Registration of Volumetric Ultrasound Images Using Siamese Neural Networks

Amir Pirhadi

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

November 2021

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Amir Pirhadi**

Entitled: **Registration of Volumetric Ultrasound Images Using Siamese Neural Networks**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. M.N.S. Swamy

_____ External Examiner
Dr. Nizar Bouguila

_____ Examiner
Dr. M.N.S. Swamy

_____ Co-supervisor
Dr. Hassan Rivaz

_____ Co-supervisor
Dr. M. Omair Ahmad

Approved by _____
Dr. Yousef R. Shayan, Chair
Department of Electrical and Computer Engineering

_____ 2021

_____
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Registration of Volumetric Ultrasound Images Using Siamese Neural Networks

Amir Pirhadi

In brain tumor resection, soft tissue deformation (i.e., brain shift) causes the pre-operative images to be invalid. Intra-operative ultrasound imaging is a non-invasive, portable, real-time, and cost-efficient alternative to track the surgery. A robust registration method is required to accurately align the post-resection and pre-resection ultrasound images for maximum tumor resection. However, registering the ultrasound images before and after resection is challenging for two main reasons. First, the tumor cavity after the surgery does not have correspondence in ultrasound images before the resection. Second, the quality of the image reduces during the surgery due to the blood clotting agents, air bubbles, and the saline water around the tumor.

This thesis proposes a robust non-rigid registration method based on a landmark tracking technique for brain shift correction in intra-operative ultrasound images. Some landmarks are selected in pre-resection ultrasound image manually that allows user-interaction. A Siamese neural network is adapted to track the annotated landmarks in the post-resection ultrasound image. The 2.5D approach enables 3D tracking and outlier detection. An optimal affine transformation is calculated using Iterative re-weighted least square (IRLS), which automatically suppresses the outliers.

The proposed method is tested on two publicly available datasets of REtroSpective Evaluation of Cerebral Tumors (RESECT) and Brain Images of Tumors for Evaluation (BITE). Mean target registration error (mTRE) is exploited for registration evaluation. In the BITE dataset, the method decreases the initial miss-alignment from 3.55±2.29 mm to 1.80±0.84 mm in pre/post-resection registration. In the RESECT dataset, mTRE is decreased from 3.55±1.76 mm and 3.49±1.56 mm to

1.26±0.57 mm and 1.12±0.46 mm in pre/post-resection registration and pre/during-resection registration, respectively. The fine-tuning effect is also assessed and is shown the generality of the method. The proposed method is compared to the state-of-the-art methods with statistical tests and showed average comparable or better results. The great accuracy, flexibility, and time-efficiency of the method make it an attractive option in real clinical applications that can increase the performance in neurosurgery.

Acknowledgments

I want to express my deepest gratitude to my knowledgeable and supportive supervisor Dr. Hassan Rivaz. His trust in me made me more confident, and his guidance helped me find the best way through my research. I also want to thank you to My co-supervisor, Dr. Omair Ahmad, who pushed me forward and made me achieve more. Dr. Yiming Xiao also played an invaluable role in making this work happen. He was always providing excellent advice and showing great enthusiasm in the project.

I also want to thank my great colleges in the IMPACT lab and my friends for their technical and emotional support. And last but not least, my dear family that was always there for me in all the critical moments of my life.

This thesis was not possible without all the help and support from these people, and this was a great pleasure for me to have this chance to know them and work with them.

Contents

List of Figures

List of Tables

List of Abbreviations

FOV:	Field of View
RF:	Radio Frequency
B-mode:	Brightness Image
MR:	Magnetic Resonance
CT:	Computed Tomography
IGS:	Image-guided Surgery
GMI:	Gometric Moment Invariant
LOG:	Laplacian of Gaussian
SIFT:	Scale-invariant Feature Transform
CNN:	Convolutional Neural Network
CC:	Correlation Coefficient
NCC:	Normalized Cross-correlation
RESECT:	RetroSpective Evaluation of Cerebral Tumors
ASMUS:	Advances in Simplifying Medical UltraSound
IJCARS:	International Journal for Computer Assisted Radiology and Surgery
iMRI:	Intra-operatie Magnetic Resonance Imaging
iUS:	Intra-operative Ultrasound
IRLS:	Iterative Re-weighted Least Squares
SGD:	Stochastic Gradient Decent
mTRE:	Mean Target Registration Error
BITE:	Brain Images of Tumors for Evaluation

DOF:	Degrees of Freedom
TPS:	Thin-plate Spline
NSERC:	Natural Sciences and Engineering Research Council of Canada
CAM:	Class Activation Mapping

Chapter 1

Introduction

1.1 Ultrasound Imaging

Ultrasound machines produce sound waves with a frequency range of 1-20 MHz, which is higher than the audible limit of human hearing. Ultrasound imaging is a non-invasive, cost-efficient, and portable modality that makes it a suitable alternative to other modalities in many medical applications. It has a variety of usage from diagnosis purposes like tumor detection and fetus monitoring during pregnancy to therapeutic applications such as hemostasis, cataract surgery, blood clots removal, cancer therapy, etc [5, 6].

If a piezoelectric material is mechanically excited, it generates an ultrasound wave. By applying the ultrasound wave to the human body, it propagates into different organs that exhibit dissimilar propagation impedance [7]. Attenuation causes a part of the wave to transfer to heat, and it has a direct relation with the frequency [8]. The impedance difference causes a part of the wave to back-propagate to a receiver. Knowing the speed of the wave and the time, we can calculate the depth of different objects and organs inside the body. These piezoelectric materials are the critical elements of transducers in ultrasound probes that transmit and receive ultrasound waves.

There are different types of ultrasound probes based on transducers arrangement. Linear, curvilinear, and phased array probes are the most popular options. They are utilized based on the depth, location, and structure of the object of interest. The linear probes are presented with a linear arrangement of the transducers as their name implied. Since they have a relatively higher frequency

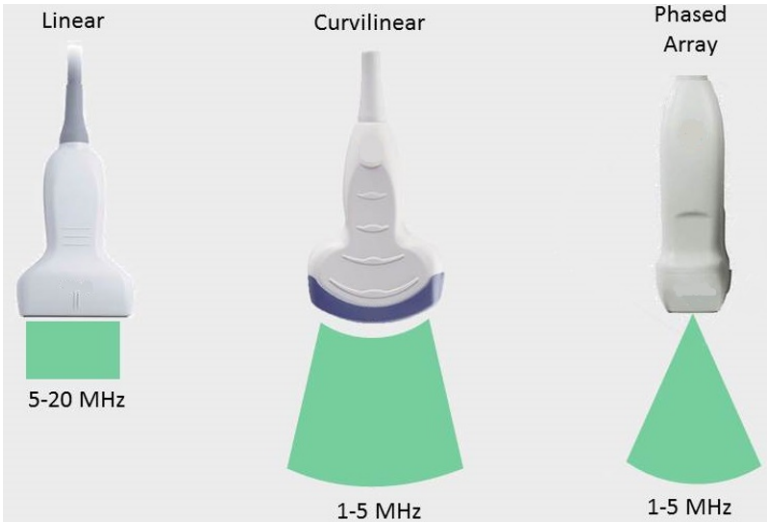

Figure 1.1: Different types of probes with specifying their frequency range. Image courtesy of: https://ameultrasounds.com/blogs/ultrasound-series/probes-and-transducers

than other types, they are good choices for near-the-body surface applications like vascular imaging. On the other hand, the curvilinear probes provide a larger field of view (FOV) and are used for deeper objects like abdominal imaging. The phased array probes were demonstrated to have two main advantages. First, by delaying the firing of different transducer elements, they can change the focus without moving the probe. Second, a smaller transducer's face gives the ability to capture images from a small acoustical window. You can see an illustration of different probes in Fig. 1.1

The received echo from each transducer element of the probe creates a radio frequency (RF) line. An RF image is not simply interpretable, so a 2D brightness image (B-mode) is computed by processing the initial RF image (see Fig. 1.2). 3D images can be constructed from the 2D images as well. High-frequency ultrasound probes provide higher resolution images than other modalities like magnetic resonance (MR) and computed tomography (CT) images. However, higher frequency means more attenuation and decreases the reachable depth. Despite the advantages of ultrasound imaging, there are some downsides, such as providing low contrast images, inability to get through the hard structures like bones, and containing noise and artifacts [9].

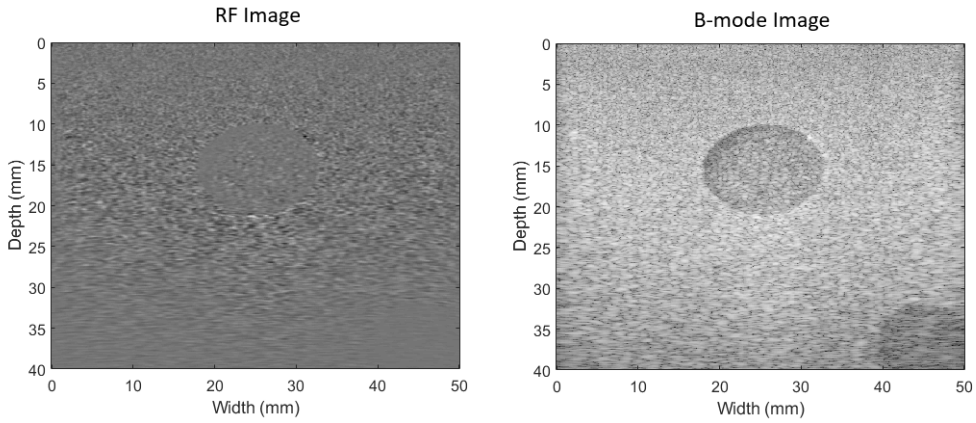

Figure 1.2: A visual comparison between RF and B-mode images.

1.2 Image registration

Image registration is the process of overlaying two or more images into one coordinate system [10]. Images are taken from different points of view, with different sensors, or at different times. Each of the images contains valuable and complementary information, so overlaying them would be of high benefit [11]. Image registration is widely used in many applications such as remote sensing, computer vision and medical imaging [12].

Image registration is a necessary part of many medical applications. Motion correction, template matching, and treatment follow-up are some of the examples. There are different imaging modalities in medical imaging that each one demonstrates a specific feature. Fusion of different modalities of the same scene would be more informative [13, 14]. In image-guided surgery (IGS), image registration was shown to help the surgeon achieve more precise results. Tracking surgical tools in the operating room, overlaying images to patient coordinate, and monitoring the operation site is crucial in IGS and is possible with registration techniques [15–17]. Different modalities and datasets show different natures, and each task requires specific accuracy and aims for a unique goal. To address all the applications, the presence of one global registration method is irrational, and there are different techniques required [18].

Image registration can be divided into different categories. Concerning the images modalities,

there is multi-modal and uni-modal registration. In the case of the user-interaction, there is automatic, semi-automatic, and manual registration. In terms of the algorithm, there are intensity-based and feature-based registration. In the intensity-based methods, voxels intensities of the images are used to be compared and matched between images. Unlike the intensity-based methods, some features are extracted in the feature-based registration, and the registration is done based on them. The feature-based methods are simpler, less computationally expensive, more robust to changing the modality, and presence of artifacts compared to the intensity-based methods [19, 20]. They are also more suitable in the cases of a large initial miss-alignment [21].

In feature-based methods, the anatomical structures of the images are captured during feature extraction, and the corresponding features are matched to find the geometrical transformation. Generating an attribute vector in a patch-based manner for all the voxels in the image is a common approach. In [22, 23], they employed geometric moment invariant (GMI) feature vector. It is shown that the GMI features are venerable to the presence of speckles in the ultrasound images. The intensity value along with the gradient magnitude and Laplacian of Gaussian (LOG) are the other features that showed a good performance in medical image registration applications [24, 25]. Utilizing the scale-invariant feature transform (SIFT), [26] is another popular method in computer vision and the medical field due to its ability to perform well in different image scales and orientations [27, 28]. Segmentation of the salient anatomical structures that keep their correspondence between the images showed promising results as well [29].

Deep learning methods have been widely used in the application of image registration in recent years. It has been shown that the convolutional neural networks (CNN) are a superior alternative to the classical feature extractor methods [12]. Their deep architecture allows deep feature extraction from the images that is more descriptive and general than hand-crafted features like SIFT. They are data-driven and end-to-end, so they usually do not need any pre-processing on the images. Although the training phase is time-consuming, they have a fast performance in the inference phase. In [30], they utilized an unsupervised learning strategy based on convolutional stacked auto-encoder to perform deep feature extraction and matching for registering MR images.

Landmark detection and tracking provide proper guidance to the registration. Classical methods like block-matching try to create a patch around the landmark and find its correspondence in the

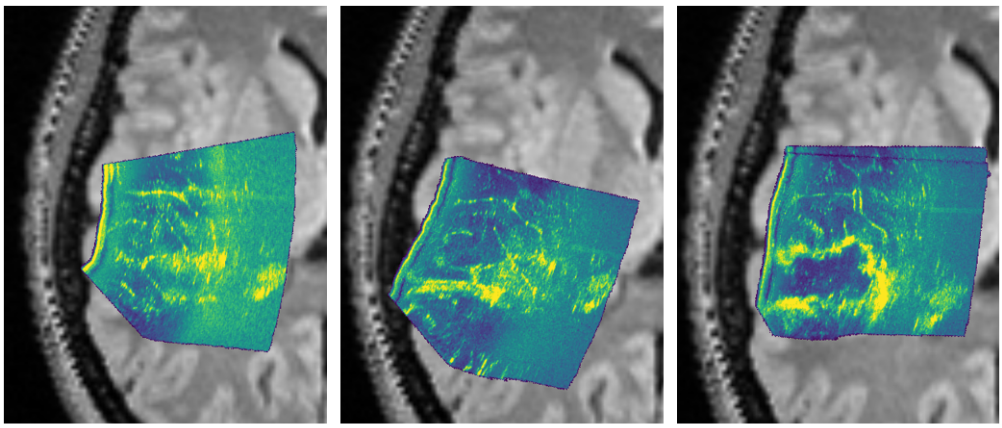

Figure 1.3: Intra-operative ultrasound images overlaied on the pre-operative MRI. Data collected from the RESECT dataset [1].

other image by using a comparison metric like correlation coefficient (CC) [31]. Since the deep neural networks showed an excellent ability to capture the spatial structures in the image, they have been used to find landmarks in varying medical image modalities [32, 33]. The Siamese neural networks have gained popularity in computer vision tasks for landmark and object tracking. They derive high-level features from the input images with respect to a similarity criteria. The Siamese networks have been utilized in medical image registration as well and showed promising results in terms of computational time and reducing the initial miss-alignment [34, 35].

Transformation is an essential part of a registration algorithm and is determined based on the task. They can be classified into two main classes of rigid and non-rigid transformations. The rigid transformation only consists of rotation and shift, so do not change the dimension of the object. Rigid transformation is used for registering rigid structures like bones or as a preceding step before non-rigid registration [19]. Non-rigid registration is more common because they can estimate more complex deformations. The affine transformation with 12 parameters in 3D is a non-rigid transformation that preserves collinearity. Ratios are stored, Lines would remain lines, and parallel lines stay unchanged. Non-linear B-spline transformation is a more complex transformation that can accomplish more accurate results. Unlike the affine transformation, it is local and more vulnerable to local minima convergence.

1.3 Brain shift correction

In tumor resection surgery after craniotomy and opening the dura, pre-operative images become invalid due to the brain deformation (i.e., brain shift). Brain shift occurred due to many factors like gravity, brain swallow, and drug administration. Intra-operative ultrasound images are a good modality to track the surgery during the procedure. They do not need a dedicated operating room, are non-invasive, and are cost-efficient. Tumor resection introduces more complex deformation around the resection site and decreases the quality of the image by producing air bubbles, debris, and blood clots. A robust non-rigid registration method is required to align the intra-operative ultrasound images acquired in different stages of the surgery. Fig. 1.3 illustrates the brain shift in intra-operative ultrasound images.

Tumor resection causes correspondence loss around the resection site and is considered as an outlier during the registration. In [36], they manually segmented the tumor cavity before registration and removed it to increase the performance. However, manual segmentation needs enormous time and effort and is not feasible. In [37,38], Automatic outlier detection was done based on jacobian of normalized cross-correlation (NCC) in different patches of the images. Automatic segmentation of the anatomical structures that keep their correspondences during the surgery was done in [4,39] by utilizing the deep neural networks. Another method that showed an acceptable result in ultrasound-ultrasound registration of neurosurgery is based on SIFT features that are tracked in 3D ultrasound images before and after resection [3].

1.4 Thesis statement

This thesis aims to track the brain shift during tumor resection surgery using intra-operative ultrasound images. Having an end-to-end method that does not require pre-processing and extracting hand-crafted features is beneficial. The other important factor in medical image registration applications is the computational complexity of the method to be feasible for clinical usage. For this purpose, we exploited a non-rigid landmark-based registration. A Siamese network has been used to extract deep features from the input images and enables accurate and fast landmark tracking. The outline of this thesis is presented in 1.5.

1.5 Thesis outline

Chapter 2 presents a robust feature-based registration method using a Siamese network to track the brain shift in intra-operative ultrasound images of neurosurgery. Two outlier suppression algorithms are employed, and the method is evaluated on a public dataset of retroSpective evaluation of cerebral tumors (RESECT) [1] with 3D ultrasound images. The proposed method is compared to the state-of-the-art methods on the same dataset and shows comparable accuracy and computation time results. This work has been published in the International Workshop of Advances in Simplifying Medical UltraSound (ASMUS) [40]

In Chapter 3, we extend our work by assessing the effect of fine-tuning the network on the method's performance. For further evaluation, the BITE public dataset is exploited along with the RESECT dataset to show the ability of the proposed method to work on the images with different acquisition settings. This work will be submitted to the International Journal for Computer Assisted Radiology and Surgery (IJCARS).

Chapter 2

Robust ultrasound-to-ultrasound registration for intra-operative brain shift correction with a Siamese neural network

2.1 Introduction

In brain tumor surgery, soft tissue deformation, or brain shift, can result from many factors, such as gravity and drug administration, and can greatly affect the quality and safety of the procedure. Intra-operative imaging is often used to track brain shift and the surgical progress. In contrast to the high cost and special setups required by intra-operatie magnetic resonance imaging (iMRI), intra-operative ultrasound (iUS) is a cost-effective and portable imaging modality that has gained popularity in the clinic [41]. However, to help account for brain shift to update pre-surgical plans [42] in commonly used surgical navigation system, robust and efficient image registration algorithms are crucial. While so far most of the previous works [21, 43, 44] focus on the alignment of pre-operative MRI and iUS obtained before dura-opening, very few have attempted to correct additional tissue deformation during the procedure, which is also important to ensure clean removal

of any residual tumour and thus increase the patient survival rate [45]. In this scenario, iUS-iUS registration is required, and poses unique challenges from the more commonly seen MRI-iUS alignment. For example, in addition to continuous tissue deformation introduced from gravity and tissues removal, the procedure can also significantly alter image features and reduce image quality in iUS by introducing air bubbles, debris, and blood clots in the surgical site, rendering registration of pre- and post-resection iUS images more challenging.

To tackle the discrepancies of iUS at different surgical stages, an attractive solution is based on matching anatomical landmarks that are consistent between scans [46]. To date, a number of medical image registration algorithms based on automatic landmark detection have been proposed. Lu *et al.* [47] and Urschler *et al.* [48] used segmentation and corner detection for finding the global shape information. Local keypoint selection and feature matching were done using scale-invariant feature transform (SIFT). They applied their method on 2D US images of kidney and thoracic 3D CT images respectively in the application of feature-based non-rigid registration. Machado *et al.* [3] presented an optimal global feature mapping in 3D iUS images in neurosurgery using 3D SIFT-Rank, which showed a large improvement in the alignment. However, the fully automatic aspect of their work can potentially make it sensitive to selection of voxels in and around the tumor, which usually do not match the post-resection scan.

Application of deep learning in medical image registration has rapidly increased during the past few years. Canalini *et al.* [4] proposed segmentation-based registration of 3D iUS images in neurosurgery. They segmented hyperechogenic regions of the brain that keep their correspondence after resection and excluded the resection cavity. An attractive alternative to this approach is to exploit Siamese networks, which require substantially less training data and have shown promising results in tracking tasks in the computer vision field [49–51]. Gomariz *et al.* [34] took advantage of this strength to develop a Siamese network for tracking 2D US images of the liver. Since landmark locations are not expected to change drastically between frames, they utilized a temporal consistency model to weight the similarity map around the previous landmark location.

In this article, we proposed a novel technique based on Siamese networks for landmark tracking in the 3D iUS images. This network was chosen for two main reasons. First, it is an end-to-end learning technique. Second, it works on new domains not seen by the network in the training stage

as the network extracts general features from the inputs that are necessary for comparison [52]. To allow interactivity and flexibility, in our method, template landmarks are first manually selected by clinicians in pre-resection images. Then, with the Siamese network, matching landmarks can be quickly identified after resection starts to continuously track brain shift. Our main contributions are listed below:

1. Using a Siamese network in the application of landmark tracking for iUS-based brain shift correction at different surgical stages (during and after resection).

2. Demonstrating the adoption of the Siamese network from natural images to US volumes without re-training. This suggests its high adaptability to scans from different machines, imaging settings, and anatomies.

3. Employing a 2.5D search scheme for efficient and robust 3D landmark tracking.

4. Fast registration with an iterative re-weighted least squares (IRLS) algorithms to ensure robustness, rendering the method an attractive choice in neurosurgery.

2.2 Methods

To automatically match clinician-defined reference landmarks from pre-resection iUS scans in those after resection starts, we are inspired by object tracking in videos with Siamese neural networks [51], where high-dimensional image features are represented robustly in more efficient form with convolutional neural networks (CNNs). With the automatically identified matching landmark pairs, an affine image alignment is then estimated. Here, we denote the pre-, during- and post-resection US images as iUS_{pre}, iUS_{during} and iUS_{post}, respectively.

2.2.1 Siamese Network

An overview of our fully convolutional Siamese network is shown in Fig. 1. In essence, the network finds an embedding function Φ to extract a representative feature map of the input image. The embedded images then would be passed through a cross-correlation layer as a similarity function, to find the location of the template image inside the search image. It has been shown that this implementation is fully-convolutional based on the search image. That means translation is commutative

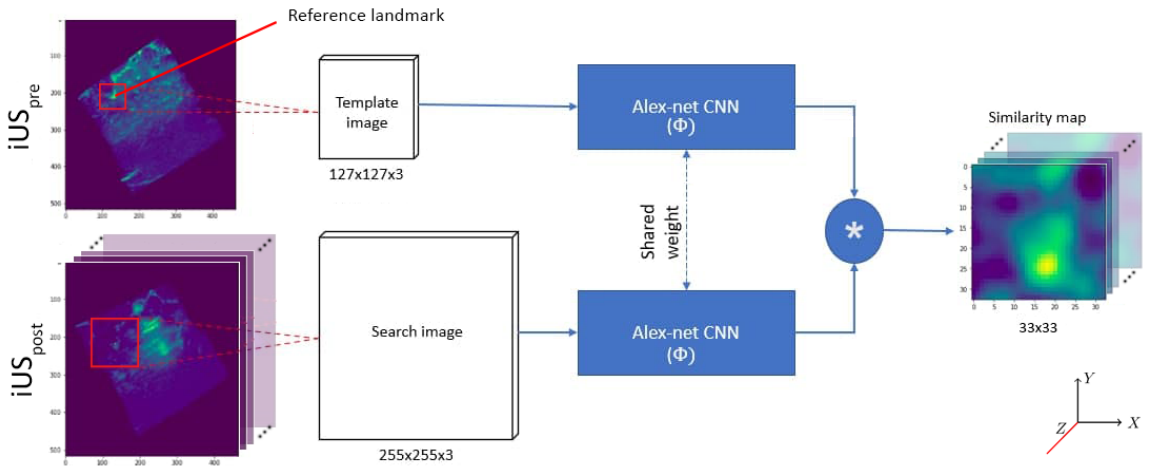

Figure 2.1: Siamese network. The convolutional stage of AlexNet [2] was chosen for feature embedding thanks to its memory efficiency and good performance, and cross-correlation is used as the similarity function. Here, the x-y plane is used for demonstration.

as shown below.

$$\Phi\left(L_\tau x\right) = L_\tau \Phi(x) \tag{1}$$

where L is the translation function with translation τ and x is the search image. The fully-convolutional nature of the network enables us to use a large search image. The search image is divided into sub-windows. These sub-windows are passed through the network. Similarity of the embedded template image and all the translated sub-windows of search image would be evaluated in the cross-correlation layer at once. In our method, we used the positions of the pre-selected landmarks in iUS_{pre} to find their correspondences in iUS_{during} and iUS_{post}.

2.2.2 Training

The network was trained on the ILSVRC17 dataset [53] for video object tracking purposes [54], and no network fine-tuning was performed for our ultrasound application. Stochastic gradient decent (SGD) with binary cross-entropy loss was used, and the learning rate was set to 0.01 with a batch size of 8. The template image size is 127x127 and the search image size is 255x255. For optimization, a binary ground truth (match vs. not a match) was generated by considering a radius

of 4 pixels around the center of the similarity map. If the maximum similarity occurs within the radius, it indicates a match. Otherwise, it is not.

2.2.3 Preprocessing

We used the RESECT database [1] to demonstrate our technique. Voxel sizes in the dataset differ among patients and even scans. As such, all US images were resampled to the smallest isotropic resolution in the database, i.e., $0.14 \times 0.14 \times 0.14mm^3$. In iUS_{pre}, pre-identified anatomical landmarks will be used as references to find the corresponding ones in iUS_{during} or iUS_{post}. In the case of iUS_{pre} to iUS_{post} registration, iUS_{pre} and iUS_{post} will serve as the template and search images, respectively, and the images were cropped around the location of the reference landmark according to the requirement of the network. Finally, all image intensity ranges were normalized to [0,1].

2.2.4 The 2.5D approach

The network we used in this chapter obtains 2D images as inputs. However, our goal is to find landmarks in 3D US images. Similar to the approach of Heinrich *et al.* [55], we performed landmark matching in three orthogonal directions, or 2.5D. In each direction, a region of interest (ROI) around a reference landmark in iUS_{pre} and a larger one in iUS_{post}/iUS_{during} were selected as template and search image respectively. To enable 3D search, the ROI in iUS_{post}/iUS_{during} was swept forward and backward in that direction with a stride of 2, so the search images have one slice overlap.

The location of the maximum correlation in the similarity map was considered as the desired result. Since we had three searching directions by using the 2.5D approach, at the end, we have 3 predictions for each landmark's location in iUS_{post} or iUS_{during}. A final decision was made base on them. If at least two of these three predicted landmarks were located near each other (< 2 mm), the average of them would be considered as the result. Otherwise, the results were treated as incorrect and were discarded.

2.2.5 Affine transformation

Affine transformation has been utilized for tissue shift correction in brain tumour resection thanks to its ability to robustly improve global misalignment. Furthermore, it is simpler and faster in comparison to more complex deformation models, such as free-form B-splines [21]. To estimate a 12-parameter 3D affine transformation, at least 4 pairs of landmarks are required to solve a linear system. Our landmark selection method usually provides at least 5 landmarks (Tables 1 and 2), resulting in an over-determined linear system that can be solved.

In order to obtain the optimal 3D affine transformation while overcoming the potential influence of outlier landmarks, we employed the iterative re-weighted least square (IRLS) method [56]. Here, the Cauchy function (Eq. 5) has been chosen as the weighing function in the IRLS algorithm, where small weights were assigned to the outliers in the linear equation [57] to mitigate their impacts. In Eq. 2, r_i is the residual after an iteration and $R = 1$ was selected manually.

$$w\left(r_i\right) = \frac{1}{1 + \left(r_i / R\right)^2} \tag{2}$$

2.2.6 Experimental setup

We validated our registration method using 17 clinical cases that have pre-, during- and post-resection iUS images in the RESECT public database [1], where matching ground truth landmark pairs have been provided by experts. Quantitative evaluation for our algorithm was performed using mean target registration errors (mTREs) before and after registration using the ground truth landmark pairs. The metric is shown in Eq.6.

$$mTRE = \frac{1}{N} \sum_{i=1}^{N} \left\| T\left(x_i\right) - x_i' \right\|, \tag{3}$$

where x_i and x_i' are the landmark pairs in the corresponding iUS scans, T is the affine transformation estimated with the proposed method, and N is the total number of landmarks. Here, we used the full set of landmark pairs from the original database to compute mTRE. The accuracy of our method was compared against two recent nonlinear techniques [3, 4] that were validated on the same database.

Table 2.1: mTREs of our method and two comparison methods [3, 4] for iUS_{pre} vs. iUS_{post} registration. Initial mTRE before registration and minimum achievable mTRE (with affine transformations) were calculated from the ground truth landmarks. Note that nonlinear registration results can be lower than min achievable mTREs.

Patient ID	No. total land-marks	No. selected land-marks	Initial mTRE (mm)	After Siamese based affine (mm)	Canalini et al. (mm)	Machado et al. (mm)	Affine Mini-mum achiev-able mTRE (mm)
1	13	10	5.80 (3.62-7.22)	1.32 (0.47-4.06)	1.03	1.48	0.97
2	10	5	3.65 (1.71-6.72)	**2.47** (0.45-4.35)	3.90	2.62	1.57
3	11	6	2.91 (1.53-4.30)	1.05 (0.28-1.79)	1.15	1.04	0.67
4	12	9	2.22 (1.25-2.94)	0.81 (0.21-1.43)	0.61	0.83	0.55
6	11	6	2.12 (0.75-3.82)	1.66 (0.29-3.72)	1.41	1.55	1.20
7	18	15	3.62 (1.19-5.93)	**1.73** (0.41-4.39)	2.03	2.38	1.50
12	11	9	3.97 (2.58-6.35)	1.36 (0.25-3.42)	0.79	1.20	0.98
14	17	17	0.63 (0.17-1.76)	0.53 (0.28-0.90)	0.46	0.53	0.45
15	15	15	1.63 (0.62-2.69)	0.72 (0.25-1.84)	0.58	0.74	0.70
16	17	13	3.13 (0.82-5.41)	1.19 (0.42-2.92)	0.92	1.94	1.08
17	11	9	5.71 (4.25-8.03)	1.40 (0.49-4.20)	1.10	1.99	0.96
18	13	6	5.29 (2.94-9.26)	1.29 (0.45-2.90)	1.13	1.69	1.14
19	13	11	2.05 (0.43-3.24)	1.23 (0.32-5.28)	1.10	2.78	0.86
21	9	7	3.35 (2.34-5.64)	1.77 (0.76-3.71)	1.80	1.07	0.76
24	14	11	2.61 (1.96-3.41)	1.02 (0.35-2.54)	0.87	1.35	0.62
25	12	11	7.61 (6.40-10.25)	**1.20** (0.29-2.42)	1.21	1.24	0.91
27	12	11	3.98 (3.09-4.82)	0.63 (0.20-1.06)	0.53	0.83	0.47
mean	13	10	3.55	1.26	1.21	1.49	0.90
std			1.76	0.48	0.81	0.67	0.33

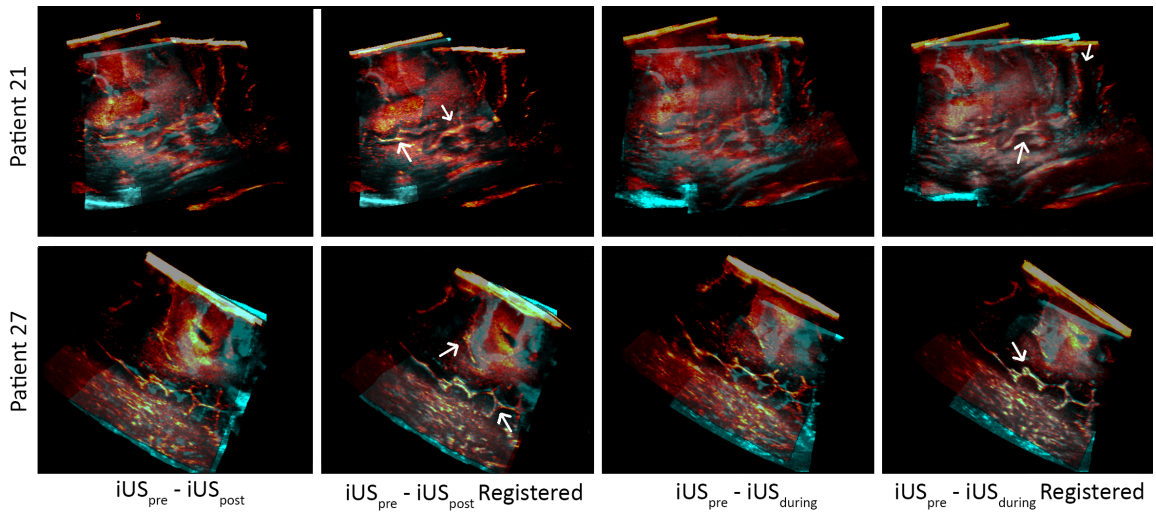

Figure 2.2: Visual comparison between image pairs before and after registration with our proposed method. Cyan color = iUS_{post} and iUS_{during} and orange color = iUS_{pre}. The arrows mark the sulci regions of improved alignment.

Statistical tests were done to confirm the performance of our method.

2.3 Results

Tables 1 and 2 show the quantitative evaluation of mTREs, with the number of landmarks selected by our method to obtain the affine transformations. On average, the tested cases in the RESECT database have 13 and 21 landmarks per patient in iUS_{pre}-iUS_{post} and iUS_{pre}-iUS_{during} cases, respectively. With automatic landmark selection using the Siamese network, we can obtain an average of 10 and 18 landmarks per patient, respectively. With our registration method, the initial misalignment of 3.55±1.76 mm was reduced to 1.26±0.48 mm between before and after resection, and from 3.49±1.56 mm to 1.16±0.49 mm between before and during the resection. Furthermore, the comparison between our method and those by Machado *et al.* [3] and Canalini *et al.* [4] are shown in Tables 1 and 2. Overall, our proposed approach showed very similar performance to that of Canalini *et al.* [39] and on average outperformed the method of Machado *et al.* [3] in iUS_{pre} vs. iUS_{post} registration. In addition, our results are also better in iUS_{pre} vs. iUS_{during} registration than Canalini *et al.* [4] on average (Machado *et al.* [3] didn't perform iUS_{pre} vs. iUS_{during} registration).

Table 2.2: mTREs of our method and a comparison method [4] for iUS_{pre} vs. iUS_{during} registration. Initial mTRE before registration and minimum achievable mTRE (with affine transformations) are calculated based on the ground truth landmarks provided. Note that nonlinear registration results can be lower than min achievable mTREs.

Patient ID	No. total landmarks	No. selected landmarks	Initial mTRE (mm)	After Siamese based affine (mm)	Canalini et al. (mm)	Affine Minimum achievable mTRE (mm)
1	34	33	2.32 (1.49-3.29)	0.88 (0.18-1.90)	0.64	0.83
2	16	11	3.10 (1.79-5.19)	**1.43** (0.33-4.42)	1.50	1.21
3	17	17	1.93 (0.67-3.02)	0.79 (0.26-1.33)	0.77	0.70
4	19	17	4.00 (3.03-5.22)	0.89 (0.30-2.44)	0.80	0.74
6	21	17	5.19 (2.60-7.18)	**1.77** (0.32-3.42)	5.17	1.47
7	22	19	4.69 (0.94-8.16)	2.46 (0.26-6.56)	1.98	1.82
12	24	23	3.39 (1.74-4.81)	1.12 (0.15-2.01)	0.84	1.04
14	22	22	0.71 (0.42-1.59)	0.50 (0.03-0.90)	0.41	0.47
15	21	21	2.04 (0.85-2.84)	0.68 (0.27-1.38)	0.60	0.58
16	19	10	3.19 (1.22-4.53)	1.51 (0.21-4.60)	1.26	1.10
17	17	11	6.32 (4.65-8.07)	1.54 (0.49-3.91)	1.49	0.97
18	23	16	5.06 (1.55-7.44)	1.33 (0.20-3.71)	1.18	1.10
19	21	20	2.06 (0.42-3.40)	1.03 (0.20-2.35)	0.96	0.90
21	18	14	5.10 (3.37-5.94)	1.25 (0.33-3.81)	1.11	0.97
24	21	19	1.76 (1.16-2.65)	0.78 (0.13-1.85)	0.67	0.64
25	20	19	3.60 (2.19-5.02)	0.72 (0.27-2.01)	0.55	0.65
27	16	16	4.93 (3.61-7.01)	0.96 (0.19-2.34)	0.87	0.57
mean	20	18	3.49	1.16	1.22	0.93
std			1.56	0.49	1.10	0.35

We performed Wilcoxon rank sum tests on the mTREs before and after registration with the proposed method, as well as on the same metric to compare between our method and those two recent works [3,4]. The statistical tests showed that the reduction in mTREs with our technique was statistically significant (p <0.001). In addition, in post-resection registration, our results are comparable (p>0.05) to those by Machado *et al.* and Canalini *et al.* [3,4] while our average mTRE reduction is better than that of Machado *et al.* [3]. In during-resection registration, our results were comparable to those of Cananili *et al.* [4](p=0.19), with a better average mTRE. Furthermore, qualitative assessment of our method is illustrated in Fig. 2.2, where pre-resection and during-/post-resection images are overlaid in cases of before and after registration for two patients. Note that anatomical features (e.g., sulci) are shown as hyperintense edges in each image.

2.4 Discussion

In this chapter, we used the 12-parameter affine transformation for brain shift correction during tumour resection. Although nonlinear deformation models, such as B-splines, can more precisely adapt to local tissue deformation, the computational complexity is much higher. In addition, robustness and reliability in intra-operative registration algorithms can be more valuable in the clinic, and thus affine transformation appears more advantageous, considering its mTRE measures are comparable to the non-linear counterparts [3,4].

In the proposed algorithm, reference landmarks for tracking need to be identified first while in previous fiducial point-based registration methods, automatic landmark selections were employed based on image feature detection (e.g., SIFT). The involvement of manual interaction with the image can help ensure the distribution of salient anatomical landmarks for optimal registration quality [58], and improve the flexibility and robustness in real clinical applications. The proposed method showed an excellent performance in reducing the initial misalignment in terms of mTREs. On average, the results are comparable to or better than the recent state-of-the-art techniques [3,4], which were validated also on the RESECT database. For our method,the whole process of landmark matching and registration on NVIDIA GeForce GTX 1050 Ti GPU took about 45 sec in post-resection registration and 70 sec in during-resection registration (due to a greater number of

landmarks). The segmentation-based registration by Canalini *et al.* [4] takes 55 sec on average and the SIFT-based registration by Machado *et al.* [3] is the fastest with 30 sec of average run time. Thus, the computational time of our technique is highly promising in real clinical applications and comparable to the state-of-the-art methods.

One limitation of our proposed method lies in the requirement of reference landmark tagging, which can cost extra clinical time. However, as with affine transformation, the number of landmarks doesn't need to be large, and with the rich experience of the clinicians, this can be performed robustly and quickly, especially considering that a typical neurosurgery lasts a few hours. This also offers more flexibility for manual interaction. Although we present the proposed technique for iUS-iUS registration, we believe that it can be adapted to MR-iUS registration, which will be investigated in the future.

2.5 Conclusion

We have proposed a robust and efficient iUS-iUS registration technique based on anatomical landmark detection to account for tissue shift in brain tumor surgery. The method demonstrated excellent performance compared to the recent works, and can potentially improve the accuracy and safety of the procedure. The Siamese network weights were trained on natural images without domain-specific fine-tuning, rendering the method robust to scanner differences.

Chapter 3

Robust landmark-based brain shift correction with a Siamese neural network in ultrasound-guided brain tumor resection

3.1 Introduction

Pre-operative planning in neurosurgery is facilitated by pre-operative imaging of the brain, with magnetic resonance imaging (MRI) being the main image modality [59]. During the surgery, brain tissue deforms because of several reasons, including gravity, drug administration, brain swelling, etc. This phenomenon is called brain shift. Intra-operative imaging is often used to track the brain shift and update the surgical plan to ensure patient safety and outcomes of surgery [42]. Among the common choices, intra-operative MRI (iMRI) is a high-cost option that, despite its good contrast, requires a dedicated operating room setup with MRI-compatible surgical tools, and thus it is not widely available [60]. Intra-operative ultrasound (iUS), on the other hand, is a cost-effective, portable, and flexible alternative [36, 37].

Most surgical navigation systems utilize rigid transformations to track surgical tools in the operating room and execute the image to patient alignment. However, the brain deforms non-rigidly during the surgery as a result of the brain shift. Most of the works in the literature focus on registering pre-operative MR images to iUS images before dura opening to re-adjust the pre-surgical planning [21, 43, 44]. The fact that the most important factor in patient survival rate is complete resection with no residual tumor [45] suggests the importance of accurate registration techniques to align intra-operative ultrasound images for the procedure.

Registration of iUS images has its unique challenges compared to MRI-iUS registration. During the resection of the tumor, on top of continuous deformation of the brain near the resection site, the quality of the iUS image decays by introducing artifacts like air bubbles, debris, and blood clots around the tumor, making the evaluation of the surgical progress hard especially towards the end of the surgery.

Landmark-based registration is a type of registration technique that is in contrast to the intensity-based methods, which use the full image content (i.e., image intensities) for feature alignment. It only requires a small number of corresponding key points annotated carefully between an image pair. This technique is flexible and efficient while providing user-interaction whenever needed, ensuring the landmark annotation quality and thus the accuracy and robustness of image registration. This is particularly attractive for clinical procedures, such as brain shift correction in iUS-guided brain tumor resection. Previously, Lu *et al.* [47] and Urschler *et al.* [48] both extracted the global structural information using corner detection, and used scale-invariant feature transform (SIFT) for local feature detection and matching. They tested their method on 2D US kidney images and 3D Thoracic computed tomography (CT) images. Other features like gradient magnitude and Laplacian of Gaussian (LOG) were also showed to be useful for landmark detection [24, 25]. Machado *et al.* [3] proposed to use SIFT points for automatic feature detection and matching in 3D iUS images. Their method is fast with good results, although the automatic key point selection makes it prone to choosing landmarks near the tumor site, which can lose their correspondence after resection. To mitigate the impact of tissue resection in image registration, Canalini *et al.* [4] employed a segmentation-base approach on iUS images using deep-learning to automatically segment the

hyperechogenic features of the image (e.g., sulci) and used them as guidance features for the registration. To avoid the tumor area, they automatically extracted it before doing the segmentation. Their results showed large improvement from the initial misalignment.

Deep-learning-based registration methods have attracted a lot of attention during the past a few years due to their high speed. They are also data-driven and can extract relevant deep-features from the input images instead of handcrafted ones that require additional processing steps. Landmark detection has been done using deep neural networks in different medical image modalities, including US images of prostate gland [32], as well as MR and CT images of the brain [33]. In terms of landmark-based image registration, several deep learning-based methods have been developed, especially with the application of the Siamese networks, which were used in real-time object tracking in videos. Gomariz *et al.* [34] utilized a Siamese network in 2D sequential liver US images to find the matching landmarks in different frames. They proposed a temporal consistency model as the landmark locations are not expected to change largely between frames. Grewal *et al.* [35] proposed an end-to-end registration approach on 2D abdominal CT images with a Siamese network where by adding a sampling layer to the network, landmark annotation is done automatically during the registration. In our previous work [40], we leveraged a pre-trained Siamese network that had only been trained on a large set of natural images to detect matching landmarks between 3D iUS scans at different surgical stages to perform iUS-iUS registration for brain shift correction in neurosurgery. Despite the fact that the employed network had not seen any US image before, our results were comparable, or in some cases, better than the state-of-the-art methods evaluated on the same RESECT database [1]. In this work, we further extended our previous work [40] in two directions. First, we investigate the effect of fine-tuning the network for landmark matching with domain-specific data (i.e., iUS) through transfer learning. Second, we thoroughly assess the adaptability and performance of the method by testing it on two public datasets of iUS-guided brain tumor resection, which were collected with different ultrasound scanners and scanning settings (i.e., imaging depth and transducer types) from different cohorts of patients and hospitals.Finally, the resulting technique is compared against the state-of-the-art algorithms.

3.2 Datasets

In this work, two publicly available datasets of iUS-guided brain tumor resection were used, including REtroSpective Evaluation of Cerebral Tumors (RESECT) [1] and Brain Images of Tumors for Evaluation (BITE) [61]. The RESECT dataset contains 23 patients with low-grade gliomas and 17 of them have pre-annotated landmarks for evaluation of iUS-iUS registration. It also contains three stages of the surgery: before the resection starts, during the resection, and after removal of the tumor. From now on in this chapter, we will refer to the iUS scans acquired at these stages as iUS_{pre}, iUS_{during} and iUS_{post} respectively. On the other hand, the BITE dataset contains 13 low-grade and high-grade glioma patients along with tagged landmarks by experts to evaluate iUs-iUS registration. BITE dataset only has iUS_{pre} and iUS_{post} and does not provide iUS_{during} for evaluation. The RESECT database used more recent ultrasound scanners to provide higher quality images than the BITE dataset, and as a result, more annotated landmark pairs are available for registration assessment.

3.3 Methods

Our proposed technique is inspired by the application of the Siamese networks in object tracking tasks proposed by Bertinetto *et al.* [51]. Similar to tracking objects in natural scenes, our proposed method matches hyperechogenic anatomical structures of the brain (e.g., sulci, gyri, and deep grooves) between iUS images at different stages of brain tumor resection.

3.3.1 Siamese Neural Network

Siamese networks were designed to extract deep features from image pairs to compare them for classification purposes [52]. They are usually constructed with two identical branches of convolutional neural networks (CNNs) that share the same weights. In this work, we used the convolutional stage of the AlexNet [2], which was built for natural image classification to extract image features from a template and a search iUS image. These features are compared using a cross-correlation layer at the end of the network. The architecture of the network is illustrated in Fig. 3.1.

This network has been shown to be fully-convolutional [51]. As a result, it enables the usage of

a larger search image compared to the template image. The template image is cropped around the target landmark in the first image, and for the search image, a larger crop as the region of interest around the same location in the second image is done. In the cross-correlation layer, the extracted deep features of all the sub-windows within the search image are simultaneously compared to that of the template image. The sub-windows are acquired with 4 voxels shift within the search image. This is known as the stride of the network. The distance of the corresponding landmark to the target landmark is determined by finding the distance of the maximum correlation to the center of the similarity map, multiplied by the stride.

We assume the spatial shift of landmarks between surgical stages are relatively limited after dura-opening, and thus utilized a constraint function to confine the predicted landmark to be close to the center of the similarity map and penalize large displacement. You can find the constraint function in Eq. 4, in which L is the search range, d is the distance of the maximum correlation to the center, and α = 0.4 was selected as a user-defined parameter.

$$C\left(d\right) = \cos\left(\frac{d \times \cos^{-1}(\alpha)}{L}\right) : d < L \tag{4}$$

In this work, manually annotated landmarks in the iUS_{pre} image are used as templates to match their location in iUS_{during} and iUS_{post} images.

3.3.2 Landmark matching with a 2.5D approach

In video object tracking, an object of interest is tracked frame-to-frame in 2D. To adopt the same network in our 3D registration application, we took advantage of the 2.5D approach that was utilized by Heinrich *et al.* [55]. This enabled us to track 3D objects with conventional 2D CNNs by using three perpendicular directions from the 3D image.

The template and search images were generated by concatenating two adjacent image slices to the target slice to imitate the 3 channels of RGB natural images. The search images were selected with 2 voxels shift in each direction of the 2.5D approach, so each of the search images had one slice in common with its neighbour search image.

In addition, we further implemented an outlier suppression technique for our 2.5D approach.

Specifically, by leveraging three perpendicular directions, three predictions for a single landmark were generated. The final spatial location of the particular landmark was calculated as the average of the two or three predictions that were located close to each other(less than 3 mm). Otherwise, in the case of no adjacent predictions, the landmark matching was not successful, and the operation will move on to the next target landmark in the reference image (i.e., iUS_{pre}).

3.3.3 Landmark-based tissue shift correction

To correct the brain shift, a spatial transformation is required to emulate the brain deformation. Affine transformations have been illustrated to be a reliable option due to its simplicity and high-speed performance in various applications. In contrast to more complex non-linear transformations, affine transformation provides more robust results since they are less susceptible to local-minima convergence in the optimization process [62].

In 3D space, a full affine transformation has 12 degrees of freedom (DOF). By providing four pairs of corresponding points, all the 12 parameters can be estimated. However, the proposed landmark selection algorithm usually provides more than five corresponding landmark pairs that can be used to solve an over-determined system of linear equations.

Since the affine transformation model can render global adjustment to the image, the presence of outliers would mis-guided the registration and potentailly cause fatal errors for brain shift correction in surgery. Therefore, we exploited an iterative re-weighted least square (IRLS) method [56] to find the affine transformation while removing the outliers. Specifically, in each iteration of the optimization, the distance of the estimated landmarks from the ground truth were calculated. The estimated landmarks with larger distance from the ground truth were penalized more, and smaller weights were assigned to them. The Cauchy function was deployed as the weighing function [57] as you can find in Eq. 5 which r_i is the residual of each point in each iteration.

$$w\left(r_i\right) = \frac{1}{1 + \left(r_i\right)^2} \tag{5}$$

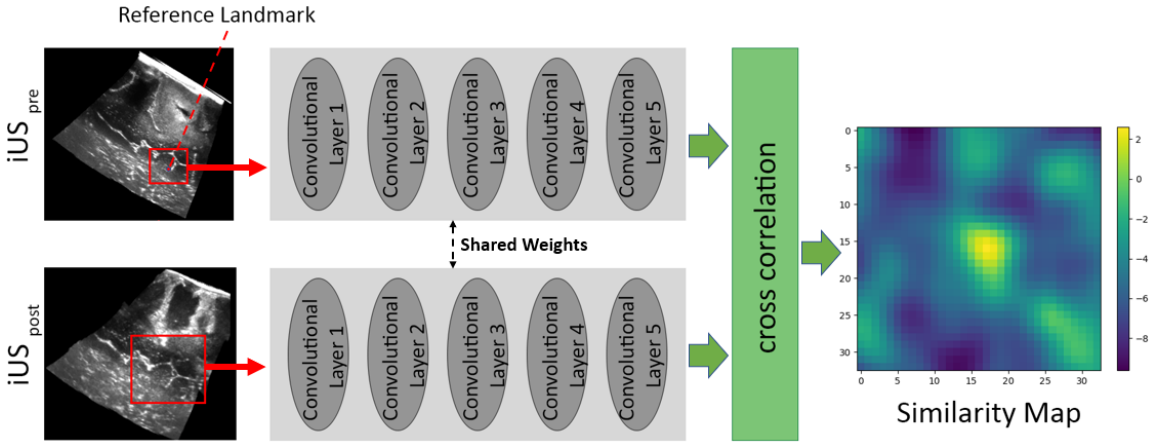

Figure 3.1: The Siamese network architecture.

3.4 Experimental setup

3.4.1 Data preprocessing

Three-dimensional iUS images in the BITE dataset have the resolution of $0.3 \times 0.3 \times 0.3mm^3$ in comparison to those of the RESECT dataset, which are between $0.14 \times 0.14 \times 0.14mm^3$ and $0.24 \times 0.24 \times 0.24mm^3$ due to the use of different ultrasound transducers and imaging depths. To unify the resolution, we resampled all the images to the lowest voxel dimension of $0.14 \times 0.14 \times 0.14mm^3$.

Here, we set the template images to have the size of $127 \times 127 \times 3$ voxels that were cropped around the pre-annotated landmarks in iUS_{pre}. Search images have the size of $255 \times 255 \times 3$ voxels that were selected in iUS_{during} or iUS_{post} as discussed in section 3.3.2 . Eventually, all the images were normalized to be in the range [0,1].

3.4.2 Network fine-tuning

Transfer learning from the pre-trained deep neural networks on natural images to medical images is advantageous. Data collection and labeling, unlike natural imaging, is a difficult task in medical imaging that restricts the available data for training. However, The difference between the modality and quality of the natural and medical images would cause a performance decrease. Transfer learning accompanied with a proper fine-tuning strategy would increase the performance

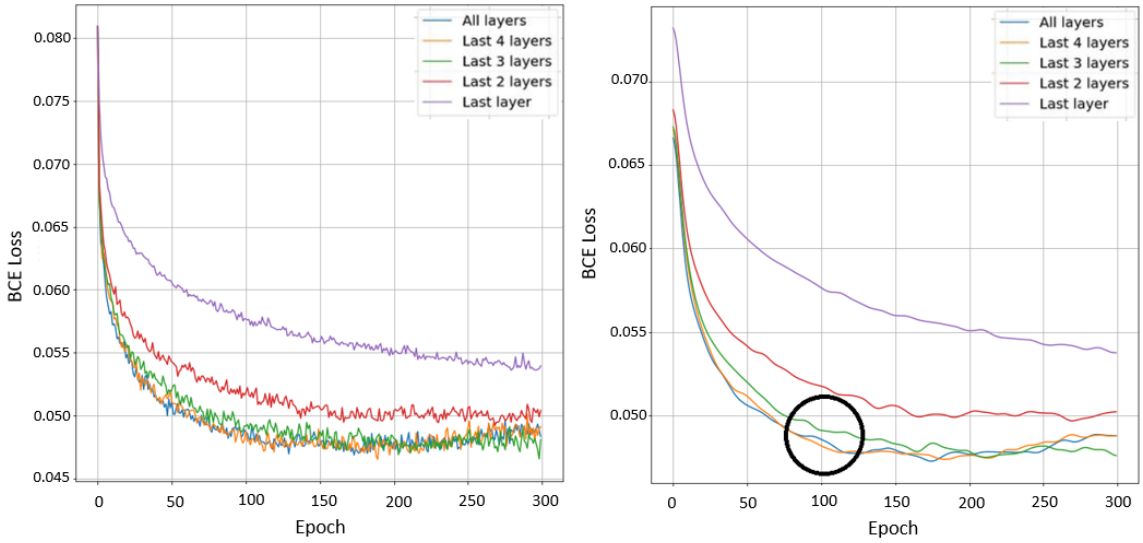

Figure 3.2: Fine-tuning assessment on different layers of the network. The right graph shows the Gaussian smoothed version of the actual graph on the left. The black circle indicates the elbow point where the best model is selected from.

and generality of the network [63].

For our method, a pre-trained Siamese network on the ILSVRC17 dataset [53] of natural images after domain-specific fine-tuning was used. Previously, we demonstrated the generalizability of the pre-trained network by adopting it for the application of iUS-iUS registration without any fine-tuning [40]. However, this work will further investigate the benefit of network fine-tuning through transfer learning on the quality of landmark selection and the final registration accuracy.

The RESECT dataset provides 17 patients with a total of 570 manually annotated landmarks in iUS_{pre} images. The BITE dataset has 13 patients with a total of 130 annotated landmarks. By using all the landmarks in both datasets, 700 landmarks were made available. Since the 2.5D approach requires three pairs of images per landmark (for three perpendicular directions), 2100 image pairs could be employed for fine-tuning. In each image pair, the template image was cropped around a target landmark and the search image was cropped around the ground truth.

For our experiments, a five-fold cross-validation was utilized by splitting the data into the training, validation, and test groups in the proportion of 60%:20%:20%. To avoid data leakage, we ensured that all the image pairs of a patient were present in just one of the three split groups. This

strategy ensures the data in the test set were never seen before by the network in the train or validation set.

For the transfer learning procedure, binary cross-entropy was utilized as a loss function. The similarity map was passed through a sigmoid function to be in the probability range of [0,1]. The loss was calculated for each element of the similarity map by considering the elements near the center (i.e., ground truth) in a radius of 4 pixels to be in a positive class (matched elements) and elements out of the 6 pixels radius to be in the negative class (not-mathched elements). The middle elements between these two radii were considered neutral, so they did not contribute in the optimization. The average loss for all the elements was considered as the total loss for each image pairs. Adam optimization with a learning rate of 5×10^{-5} was used with a batch size of 8. To prevent over-fitting, we used an early stopping strategy in which the elbow point is determined in an epoch that the loss would not decrease more than 0.0005 after 30 epochs. Data-augmentation was employed to increase the number of training data and prevent over-fitting. More specifically, we used random vertical and horizontal flipping and a random rotation in the range of [0-180] degrees in each epoch on the training images.

We used pre-trained weights for the Siamese network for initialization. It was shown that fine-tuning the last few layers and freezing the initial layers would decrease the chance of over-fitting unless the source and target datasets are very different [64, 65]. To explore which of the five layers in the CNN branch to fine-tune, we conducted experiments to perform transfer learning on different ensemble of layers, beginning from just the last layer till all five layers. As the validation curve in Fig. 3.3 shows, fine-tuning all the layers or the last four layers led to better performance. The first layer of the CNN branch usually corresponds to low-level features that are the similar across different datasets [66], so in our final model, the last four layers were used for fine-tuning, and the weights of the first layer were frozen.

3.4.3 Quantitative evaluation criteria

A commonly used metric for quantitative evaluation of image regisration is mean target registration error (mTRE). The mTRE is the average of the Euclidean distances between landmarks of

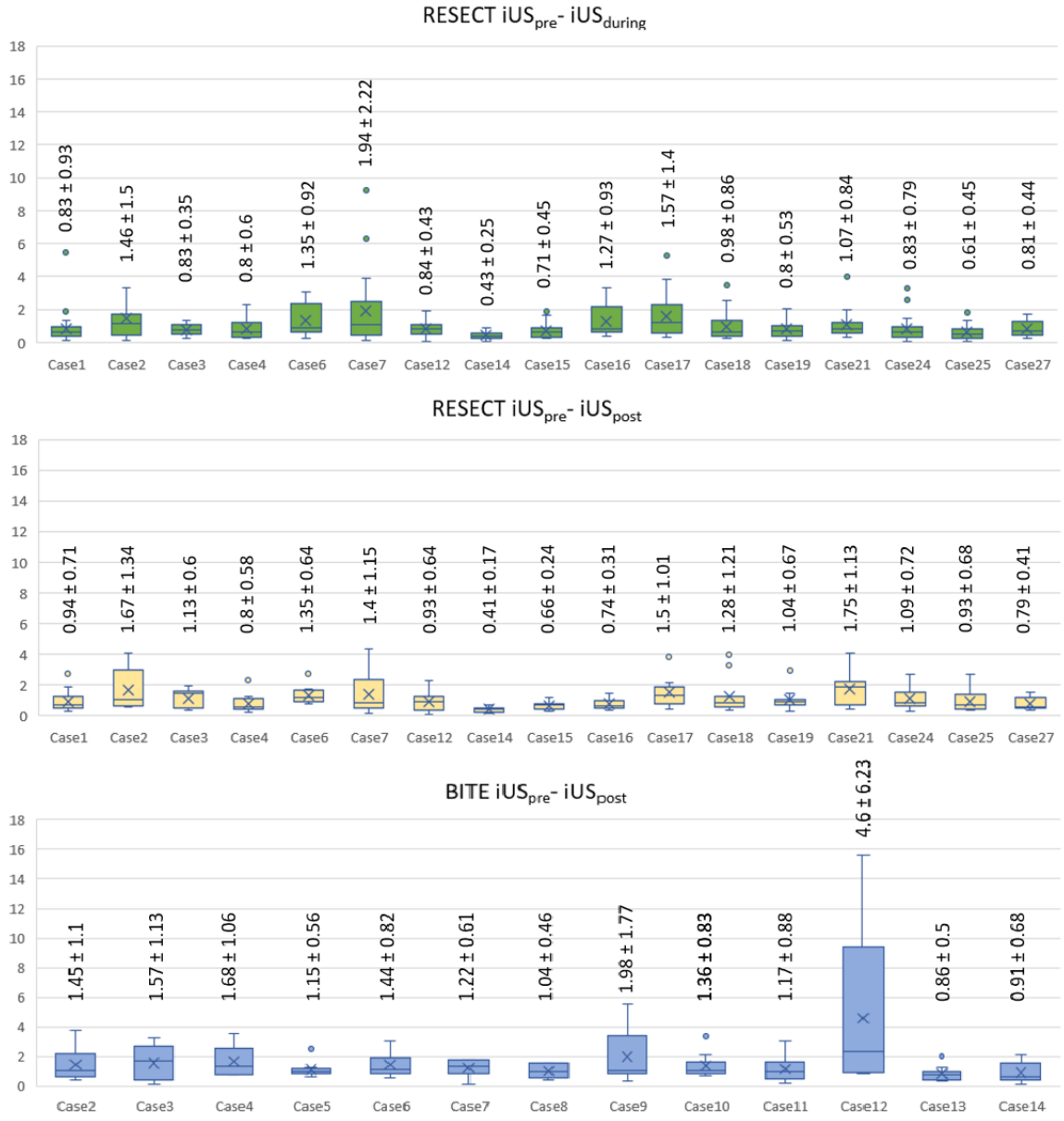

Figure 3.3: Landmark selection error with respect to the available ground truth. The cross sign in each bar represents the average, and the small dots show the outliers. The average error is 1.01±0.39 mm, 1.08±0.37 mm, and 1.57±0.96 mm for the first, the second, and the third row, respectively.

Table 3.1: mTREs of different methods [3,4] in comparison to the proposed method for iUS_{pre} vs. iUS_{post} registration of the BITE dataset.

Patient ID	No. total land-marks	No. selected land-marks	Initial mTRE (mm)	Proposed method (mm)	Canalini et al. (mm)	Machado et al. (mm)	Affine Mini-mum achiev-able mTRE (mm)
2	10	9	2.30 (0.57-5.42)	1.56 (0.47-4.03)	1.70	1.66	0.97
3	10	9	3.40 (0.00-5.09)	1.61 (0.35-3.45)	1.49	1.30	1.16
4	10	8	4.60 (2.96-5.88)	1.47 (0.56-2.83)	5.34	1.13	0.61
5	10	9	4.11 (2.58-5.52)	1.04 (0.42-2.19)	1.17	1.24	0.74
6	10	10	2.26 (1.36-3.10)	1.29 (0.43-2.62)	1.08	1.18	0.76
7	10	8	3.87 (2.60-5.07)	1.29 (0.28-2.14)	1.23	0.86	0.66
8	10	8	2.51 (0.67-3.93)	1.26 (0.36-2.93)	1.21	1.12	0.85
9	10	9	2.21 (1.00-4.59)	1.73 (0.11-4.7)	1.57	1.38	0.71
10	10	10	3.87 (0.98-6.68)	1.4 (0.3-2.91)	1.18	2.59	1.26
11	10	9	2.74 (0.44-8.22)	1.59 (0.32-3.93)	2.29	2.52	1.34
12	10	5	10.54 (7.85-13.04)	6.62 (1.08-15.15)	10.79	2.37	0.97
13	10	10	1.62 (1.33-2.21)	0.83 (0.54-1.47)	0.71	1.01	0.56
14	10	10	2.19 (0.59-3.99)	1.22 (0.34-2.05)	1.17	1.04	1.01
mean	10.00	8.77	3.55	1.76	2.38	1.49	0.89
std			2.29	1.48	2.78	0.60	0.25

the fixed image and transformed landmarks of the moving image as demonstrated in Eq. 6

$$mTRE = \frac{1}{N}\sum_{i=1}^{N}\left\|T\left(x_i\right) - x_i'\right\|, \tag{6}$$

where x_i' and x_i represent landmarks of the fixed and moving images, respectively. T is the transformation, and N shows the total number of landmarks in each case.

3.5 Results

3.5.1 Landmark selection accuracy

Figure 3.3 shows the landmark selection accuracy in all the cases in both public datasets. We used pre-annotated landmarks provided in iUS_{pre} images to track them after the resection starts in iUS_{during} and iUS_{post} images. The errors were obtained by computing the distance between the predicted landmarks and the ground truth. Overall, our landmark selection method could find the corresponding landmarks with the accuracy of 1.08±0.37 mm and 1.57±0.96 mm for iUS_{post} tracking in RESECT and BITE databases, respectively. In the case of iUS_{during} landmark localization, our method achieved an accuracy of 1.01±0.39 mm. By minimizing the landmark selection error, a more accurate registration result is obtainable.

3.5.2 Image registration accuracy

Quantitative assessment of the registration is shown in Tables 3.1, 3.2 and 3.3. We compared our method with two state-of-the-art methods [3, 4]. For our technique, selected landmarks from the fine-tuned Siamese network were used to obtain the affine transformation, and we used the total number of landmarks to calculate mTRE. In addition, the minimum achievable registration results were also calculated by using the ground truth landmark pairs. In summary, our method could reduce the mTRE from 3.55±2.29 mm to 1.76±1.48 mm in the BITE dataset for iUS_{pre} to iUS_{post} registration. In the RESECT dataset, we could decrease the mTRE from 3.55±1.76 mm to 1.22±0.46 mm and 3.49±1.56 mm to 1.11±0.43 mm in iUS_{pre} to iUS_{post} registration and iUS_{pre} to iUS_{during} registration, respectively.

Table 3.2: mTREs of different methods [3, 4] in comparison to the proposed method for iUS_{pre} vs. iUS_{post} registration of the RESECT dataset.

Patient ID	No. total land-marks	No. selected land-marks	Initial mTRE (mm)	Proposed method (mm)	Canalini et al. (mm)	Machado et al. (mm)	Affine Mini-mum achiev-able mTRE (mm)
1	13	13	5.80 (3.62-7.22)	1.1 (0.43-2.82)	1.03	1.48	0.97
2	10	7	3.65 (1.71-6.72)	2.23 (0.57-5.05)	3.90	2.62	1.57
3	11	9	2.91 (1.53-4.30)	1.2 (0.48-2.21)	1.15	1.04	0.67
4	12	12	2.22 (1.25-2.94)	0.59 (0.08-1.11)	0.61	0.83	0.55
6	11	8	2.12 (0.75-3.82)	1.44 (0.68-2.95)	1.41	1.55	1.20
7	18	15	3.62 (1.19-5.93)	1.6 (0.52-4.87)	2.03	2.38	1.50
12	11	11	3.97 (2.58-6.35)	1.28 (0.47-4.17)	0.79	1.20	0.98
14	17	17	0.63 (0.17-1.76)	0.5 (0.23-0.86)	0.46	0.53	0.45
15	15	15	1.63 (0.62-2.69)	0.77 (0.36-2.16)	0.58	0.74	0.70
16	17	14	3.13 (0.82-5.41)	1.18 (0.41-2.47)	0.92	1.94	1.08
17	11	9	5.71 (4.25-8.03)	1.47 (0.55-4.06)	1.10	1.99	0.96
18	13	11	5.29 (2.94-9.26)	1.36 (0.45-3.27)	1.13	1.69	1.14
19	13	12	2.05 (0.43-3.24)	1.45 (0.37-5.06)	1.10	2.78	0.86
21	9	9	3.35 (2.34-5.64)	1.85 (0.85-4.03)	1.80	1.07	0.76
24	14	14	2.61 (1.96-3.41)	0.89 (0.09-2.61)	0.87	1.35	0.62
25	12	12	7.61 (6.40-10.25)	1.09 (0.34-1.96)	1.21	1.24	0.91
27	12	12	3.98 (3.09-4.82)	0.64 (0.19-1.21)	0.53	0.83	0.47
mean	12.88	11.76	3.55	1.22	1.21	1.49	0.90
std			1.76	0.46	0.81	0.67	0.33

The readers can find the comparison of our method compared to other two methods in Fig. 3.5. In terms of the mean values, in the BITE dataset, our proposed technique outperformed the method of Canalini *et al.* [4], but the method of Machado *et al.* [3] showed a better performance. However, in the RESECT dataset, our method presented a comparable result to Canalini *et al.* [4] method, and a better result than Machado *et al.* [3]. The proposed method also demonstrated more robust results by achieving a lower standard deviation in the average mTREs except in the case of the BITE dataset that the method of Machado *et al.* [3] showed more robust results.

To confirm the observation, non-parametric Wilcoxon rank-sum test was deployed to further evaluate the performance of our method in reducing the target registration error from the initial misalignment. We also utilized the same test to compare our method against the two other recent works as well [3, 4]. From the tests, our method showed a significant reduction from the initial mTRE (p <0.001). In the case of iUS_{pre} to iUS_{during} registration our method showed comparable result with Canalini *et al.* [4] (p=0.41), although we achieved a lower mTRE on average. In the iUS_{pre} to iUS_{post} registration, our proposed method presented comparable results with the two other methods for both the RESECT and BITE datasets (p >0.1). However, the average reduction in the mTRE by our method is better than the method by Canalini *et al.* [4] in the BITE dataset. On the other hand, in the RESECT dataset, the average target error of our method is lower than that of Machado *et al.* [3], and comparable to Canalini *et al.* [4] method.

Fig. 3.4 illustrates the qualitative evaluation of the method. For easier interpretation, The improved regions in the registered images are shown by arrows.

3.5.3 Impact of transfer learning

To better examine the benefit of the transfer learning, we compared the mTREs and landmark selection error for the proposed system with and without transfer learning in Table 3.4 and 3.5, respectively. As we can see, network fine-tuning improved the performance of the method on average. However, the mTRE results without fine-tuning were already close to the minimum achievable results by the affine transformation, so we did not expect a significant improvement by utilizing the fine-tuning. With the Wilcoxon rank-sum test, we compared the mTREs obtained with the two different setups, and their differences were not statistically significant ($p > 0.5$) This is largely due

Table 3.3: mTREs of a method [4] in comparison to the proposed method for iUS_{pre} vs. iUS_{during} registration of the RESECT dataset.

Patient ID	No. total land-marks	No. selected land-marks	Initial mTRE (mm)	Proposed method (mm)	Canalini et al. (mm)	Affine Mini-mum achiev-able mTRE (mm)
1	34	34	2.32 (1.49-3.29)	0.87 (0.15-2.82)	0.64	0.83
2	16	14	3.1 (1.79-5.19)	1.61 (0.26-5.05)	1.50	1.21
3	17	17	1.93 (0.67-3.02)	0.79 (0.29-2.21)	0.77	0.70
4	19	19	4 (3.03-5.22)	0.88 (0.12-1.11)	0.80	0.74
6	21	20	5.19 (2.6-7.18)	1.57 (0.46-2.95)	5.17	1.47
7	22	22	4.69 (0.94-8.16)	2.2 (0.21-4.87)	1.98	1.82
12	24	24	3.39 (1.74-4.81)	1.15 (0.33-4.17)	0.84	1.04
14	22	22	0.71 (0.42-1.59)	0.51 (0.13-0.86)	0.41	0.47
15	21	21	2.04 (0.85-2.84)	0.67 (0.21-2.16)	0.60	0.58
16	19	19	3.19 (1.22-4.53)	1.31 (0.62-2.47)	1.26	1.10
17	17	15	6.32 (4.65-8.07)	1.52 (0.62-4.06)	1.49	0.97
18	23	19	5.06 (1.55-7.44)	1.28 (0.22-3.27)	1.18	1.10
19	21	20	2.06 (0.42-3.4)	1 (0.27-5.06)	0.96	0.90
21	18	18	5.1 (3.37-5.94)	1.09 (0.35-4.03)	1.11	0.97
24	21	21	1.76 (1.16-2.65)	0.86 (0.2-2.61)	0.67	0.64
25	20	20	3.6 (2.19-5.02)	0.71 (0.27-1.96)	0.55	0.65
27	16	16	4.93 (3.61-7.01)	0.84 (0.19-1.21)	0.87	0.57
mean	20.65	20.06	3.49	1.11	1.22	0.93
std			1.56	0.43	1.10	0.35

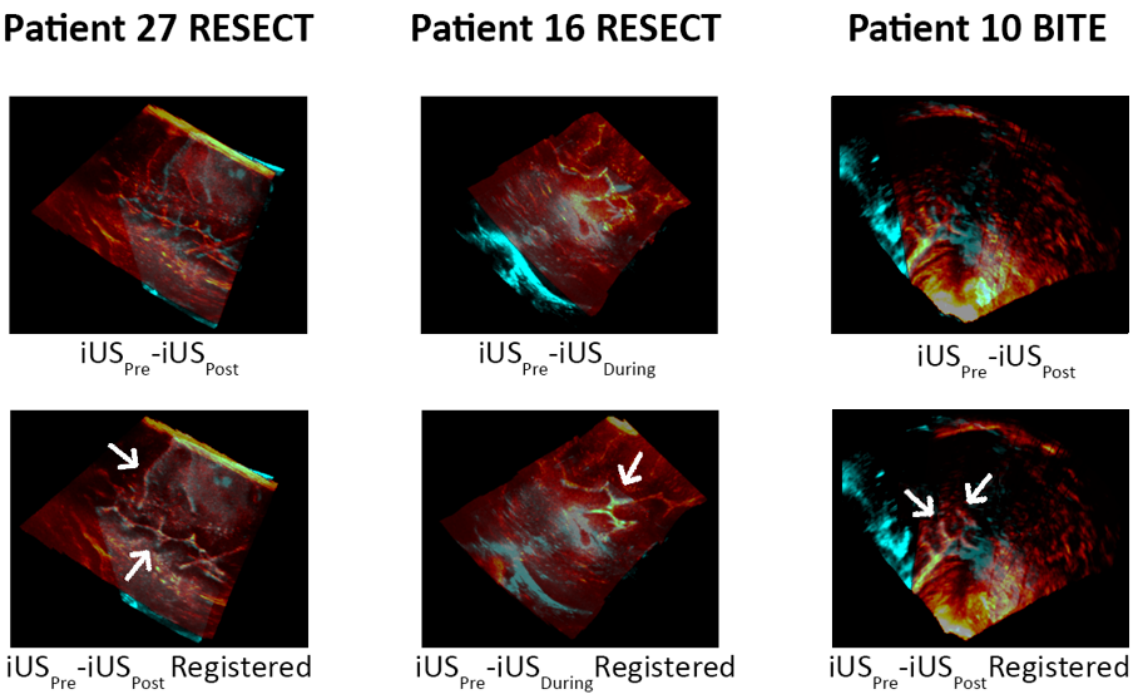

Figure 3.4: Qualitative evaluation of the method. The cyan color image represents iUS_{post} in the first and the third column and iUS_{during} in the second column. The red color image represents the iUS_{pre} before registration and after registration in the first and second row, respectively.

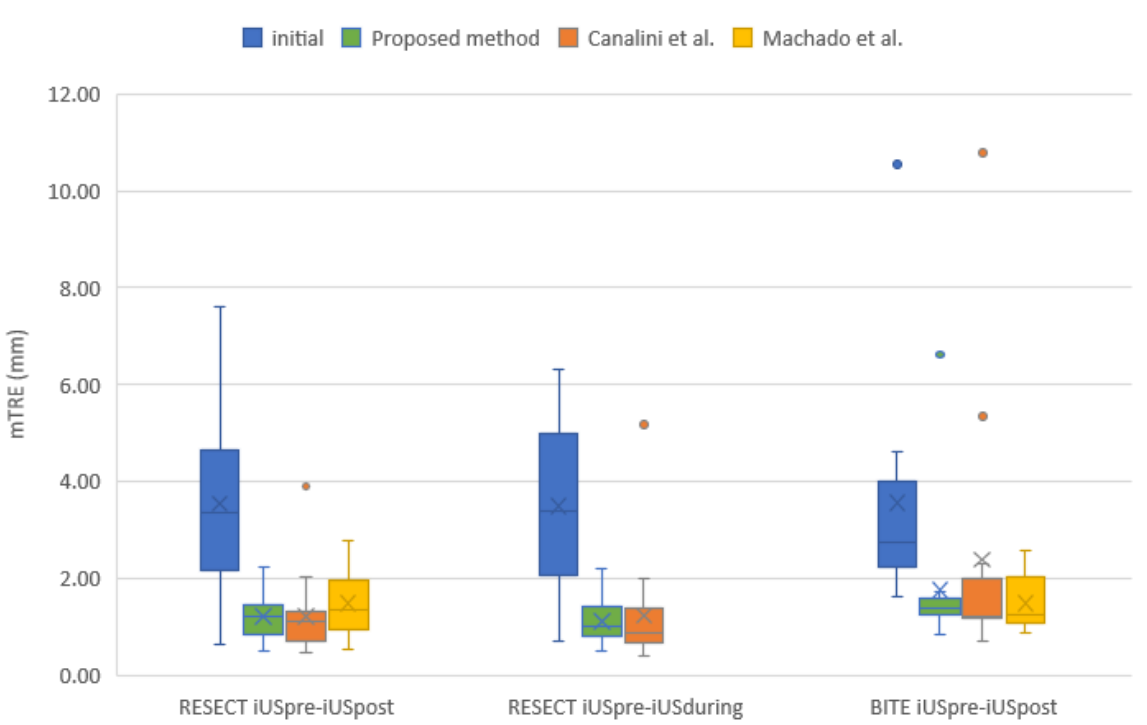

Figure 3.5: Overal comparison between different methods on BITE and RESECT datasets. The cross signs represent the mean values and the small dots demonstrate the outliers in the distributions.

to the spatial transformation model that we employed.

3.6 Discussion

In this work, the affine transformation was employed instead of non-linear counterparts because it is faster and more robust, which is favorable in the clinical application, where reliability is highly valued. Although the compared methods [3, 4] took advantage of non-linear thin-plate spline (TPS) transformation, our proposed method with affine transformation achieved comparable or better results on average. In addition, we used a 2.5D approach with an outlier detection strategy to localize matching anatomical landmarks. Compared with 3D patch approaches for the same goal, our approach is more computationally efficient and allows us to leverage a large quantity of natural images to pre-train the target networks to address the common issue of limited databases for clinical scans.

We validated our method with two datasets with distinct image characteristics due to different choices of scanners, scanner settings, and patients and showed its high adaptability to work on

Table 3.4: The average mTREs of different databases in millimeters, with and without fine tuning. The fine tuning was done with all the data available in both databases.

	With fine tuning	without fine tuning	p-value
RESECT iUS_{pre}/iUS_{post}	**1.22±0.46**	1.27±0.48	0.82
RESECT iUS_{pre}/iUS_{during}	**1.11±0.43**	1.14±0.47	0.82
BITE iUS_{pre}/iUS_{post}	**1.76±1.48**	1.79±1.54	0.77

Table 3.5: The average landmark selection error of different databases in millimeters, with and without fine tuning. The fine tuning was done with all the data available in both databases.

	With fine tuning	without fine tuning	p-value
RESECT iUS_{pre}/iUS_{post}	**1.08±0.37**	1.13±0.36	0.64
RESECT iUS_{pre}/iUS_{during}	1.01±0.39	**0.99±0.37**	0.88
BITE iUS_{pre}/iUS_{post}	**1.57±0.96**	1.72±1.29	0.85

images acquired with different setups and quality. On average, the technique's performance is better in the RESECT dataset as it provides higher image quality and has more landmarks as well.

To further adapt the pre-trained networks for domain specific applications, we performed transfer learning with iUS images. Through our experiments, we decided that for the application, it is beneficial to fine-tune the last four layers of the CNN branch while keeping the first frozen. This demonstrate that there is still relatively large discrepancy between features in natural images and ultrasound scans. Although this results imply that domain-specific fine-tuning is necessary, the improvement in terms of registration accuracy is limited. This is partially due to the fact that we used an affine transformation model. Based on the experimental results, depending on the accuracy requirement of the application, the proposed framework can be deployed with a slight fine-tuning with application-specific data or even without the procedure. In terms of registration speed, the total landmark selection and registration took about 50 seconds on a NVIDIA GeForce GTX 1050 Ti GPU. In comparison to the segmentation-based method of Canalini *et al.* [4], and the SIFT-based method of Machado *et al.* [3] that reported 55 and 30 seconds on average, the computational time of the proposed method is promising.

There are still some limitations to our method. First, to achieve the registration, it requires manual landmark annotation in iUS_{pre} images as a reference, and is classified as a semi-automatic image alignment technique. Although the manual landmark tagging adds extra time to the process,

the user-interaction can ensure the optimal landmark annotation in salient regions of the image to increase the flexibility and robustness of the method [58]. The extra time added by manual landmark tagging is also negligible compared to a typical resection surgery that generally takes several hours. In our future work, we will seek techniques to allow automatic landmark annotation in reference images that captures salient key points that contain good anatomical significance and spatial distribution. Second, although we have performed method validation on two existing public databases, it will be ideal to further examine the impact of increased anatomical and image feature variability with more patients on the improvement of transfer learning and registration accuracy. However, this is currently confounded by the data availability. Finally, we will further extend our work for inter-modal landmark matching, such as in the case of pre-operative MRI vs. iUS registration.

3.7 Conclusion

This chapter proposed a robust landmark-based registration method that can effectively correct the brain shift during resection surgery. The method demonstrated a comparable performance related to other state-of-the-art methods. We also assessed the impact of transfer learning for our proposed network on the registration results and showed high adaptability of the Siamese neural networks in different domains. Finally, the low computational time and high accuracy of our proposed method makes it a robust choice for the relevant potential clinical applications.

3.8 Acknowledgement

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Chapter 4

Conclusion and Future Work

4.1 Conclusion

In this thesis, a robust feature-based registration method for brain shift correction in neurosurgery has been proposed. A Siamese network architecture was used with convolutional stage of AlexNet as a feature extractor and cross-correlation as a similarity metric. Although the previous works on medical image registration with the Siamese networks [34, 35] showed promising result, they usually suffer from some limitations. They have utilized 2D images and assumed small deformation in sequential images of liver and abdomen. The 2.5D technique enabled the method to be used on 3D images. However, 3D models could be exploited but the 2.5D approach is relatively faster and more memory efficient. By selecting the corresponding parts of the image like sulci, small deformation assumption is no longer required and the method could track the large deformation during the tumor resection surgery.

The 2.5D technique followed by an IRLS algorithm provided a two-level outlier suppression block that helped the method achieve more robust result. the proposed method was compared to two other state of the are methods [3,4] tested on the same datasets and on average showed comparable or better results. Although the compared methods utilized automatic landmark selection, the proposed method by manual landmark annotation in iUS_{pre} enabled more robust distribution of the landmarks in salient parts of the image and gave more flexibility to the process. In terms of the computational time, the proposed method provides similar results compared to the other methods.

We performed the fine-tuning assessment by an intensive experiment to find the best tuning layer and learning rate. The Siamese network presented an excellent performance in domain adaptability. However, utilizing fine-tuning could make some improvement in general.

In conclusion, the proposed method demonstrated a significant result in brain shift correction in favorable time duration. We showed that the method could achieve perfect results even without any further fine-tuning, which implies the ability of the model to work on the images collected with different acquisition settings.

4.2 Future Work

In this work, we investigated the application of landmark tracking with a Siamese neural network in US-US registration. Although the method showed an acceptable computation time compared to the actual duration of a resection surgery, it can be optimized further by a parallel implementation.

Decreasing the stride of the network by adding decoding layers on the extracted features would help the method to achieve more accurate results. However, this technique will be time-consuming and need to be thoroughly investigated in the future.

Manual landmark annotation gives flexibility to the clinician, but having a robust automatic landmark detection method would make the process faster. Adding an automatic landmark detection algorithm based on deep learning methods could provide more landmarks in less time. More properly distributed landmarks in the image enable more complex transformations to be used, such as non-linear free-form B-spline deformation. However, in this work, we demonstrated the ability of the affine transformation to track the brain shift with comparable and better results than its non-linear counterparts.

Domain adaptability of the Siamese network has been investigated in this work. Assessment of the method on the other uni-modality registration applications like MRI to MRI registration would be interesting. We believe the method can be extended to multi-modality applications by performing some modifications to the model.

Limited training data and available ground truth in medical imaging restrict the application of

deep learning. Utilizing unsupervised learning methods would help the method to use more training data without any grand truth requirement. Another desirable modification is achievable by using explainable deep learning methods. One of the drawbacks of deep learning methods in healthcare arises from their inability to justify their results, where explainable models can be utilized to shed light on inner workings of the deep models [67].

References

[1] Yiming Xiao, Maryse Fortin, Geirmund Unsgård, Hassan Rivaz, and Ingerid Reinertsen. Retrospective evaluation of cerebral tumors (resect): A clinical database of pre-operative mri and intra-operative ultrasound in low-grade glioma surgeries. *Medical physics*, 44(7):3875–3882, 2017.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[3] Inês Machado, Matthew Toews, Jie Luo, Prashin Unadkat, Walid Essayed, Elizabeth George, Pedro Teodoro, Herculano Carvalho, Jorge Martins, Polina Golland, et al. Non-rigid registration of 3d ultrasound for neurosurgery using automatic feature detection and matching. *International journal of computer assisted radiology and surgery*, 13(10):1525–1538, 2018.

[4] Luca Canalini, Jan Klein, Dorothea Miller, and Ron Kikinis. Enhanced registration of ultrasound volumes by segmentation of resection cavity in neurosurgical procedures. *International Journal of Computer Assisted Radiology and Surgery*, 15(12):1963–1974, 2020.

[5] Douglas L Miller, Nadine B Smith, Michael R Bailey, Gregory J Czarnota, Kullervo Hynynen, Inder Raj S Makin, and Bioeffects Committee of the American Institute of Ultrasound in Medicine. Overview of therapeutic ultrasound applications and safety considerations. *Journal of ultrasound in medicine*, 31(4):623–634, 2012.

[6] Timothy J Mason. Therapeutic ultrasound an overview. *Ultrasonics sonochemistry*, 18(4):847–852, 2011.

[7] Vincent Chan and Anahi Perlas. Basics of ultrasound imaging. In *Atlas of ultrasound-guided procedures in interventional pain management*, pages 13–19. Springer, 2011.

[8] Zara Vajihi, Ivan M Rosado-Mendez, Timothy J Hall, and Hassan Rivaz. Low variance estimation of backscatter quantitative ultrasound parameters using dynamic programming. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 65(11):2042–2053, 2018.

[9] Robert F Wagner and David G Brown. Unified snr analysis of medical imaging systems. *Physics in Medicine & Biology*, 30(6):489, 1985.

[10] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

[11] JB Antoine Maintz and Max A Viergever. An overview of medical image registration methods: Imaging science department. *Imaging Center Utrecht*, 1996.

[12] Bernd Fischer and Jan Modersitzki. Ill-posed medicine—an introduction to image registration. *Inverse Problems*, 24(3):034008, 2008.

[13] Alex Pappachen James and Belur V Dasarathy. Medical image fusion: A survey of the state of the art. *Information fusion*, 19:4–19, 2014.

[14] Gaurav Bhatnagar, QM Jonathan Wu, and Zheng Liu. A new contrast based multimodal medical image fusion framework. *Neurocomputing*, 157:143–152, 2015.

[15] Frank Sauer. Image registration: enabling technology for image guided surgery and therapy. In *2005 IEEE engineering in medicine and biology 27th annual conference*, pages 7242–7245. IEEE, 2006.

[16] Diane M Muratore, Jeannette Herring Russ, Benoit M Dawant, and Robert L Galloway. Three-dimensional image registration of phantom vertebrae for image-guided surgery: a preliminary study. *Computer Aided Surgery*, 7(6):342–352, 2002.

[17] Fakhre Alam, Sami Ur Rahman, Sehat Ullah, and Kamal Gulati. Medical image registration in image guided surgery: Issues, challenges and research opportunities. *Biocybernetics and Biomedical Engineering*, 38(1):71–89, 2018.

[18] Barbara Zitova. Mathematical approaches for medical image registration. 2019.

[19] Shao-Ya Guan, Tian-Miao Wang, Cai Meng, and Jun-Chen Wang. A review of point feature based medical image registration. *Chinese Journal of Mechanical Engineering*, 31(1):1–16, 2018.

[20] Stefan Matl, Richard Brosig, Maximilian Baust, Nassir Navab, and Stefanie Demirci. Vascular image registration techniques: A living review. *Medical image analysis*, 35:1–17, 2017.

[21] Nima Masoumi, Yiming Xiao, and Hassan Rivaz. Arena: Inter-modality affine registration using evolutionary strategy. *International journal of computer assisted radiology and surgery*, 14(3):441–450, 2019.

[22] Dinggang Shen and Christos Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE transactions on medical imaging*, 21(11):1421–1439, 2002.

[23] Dinggang Shen. Image registration by hierarchical matching of local spatial intensity histograms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 582–590. Springer, 2004.

[24] Pezhman Foroughi, Purang Abolmaesumi, and Keyvan Hashtrudi-Zaad. Intra-subject elastic registration of 3d ultrasound images. *Medical image analysis*, 10(5):713–725, 2006.

[25] Robert J Schneider, Douglas P Perrin, Nikolay V Vasilyev, Gerald R Marx, J Pedro, and Robert D Howe. Real-time image-based rigid registration of three-dimensional ultrasound. *Medical image analysis*, 16(2):402–414, 2012.

[26] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[27] Jørn Bersvendsen, Matthew Toews, Adriyana Danudibroto, William M Wells III, Stig Urheim, Raúl San José Estépar, and Eigil Samset. Robust spatio-temporal registration of 4d cardiac ultrasound sequences. In *Medical Imaging 2016: Ultrasonic Imaging and Tomography*, volume 9790, page 97900F. International Society for Optics and Photonics, 2016.

[28] Dong Ni, Yingge Qu, Xuan Yang, Yim Pan Chui, Tien-Tsin Wong, Simon SM Ho, and Pheng Ann Heng. Volumetric ultrasound panorama based on 3d sift. In *International conference on medical image computing and computer-assisted intervention*, pages 52–60. Springer, 2008.

[29] Ingerid Reinertsen, Frank Lindseth, Christian Askeland, Daniel Høyer Iversen, and Geirmund Unsgård. Intra-operative correction of brain-shift. *Acta neurochirurgica*, 156(7):1301–1310, 2014.

[30] Guorong Wu, Minjeong Kim, Qian Wang, Brent C Munsell, and Dinggang Shen. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE transactions on biomedical engineering*, 63(7):1505–1516, 2015.

[31] Guofang Xiao, J Michael Brady, J Alison Noble, Michael Burcher, and Ruth English. Nonrigid registration of 3-d free-hand ultrasound images of the breast. *IEEE Transactions on medical imaging*, 21(4):405–412, 2002.

[32] Ahmet Tuysuzoglu, Jeremy Tan, Kareem Eissa, Atilla P Kiraly, Mamadou Diallo, and Ali Kamen. Deep adversarial context-aware landmark detection for ultrasound imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 151–158. Springer, 2018.

[33] Jun Zhang, Mingxia Liu, and Dinggang Shen. Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. *IEEE Transactions on Image Processing*, 26(10):4753–4764, 2017.

[34] Alvaro Gomariz, Weiye Li, Ece Ozkan, Christine Tanner, and Orcun Goksel. Siamese networks with location prior for landmark tracking in liver ultrasound sequences. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1757–1760. IEEE, 2019.

[35] Monika Grewal, Timo M Deist, Jan Wiersma, Peter AN Bosman, and Tanja Alderliesten. An end-to-end deep learning approach for landmark detection and matching in medical images.

In *Medical Imaging 2020: Image Processing*, volume 11313, page 1131328. International Society for Optics and Photonics, 2020.

[36] Laurence Mercier, David Araujo, Claire Haegelen, Rolando F Del Maestro, Kevin Petrecca, and D Louis Collins. Registering pre-and postresection 3-dimensional ultrasound for improved visualization of residual brain tumor. *Ultrasound in medicine & biology*, 39(1):16–29, 2013.

[37] Hassan Rivaz and D Louis Collins. Near real-time robust non-rigid registration of volumetric ultrasound images for neurosurgery. *Ultrasound in medicine & biology*, 41(2):574–587, 2015.

[38] Hang Zhou and Hassan Rivaz. Registration of pre-and postresection ultrasound volumes with noncorresponding regions in neurosurgery. *IEEE journal of biomedical and health informatics*, 20(5):1240–1249, 2016.

[39] Luca Canalini, Jan Klein, Dorothea Miller, and Ron Kikinis. Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery. *International journal of computer assisted radiology and surgery*, 14(10):1697–1713, 2019.

[40] Amir Pirhadi, Hassan Rivaz, M Omair Ahmad, and Yiming Xiao. Robust ultrasound-to-ultrasound registration for intra-operative brain shift correction with a siamese neural network. In *International Workshop on Advances in Simplifying Medical Ultrasound*, pages 85–95. Springer, 2021.

[41] G Unsgaard, OM Rygh, T Selbekk, TB Müller, F Kolstad, F Lindseth, and TA Nagelhus Hernes. Intra-operative 3d ultrasound in neurosurgery. *Acta neurochirurgica*, 148(3):235–253, 2006.

[42] Yiming Xiao, Live Eikenes, Ingerid Reinertsen, and Hassan Rivaz. Nonlinear deformation of tractography in ultrasound-guided low-grade gliomas resection. *International journal of computer assisted radiology and surgery*, 13(3):457–467, 2018.

[43] Wolfgang Wein. Brain-shift correction with image-based registration and landmark accuracy evaluation. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*, pages 146–151. Springer, 2018.

[44] Mattias P Heinrich. Intra-operative ultrasound to mri fusion with a public multimodal discrete registration tool. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*, pages 159–164. Springer, 2018.

[45] Nicholas F Marko, Robert J Weil, Jason L Schroeder, Frederick F Lang, Dima Suki, and Raymond E Sawaya. Extent of resection of glioblastoma revisited: personalized survival modeling facilitates more accurate survival prediction and supports a maximum-safe-resection approach to surgery. *Journal of clinical oncology*, 32(8):774, 2014.

[46] Yiming Xiao, Hassan Rivaz, Matthieu Chabanas, Maryse Fortin, Ines Machado, Yangming Ou, Mattias P. Heinrich, Julia A. Schnabel, Xia Zhong, Andreas Maier, Wolfgang Wein, Roozbeh Shams, Samuel Kadoury, David Drobny, Marc Modat, and Ingerid Reinertsen. Evaluation of mri to ultrasound registration methods for brain shift correction: The curious2018 challenge. *IEEE Transactions on Medical Imaging*, 39(3):777–786, 2020.

[47] Xuesong Lu, Su Zhang, Wei Yang, and Yazhu Chen. Sift and shape information incorporated into fluid model for non-rigid registration of ultrasound images. *Computer methods and programs in biomedicine*, 100(2):123–131, 2010.

[48] Martin Urschler, Joachim Bauer, Hendrik Ditt, and Horst Bischof. Sift and shape context for feature-based nonlinear registration of thoracic ct images. In *International Workshop on Computer Vision Approaches to Medical Image Analysis*, pages 73–84. Springer, 2006.

[49] Qing Guo, Wei Feng, Ce Zhou, Rui Huang, Liang Wan, and Song Wang. Learning dynamic siamese network for visual object tracking. In *Proceedings of the IEEE international conference on computer vision*, pages 1763–1771, 2017.

[50] Anfeng He, Chong Luo, Xinmei Tian, and Wenjun Zeng. A twofold siamese network for real-time object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4834–4843, 2018.

[51] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pages 850–865. Springer, 2016.

[52] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille, 2015.

[53] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[54] Pytorch-SiamFC. https://github.com/rafellerc/Pytorch-SiamFC. Accessed 29 June 2021.

[55] Mattias P Heinrich and Lasse Hansen. Highly accurate and memory efficient unsupervised learning-based discrete ct registration using 2.5 d displacement search. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 190–200. Springer, 2020.

[56] Paul W Holland and Roy E Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9):813–827, 1977.

[57] Hassan Rivaz, Emad M Boctor, Michael A Choti, and Gregory D Hager. Real-time regularized ultrasound elastography. *IEEE transactions on medical imaging*, 30(4):928–945, 2010.

[58] Jie Luo, Guangshen Ma, Sarah Frisken, Parikshit Juvekar, Nazim Haouchine, Zhe Xu, Yiming Xiao, Alexandra Golby, Patrick Codd, Masashi Sugiyama, et al. Do public datasets assure unbiased comparisons for registration evaluation? *arXiv preprint arXiv:2003.09483*, 2020.

[59] Richard D Bucholz, Kurt R Smith, Keith A Laycock, and Leslie L McDurmont. Three-dimensional localization: from image-guided surgery to information-guided therapy. *Methods*, 25(2):186–200, 2001.

[60] Hassan Rivaz and D Louis Collins. Deformable registration of preoperative mr, pre-resection ultrasound, and post-resection ultrasound images of neurosurgery. *International journal of computer assisted radiology and surgery*, 10(7):1017–1028, 2015.

[61] Laurence Mercier, Rolando F Del Maestro, Kevin Petrecca, David Araujo, Claire Haegelen, and D Louis Collins. Online database of clinical mr and ultrasound images of brain tumors. *Medical physics*, 39(6Part1):3253–3261, 2012.

[62] Sanam Maknojia, Fred Tam, Sunit Das, Tom Schweizer, and Simon J Graham. Visualization of brain shift corrected functional magnetic resonance imaging data for intraoperative brain mapping. *World neurosurgery: X*, 2:100021, 2019.

[63] Laith Alzubaidi, Mohammed A Fadhel, Omran Al-Shamma, Jinglan Zhang, J Santamaría, Ye Duan, and Sameer R Oleiwi. Towards a better understanding of transfer learning for medical imaging: a case study. *Applied Sciences*, 10(13):4523, 2020.

[64] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[65] Yunhui Guo, Honghui Shi, Abhishek Kumar, Kristen Grauman, Tajana Rosing, and Rogerio Feris. Spottune: transfer learning through adaptive fine-tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4805–4814, 2019.

[66] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *arXiv preprint arXiv:1411.1792*, 2014.

[67] Hamza Rasaee and Hassan Rivaz. Explainable ai and susceptibility to adversarial attacks: a case study in classification of breast ultrasound images. *arXiv preprint arXiv:2108.04345*, 2021.