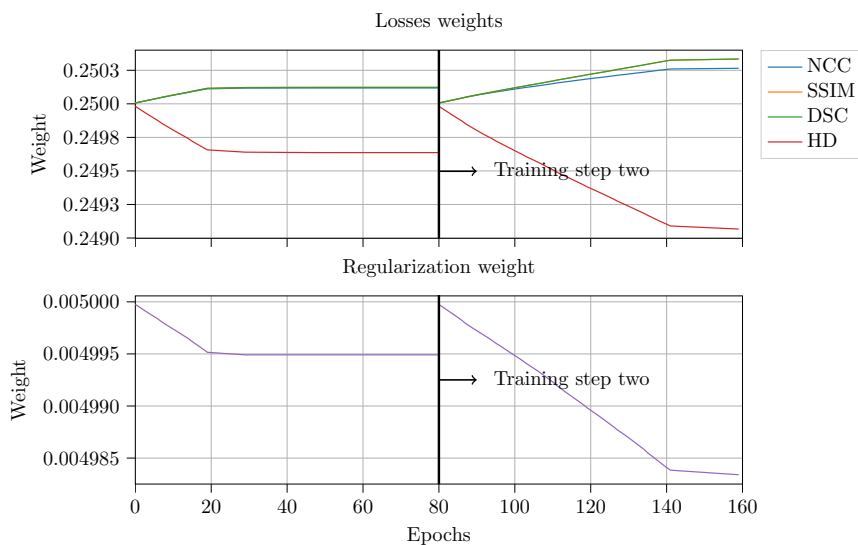
Figure L. Loss weights of the model UW-NSDH trained on the Oslo-CoMet dataset from finetuning in two steps.

S4: Statistical analysis

Javier Pérez de Frutos[1*], André Pedersen[1,2,3], Egidijus Pelanis[4], David Bouget[1], Shanmugapriya Survarachakan[5], Thomas Langø[1,6], Ole-Jakob Elle[4], and Frank Lindseth[5]

[1]Department of Health Research, SINTEF, Trondheim, Norway
[2]Department of Clinical and Molecular Medicine, Norwegian University of Science and University (NTNU), Trondheim, Norway
[3]Clinic of Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[4]Intervention Centre, Oslo University Hospital, Oslo, Norway
[5]Department of Computer Science, Norwegian University of Science and University (NTNU), Trondheim, Norway
[6]Research Department, Future Operating Room, St. Olavs hopsital, Trondheim University Hospital, Trondheim, Norway
[*]Corresponding author: Javier Pérez de Frutos, javier.perezdefrutos@sintef.no

Document description

This document contains tables showing detailed results of the statistical tests described in the manuscript, and a comparison between manually and automatically generated segmentations.

Statistical analysis

Five sets of statistical tests were conducted to assess: 1) performance contrasts between model designs (see Table A), 2) benefit of transfer learning (see Table B), 3) benefit of segmentation-guiding, and 4) benefit of uncertainty weighting (see Table C), and 5) performance contrasts between the baseline and segmentation-guided models, and the traditional methods in ANTs (SyN and SyNCC) (see Table D). For the tests, The TRE metric was used, as it is considered the gold standard for surgical practitioners. Tests were conducted on the evaluations of the IXI and Oslo-CoMet test datasets only. For the tests only involving the Oslo-CoMet dataset, the two-step transfer learning approach was used as reference. A significance level of 5% was used to determine statistical significance.

Table A. Results of multiple comparisons comparing TRE between all designs evaluated on the IXI dataset.

	BL-NS	SG-ND	SG-NSD	UW-NSD	UW-NSDH
BL-N	0.9	<0.001	<0.001	<0.001	<0.001
BL-NS	-	<0.001	<0.001	<0.001	<0.001
SG-ND	-	-	0.9	0.9	0.9
SG-NSD	-	-	-	0.9	0.9
UW-NSD	-	-	-	-	0.9

Table B. Results of hypotheses tests assessing the benefit of transfer learning on the Oslo-CoMet dataset in terms of TRE for the three designs: BL-N, SG-NSD, and UW-NSD.

Model	p-value
BL-N	0.8608
SG-NSD	0.0021
UW-NSD	0.0014

Table C. Results of hypotheses tests assessing the added value of segmentation-guiding and uncertainty weighting on the Oslo-CoMet dataset.

Model	p-value
Segmentation-guiding	<0.001
Uncertainty-weighting	0.0093

Table D. Results of hypotheses tests assessing the performance of the baseline and segmentation-guiding models compared to traditional method ANTs (SyN and SyNCC approaches), on the Oslo-CoMet dataset.

AI Model	ANTs approach	p-value
BL-N	SyN	0.5845
	SyNCC	0.5845
SG-NSD	SyN	<0.001
	SyNCC	<0.001

Comparison between manual and automatic annotations

Manual and automatic segmentations of the parenchyma and the vascular structures were performed on the Oslo-CoMet test set images. In Table E the DSC and HD95 are reported.

Table E. DSC and HD95 comparing manual and automatic annotations on the Oslo-CoMet test set images.

Label	DSC	HD95
Parenchyma	0.946±0.046	10.122±11.032
Vessels	0.355±0.090	24.872±5.161

S5: Qualitative results

Javier Pérez de Frutos[1*], André Pedersen[1,2,3], Egidijus Pelanis[4], David Bouget[1], Shanmugapriya Survarachakan[5], Thomas Langø[1,6], Ole-Jakob Elle[4], and Frank Lindseth[5]

[1]Department of Health Research, SINTEF, Trondheim, Norway
[2]Department of Clinical and Molecular Medicine, Norwegian University of Science and University (NTNU), Trondheim, Norway
[3]Clinic of Surgery, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[4]Intervention Centre, Oslo University Hospital, Oslo, Norway
[5]Department of Computer Science, Norwegian University of Science and University (NTNU), Trondheim, Norway
[6]Research Department, Future Operating Room, St. Olavs hopsital, Trondheim University Hospital, Trondheim, Norway
[*]Corresponding author: Javier Pérez de Frutos, javier.perezdefrutos@sintef.no

Document description

This document contains examples of predictions on the IXI and Oslo-CoMet test datasets.
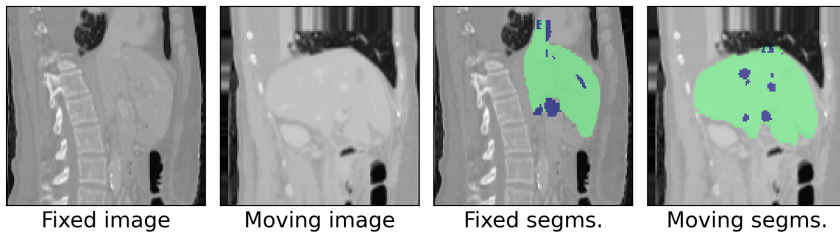

Fixed image   Moving image   Fixed segms.   Moving segms.

(a) Sample test image.

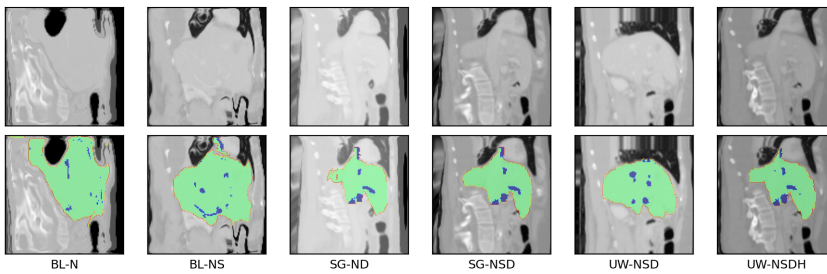BL-N   BL-NS   SG-ND   SG-NSD   UW-NSD   UW-NSDH
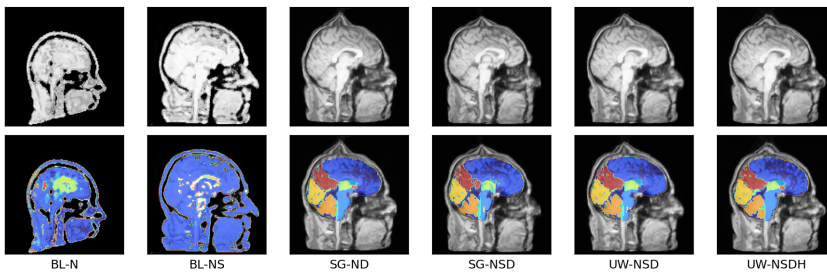
(b) Predictions of the evaluated models on the sample image.

Figure A. Sample image of the IXI test set, and predictions of the models.
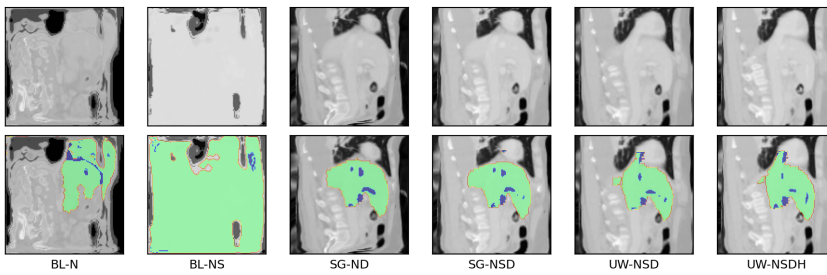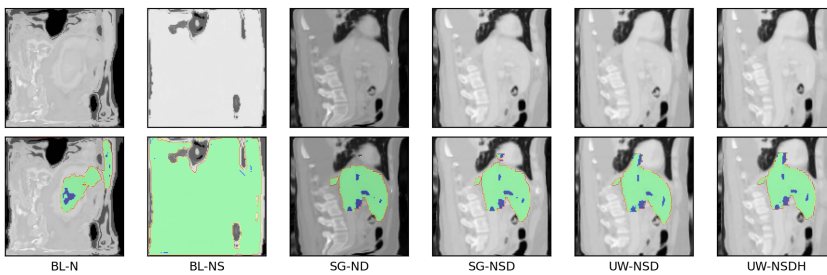
(a) Sample test image.

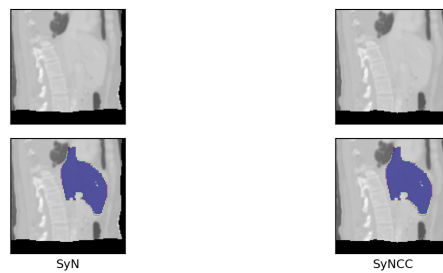(b) Predictions of the models trained on the Oslo-CoMet dataset.

Figure B. Sample image of the Oslo-CoMet test set, and predictions of the models trained on the Oslo-CoMet dataset.

(a) Predictions of the models trained on the Oslo-CoMet dataset from finetuning the entire architecture.

(b) Predictions of the models trained on the Oslo-CoMet dataset from finetuning in two steps.

(c) Predictions of the SyN and SyNCC algorithms.

Figure C. Predictions of the finetuned models and ANTs on the Oslo-CoMet test set sample.

NTNU

Norwegian University of
Science and Technology