Inter-Contrast and Inter-Modal Medical Image Registrations: From Traditional Energy-Based to Deep Learning Methods

NIMA MASOUMI

A THESIS
IN
THE DEPARTMENT
OF
ELECTRICAL AND COMPUTER ENGINEERING

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (ELECTRICAL AND COMPUTER ENGINEERING) AT
CONCORDIA UNIVERSITY
MONTRÉAL, QUÉBEC, CANADA

APRIL 2023

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Nima Masoumi**
Entitled: **Inter-Contrast and Inter-Modal Medical Image Registrations: From Traditional Energy-Based to Deep Learning Methods**

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

Doctor of Philosophy (Electrical 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. Govind Gopakumar

_____ External Examiner
Dr. Frédéric Lesage

_____ External to Program
Dr. Adam Krzyzak

_____ Examiner
Dr. Wei-Ping Zhu

_____ Examiner
Dr. M. N. Srikanta Swamy

_____ Thesis Supervisor
Dr. Hassan Rivaz

_____ Thesis Supervisor
Dr. M. Omair Ahmad

Approved _____
Dr. Jun Cai, Graduate Program Director

28 April 2023
Date of Defence

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

Abstract

Inter-Contrast and Inter-Modal Medical Image Registrations: From Traditional Energy-Based to Deep Learning Methods

Nima Masoumi, Ph.D.
Concordia University, 2023

Image registration is a crucial step in many medical image processing pipelines. The process aligns images of the same tissue taken at different times or with different imaging modalities. The first focus of this thesis is on the registration of ultrasound (US) images, which are low-cost, portable, safe, real-time, and commonly employed in several image-guided operations. Image registration of intraoperative US with preoperative images is required in image-guided surgeries. Computed Tomography (CT) scans and Magnetic Resonance Imaging (MRI) generally visualize the bones and soft tissues with better spatial details than US. Therefore, surgeons and interventionalists prefer them to US for the preoperative planning. These preoperative images should be registered to the intraoperative US images in image-guided interventions, which is a challenging task and an open area of research. Beyond image-guided interventions, image registration is a critical step in several other medical image analysis pipelines. The second focus of this work is on inter-contrast CT and MRI registrations. MRI is the primary modality for diagnosing neurodegenerative diseases such as Alzheimer's Disease. MRI comes with various contrasts, and the fusion of these contrasts taken at different times or from many subjects can give clinicians valuable information. However, MRI has a longer waiting time and less availability than CT. Thus, designing inter-modal image registration techniques to align MRI data with CT scans is essential in medical image analysis. Novel methods to tackle this problem are proposed in this thesis. The traditional image registration methods, which solve an optimization problem iteratively, can be time-inefficient for analyzing large datasets. Image registration using Deep Learning (DL) can accelerate the process but usually require training data. In this thesis, several novel methods for performing inter-contrast image registration are proposed in Chapters 3 to 5. These methods span both energy- and DL-based techniques with DL-based methods being more computationally efficient. We conclude the thesis in Chapter 6 by providing possible future research directions.

Acknowledgments

I want to offer my sincerest thanks to my knowledgeable, supportive, and genius supervisor, Dr. Hassan Rivaz. His support allowed me to expand my academic knowledge during my Ph.D. at Concordia University. In the difficult time of my research, he guided me to mature and face the reality of academic life. His attention to detail and intuition for supervising academic research resulted in great publications. Thanks to my supervisor Dr. M. Omair Ahmad, for the support, constructive discussions, and valuable comments for improving this thesis and my research papers.

I especially appreciate Dr. Yiming Xiao's invaluable support in editing my research papers and mentoring my academic career. I also extend my gratitude to my colleagues in the IMPACT lab for their guidance and for sharing their experiences with me. Special thanks to funding resources (Natural Science Engineering Council of Canada (NSERC) and in part by the Regroupment Strategique en Microelectronique du Quebec) which facilitated and encouraged me to continue strongly in the rise and falls of my research life. Many thanks to the thesis committee. You provided valuable hints for improving this thesis.

I am deeply indebted to my parents and wife for their emotional and financial support. I wish they would always be proud of me and that I could repay their deeds with my success.

Contents

List of Figures

List of Tables

List of Abbreviations

ANTs Advanced Normalization Tools

AD Alzheimer's Disease

ANOVA ANalysis Of VAriance

CSF CerebroSpinal Fluid

CT Computed Tomography

CNN Convolutional Neural Network

CR Correlation Ratio

CC Cross-Correlation

DL Deep Learning

DLCR Deep Learning-based deformable image registration using Correlation Ratio

DDD Degenerative Disc Disease

DOF Degree Of Freedom

XAI eXplainable Artificial Intelligence

FRE Fiducial Registration Error

FLASH Fourier-Approximated Lie Algebras

GD Gradient Descent

GPU Graphical Processing Unit

GM Gray Matter

HDR High-Dose-Rate

IGS Image-Guided Surgery

IVDD Interventional Disc Disease

KDE Kernel Density Estimation

LDDMM Large-Deformation Diffeomorphic Metric Mapping

ML Machine Learning

MRI Magnetic Resonance Imaging

MSE Mean Squared Error

mTRE mean Target Registration Error

MWA MicroWave Ablation

MI Mutual Information

NCC Normalized Cross-Correlation

NGF Normalized Gaussian Fields

NMI Normalized Mutual Information

ODE Ordinary Differential Equation

PD Prakinson's Disease

PDw PD-weighted

RF Radio Frequency

R-CNN Region-based Convolutional Neural Network

RaPTOR Robust PaTch-based cOrrelation Ratio

RMSD Root-Mean-Square Deviation

std standard deviation

SSD Sum-of-Squared Differences

SyN Symmetric Normalization

Mattes MI+SyN Symmetric image Normalization using Mattes Mutual Information

T1w T1-weighted

T2w T2-weighted

TCGA-SARC The Cancer Genome Atlas SARComa

TCIA The Cancer Imaging Archive

TAVI Transcatheter Aortic Valve Implantation

TRUS TransRectal UltraSound

TVUS TransVaginal UltraSound

US UltraSound

ViT Vision Transformers

vox voxel

VM-MSE VoxelMorph using Mean Square Error

WM White Matter

List of Symbols

η Correlation ratio

$\boldsymbol{Y}$ Moving image

$\boldsymbol{X}$ Fixed image

$\boldsymbol{\sigma^2}$ Variance

$\boldsymbol{T}$ Transformation

ϕ Deformation field

$\boldsymbol{\Delta}$ Laplacian operator

$\boldsymbol{V}$ Velocity field

$\boldsymbol{D}$ Derivative operator

$\boldsymbol{\nabla}$ Gradient operator

$\mathcal{A}$ σ-algebra

$\mathcal{P}$ Probability function

$\mathcal{M}$ U-Net model

$\mathcal{L}$ Loss function

ζ Logistic regression function

$\mathcal{Y}$ Segmentation of the moving image

$\mathcal{X}$ Segmentation of the fixed image

Chapter 1

Introduction

This chapter provides a necessary background to image registration. It motivates the requirement of medical image alignments by reviewing some clinical applications. Then, Section 1.2 and Section 1.3 elaborate on the thesis statement and the objective of the thesis, respectively. The organization of this thesis is provided in Section 1.4. Lastly, the publications arising from the thesis are listed in Section 1.5.

1.1 Image registration

Image registration is the process of aligning data acquired from different sensors. This process is required in various applications of engineering and clinics. Aligning images taken by medical imaging devices, namely medical image registration, is an active area of research. Thanks to the numerous radiological modalities for image-guided diagnosis and interventions, the development of image registration algorithms has been expanded and accelerated.

Magnetic Resonance Imaging (MRI) visualizes the connective and muscle tissues with excellent spatial details. MRI acquisition has multiple controllable parameters offering the practitioners a range of MRI contrasts. MR images with T1-weighted (T1w) contrast are commonly acquired for brain analysis, diagnosis, and surgical planning. Brain tissues and subcortical structures are distinguishable in standard T1w MR images. T1w MR images

of patients are routinely acquired for analysis of neurodegenerative disease progression. Early diagnosis of Parkinson's Disease (PD) and Alzheimer's Disease (AD) potentially enhances the life expectancy of patients. Registration algorithms play a crucial role in studying the inter-variability of these data. The techniques help align images with monitoring the progress of disorders in a longitudinal study. Besides, the inter-subject study of brain inter-variability by constructing a template from a large cohort of anatomical brain images has provided essential information about the human cerebrum. Building such a template involves several image registrations of the data.

In some clinical applications, multiple MRI contrasts are acquired. Other popular contrasts are T2-weighted (T2w) and Proton Density-weighted (PDw) MR. Some brain anatomical structures or anomalies may be better highlighted with T2w than T1w. For example, brain gliomas are hypo-intense in T1w but hyper-intense in T2w. Therefore, it is easier to identify them from normal tissues in T2w MR. Given this example, inter-contrast MR image registration can be helpful. Although radiologists and clinicians prefer MRI over other modalities in many applications, MRI has a longer patient waiting and acquisition time and less availability than Computed Tomography (CT). For instance, in minimally invasive abdominal surgeries, patients may be offered only one session of preoperative MRI. Nevertheless, several CT scans should be done preoperatively, intraoperatively, and postoperatively to monitor patients' prognoses. Ultrasound (US) imaging is the other common modality combined with MRI or CT. It is more accessible and non-ionizing. Designing inter-modal image registration techniques to match MRI data with CT scans and US images can be helpful. Nevertheless, inter-contrast/modal image registrations are challenging, and the current algorithms are generally application specific. Some of these registration problems are ill-posed and demand multiple image pre-processing steps. Furthermore, more public datasets for developing these methods are needed.

Multiple categorizations of image registration algorithms are feasible. In group-wise registration, more than two images are usually registered at once. In pairwise registrations, two images are registered; one image is fixed, and the other image, namely the moving image, is transformed to be aligned with the fixed image. The transformation model can

be rigid (rotation and translation), affine (rotation, translation, sheering, and scaling), and deformable. The pairwise registration problems are conventionally broken into rough alignment by a rigid and affine transformation, followed by a deformable transformation of the moving image. For densely sampled images, deformable transformations are free-form deformations formulable by B-splines. Several clinical applications require the transformations to be smooth and invertible. Invertible transformation is possible by setting constraints while estimating the deformation fields. The time-dependent smooth deformation fields are called diffeomorphisms. The process to align the images via diffeomorphisms is diffeomorphic image registration.

Several studies have been conducted over the last decade to tailor inter-modal diffeomorphic image registration techniques for clinical applications. Mitra *et al.* [1] implemented diffeomorphic registration of 2D Transrectal US (TRUS) images to MR slices. Kutten *et al.* [2] built the Mutual Information (MI)-based Large-Deformation Diffeomorphic Metric Mapping (LDDMM) [3] on a Hamiltonian framework to align inter-modal images. Reaungamornrat *et al.* [4] proposed a diffeomorphic MRI-CT registration for image-guided surgery by leveraging Symmetric Normalization (SyN) [5], diffeomorphic Demons [6], and MIND features [7]. However, performing inter-modal image registration of medical images is still an open area of research, given that the strategies are application-dependent.

Traditional image registration methods optimize a cost function to estimate the deformation field. These methods can be intensity-based or feature-based. Feature-based methods extract some features from the input images and then match them in the feature space. Intensity-based methods are formulated to register the data in the spatial domain of images. Traditional methods are computationally expensive and generally time-inefficient in analyzing large datasets. As mentioned earlier, Deep Learning (DL) approaches have been invented to tackle the issue. DL-based methods are usually faster in test time than traditional methods. However, training DL-based methods faces many challenges, such as unavailability and the small number of publicly annotated datasets. Although unsupervised DL methods aim to tackle the annotation issue, training such methods is still a challenging problem to solve.

Recently, DL-based approaches could perform excellently in various image registration problems [8–10]. Yang *et al.* [11] built on LDDMM to develop QuickSilver. Their method performs atlas-to-subject and inter-subject registrations. Balakrishnan *et al.* [12] proposed VoxelMorph, a diffeomorphic image registration technique using unsupervised learning. VoxelMorph was validated in several datasets and compared with the state-of-the-art SyN and NiftyReg. Mok *et al.* [13] developed a symmetric diffeomorphic algorithm to execute atlas-to-subject registration of brain images. Wang *et al.* [14] leveraged the Fourier-approximated Lie Algebras for Shooting (FLASH) [15] to implement DeepFLASH. DeepFLASH computationally outperformed FLASH, Quicksilver, and VoxelMorph, while achieving better image alignments regarding Dice scores of overlapping brain structures. Zhang *et al.* [16] trained a novel dual-transform network utilizing large datasets of brain images to perform monomodal registration efficiently.

The DL-based methods can assist radiologists with time-consuming tasks. The review of existing DL-based methods can be helpful for guidance to possible future research directions. The methods that can potentially assist US-guided surgeries/interventions are reviewed in several publications. Some papers highlight the clinical applications of US-guided surgeries/interventions. Pino *et al.* [17] and Dixon *et al.* [18] provided reviews of intraoperative US in brain glioma resection surgeries. Intraoperative US practices in breast cancer surgeries were reviewed in [19]. Sorenson *et al.* [20] and Gaudino *et al.* [21] studied the role of US-guided femoral and coronary arteries access in interventional neuroradiology, respectively. A review of US-guided percutaneous nephrolithotomy was provided by Alexandru *et al.* [22]. Danilo *et al.* [23] reviewed the intraoperative US practices of liver metastasis detection during the resection of colorectal cancer. Karagyozov *et al.* [24] studied US-guided biliary drainage practices. In a recent publication, Carstensen *et al.* [25] reviewed the US-guided injection in musculoskeletal interventions. Pain management in interventions using US imaging was reviewed by Diep *et al.* [26] and Gomez *et al.* [27]. The reviews of US-guided spinal surgeries were provided by Jiao *et al.* [28] and Viderman *et al.* [29]. The focus of some review papers was on the US imaging acquisition styles. Kessner *et al.* [30] and Nolsoe *et al.* [31] reviewed contrast-enhanced US practices for

interventional procedures. Dewitt *et al.* [32] reviewed the endoscopic interventional US cases. Antico *et al.* [33] reviewed the cases of robot-assisted US-guided interventions in an interesting study. In some other review publications, US guidance's goal was focused rather than the clinical application and imaging styles. For instance, Yang *et al.* [34] and Beigi *et al.* [35] reviewed medical instrument detection and localization techniques in US-guided surgeries/interventions. There are also reviews focusing on artificial intelligence and DL in US-guided interventions, such as the work of Shen *et al.* [36] and Jia *et al.* [37]. The works mentioned above included both diagnostic and therapeutic interventional US. However, the literature needs to include the focused study of DL-based methods in therapeutic interventional US.

1.2 Thesis statement

This thesis showcases novel techniques to register medical images of several modalities. The proposed algorithms have been rigorously validated to perform CT-US, MR T1w-T2w and T1w-PDw, and CT-MR T1w image registrations. These methods extend across energy- and DL-based techniques, with DL-based approaches being more computationally efficient. The dissimilarity metrics are primarily based on Correlation Ratio (CR), and the deformation models span rigid, diffeomorphic, and free-form transformations. The optimization strategies are carefully adopted according to the complexity of the registration task. Some clinical applications of novel algorithms developed in the thesis include US-guided spinal surgery, quantifying inter-subject variability of the brain, diagnosis and study of AD, estimation of tissue deformation for surgery, and diagnostic and follow-up of abdominal conditions.

1.3 Objective of the thesis

The methods that can potentially assist US-guided surgeries are reviewed in a multitude of publications. However, there are papers that highlight the clinical applications of US-

guided interventions [18,23,28], and the focus of some review papers is on the US imaging acquisition [30–33]. Given that the DL-based methods showed promising capability in enhancing the value of intra-operative US, we found that the review on advancing DL-based techniques in US-guided surgeries is beneficial. Therefore, Chapter 2 comprehensively reviews DL algorithms in US-guided interventions.

Previous works have developed CT-US image registration algorithms [38, 39], and we spotted several shortcomings. The datasets are private, their single-scale methods may not tackle larger misalignment, and exhaustive image registrations have not been performed for each range of initial misalignment. Chapter 3 addresses these research gaps by developing novel registration methods validated on the proposed public database.

Implementing inter-contrast and inter-modal medical image registrations remains a challenging task. Robust PaTch-based cOrrelation Ratio (RaPTOR) could successfully perform inter-modal image registrations [40–42], and FLASH computationally outperformed vector momentum LDDMM [43] in diffeomorphic registrations [15]. However, RaPTOR and FLASH have some pitfalls that should be implied. First, the transformation model of RaPTOR may not have a smooth inverse. Second, the dissimilarity metric of FLASH cannot provide a reliable measurement for inter-modal images [44]. Third, FLASH does not implement the standard multiresolution image pyramids to tackle larger deformations. Given these facts, Chapter 4 intends to design a diffeomorphic image registration to align images of different modalities and mitigate the drawbacks of RaPTOR and FLASH.

DL-based image registration techniques were designed to estimate a deformation field for image alignments. Recent studies perform image registrations in a number of challenging scenarios. Huang *et al.* [45] and Hoffmann *et al.* [46] proposed algorithms to perform inter-contrast MR image registrations. Nevertheless, their strategies are prone to fail in estimating a larger deformation fields, including the registration of AD patients' brain images. Given the success of CR [47,48]-based methods, such as RaPTOR, in traditional image registrations, we intend to design a DL-based algorithm to estimate correct deformation fields in challenging registration problems. Chapter 5 proposes Deep Learning-based deformable image registration using Correlation Ratio (DLCR) by employing semi-supervised learning

to minimize a CR-based loss function.

1.4 Organization of the thesis

The remaining thesis is organized as follows. Chapter 2 presents a comprehensive review of deep-learning algorithms in US-guided intervention applications, summarizes current trends, and suggests future directions. Chapter 3 showcases three different datasets of vertebrae with corresponding CT, US, and simulated US images. Besides, two patch-based rigid image registration algorithms are proposed, one based on Normalized Cross Correlation (NCC) and the other based on CR to register misaligned CT and US images. Chapter 4 presents an inter-modal/contrast registration algorithm that leverages the RaPTOR metric to allow inter-modal/contrast image alignment and bandlimited geodesic shooting demonstrated in FLASH algorithm for fast diffeomorphic registration. Chapter 5 elaborates on our newly developed method DLCR which stands for Deep Learning-based deformable image registration using Correlation Ratio. A proposed loss function consists of a CR-based dissimilarity metric and Dice loss. A model with U-Net architecture is trained in a semi-supervised setup to estimate a displacement field for 3D pairwise image registration of MRI data. The method is validated and tested by inter-subject registration of OASIS3 [49] and OASIS1 [50] datasets, respectively.

1.5 Publications arising from the thesis

- N. Masoumi, C. J. Belasso, M. O. Ahmad, H. Benali, Y. Xiao, H. Rivaz, "Multimodal 3D ultrasound and CT in image-guided spinal surgery: public database and new registration algorithms", In *IJCARS*, 16, 555–565 (2021).
  https://doi.org/10.1007/s11548-021-02323-2

- N. Masoumi, H. Rivaz, M.O. Ahmad, and Y. Xiao. DiffeoRaptor: diffeomorphic inter-modal image registration using RaPTOR. Int J CARS 18, 367–377 (2023).
  https://doi.org/10.1007/s11548-022-02749-2

- N. Masoumi, H. Rivaz, I. Hacihaliloglu, M. Omair Ahmad, I. Reinertsen and Y. Xiao, "The big bang of deep learning in ultrasound-guided surgery: a review," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,
  doi: 10.1109/TUFFC.2023.3255843.

- N. Masoumi, H. Rivaz, M.O. Ahmad, and Y. Xiao, "DLCR: Deep Learning-based deformable image registration using Correlation Ratio", Submitted to *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 2023. Lecture Notes in Computer Science, Springer, Cham.

Chapter 2

The big bang of deep learning in ultrasound-guided surgery: a review

US is a non-ionizing imaging modality that is commonly employed in the clinic, offering 2D, 3D, and 4D data. Although ultrasound probes are often operated in a free-hand manner by a physician or technician, to ensure image quality, semi-automatic or fully automatic image acquisitions are performed with the assistance of robotic arms in some applications [51]. While avoiding radiation risks, US scanners are portable and cost-effective as opposed to other staple imaging techniques, such as MRI and CT. In addition, US offers real-time anatomical and physiological information with great flexibility in applications, such as endoscopic, laparoscopic, transrectal, and transvaginal imaging. In addition to the most commonly seen B-mode contrast for structural imaging, US also provides additional contrasts, including Doppler US for flow imaging and ultrasound elastography computed from raw Radio Frequency (RF) scans to visualize biophysical properties of tissues. These advantages of US imaging make it a favourable modality for image-guided interventions, where it is used for instrument and biological tissue (e.g., lesions) detection and tracking [52, 53].

Despite multiple benefits, ultrasound still faces several drawbacks primarily as a result of its inherent imaging principle. First, US scans are often noisy and prone to imaging artifacts such as reverberations, refraction, and shadowing, making recognition of anatomy

and surgical tools difficult at times. Second, US usually has limited imaging depth, which can restrict the field of view for inspecting the pathological region. Lastly, unlike modalities such as MRI and CT that have standardized planes, the unique image contrast and arbitrary and unfamiliar imaging planes make it challenging to interpret US images. So far, a great number of image processing techniques were proposed to tackle these aforementioned drawbacks, including denoising [54], structure or instrument detection [55, 56], segmentation [57], and image registration [40–42, 58]. Before the advances in frameworks of Graphical Processing Units (GPUs) for model training, these techniques heavily rely on time-consuming iterative optimization methods or sub-optimal hand-crafted features for class Machine Learning (ML) algorithms. In comparison to conventional techniques, DL-based methods have shown excellent results in many US processing tasks by leveraging the computing power of modern GPUs [8, 59]. In addition, DL-based methods are faster in test time especially for large images [12]. With high requirement in accuracy, robustness, and efficiency, deep learning is well suited to facilitate US-guided interventions.

To date, a number of literature reviews have been conducted on the topic of US-guided interventions. However, most of these previous reviews focus on the survey of clinical applicability of intra-operative US [18, 23, 28] or related acquisition techniques [30–33]. With the great promise of DL techniques to enhance the value of intra-operative US, it is beneficial to provide a comprehensive review on the advancement of DL techniques in therapeutic interventional US. Based on the survey, we also identify the unmet clinical needs and suggest future research directions in the domain. This study intends to conduct a comprehensive review on deep-learning algorithms in the applications of US-guided interventions, summarize the current trends, and suggest future directions on the topic [60].

2.1 Literature selection

We searched the literature using Google Scholar database. The search was performed for publications from January 2015 to December 2022, in the period that DL-based techniques gain popularity in medical imaging. The search criteria “Ultrasound AND (Guided OR

Surgery OR Intraoperative) OR (Convolution OR Deep Learning)" was utilized. The papers reviewed are on the technical development and validation of the algorithms, and review articles, case reports, and clinical reports are excluded from the search. The selected papers were carefully read to ensure they were relevant to US-guided surgery and percutaneous interventions. The US-guided diagnosis and biopsies were excluded from our search to focus our review on the therapeutic application of US imaging. The survey resulted in 58 papers. A breakdown of reviewed papers' numbers for each year is shown in Fig 2.1. A breakdown of DL methods in this study is shown in Fig 2.2.

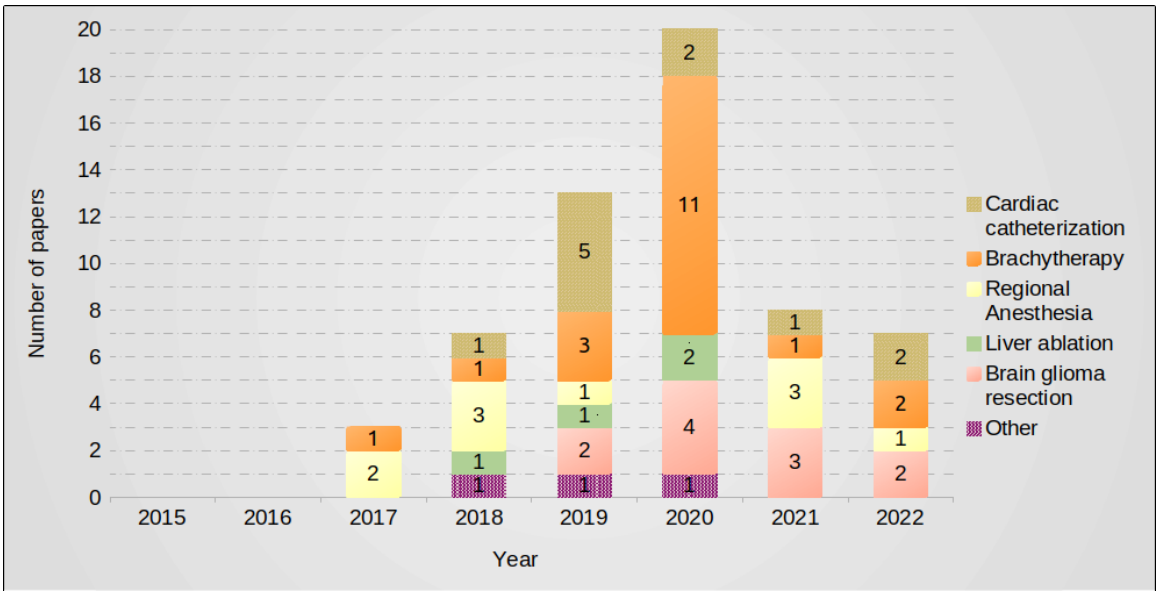

Figure 2.1: A breakdown of reviewed papers' numbers for each year is presented. In total, 58 papers were studied. We did not find relevant publications in 2015 and 2016. The number of DL-based approaches in US-guided therapeutic interventions has grown from 2016 until 2020. The drop in publications in the year 2021 is likely due to the COVID-19 pandemic, which substantially impacted performing US experiments.

2.2 Clinical applications

The main clinical applications of the reviewed papers are US-guided cardiac catheterization, brachytherapy, regional anesthesia, liver ablation, and brain glioma resection. While most papers focus on one applications, the others validated the proposed techniques in mul-

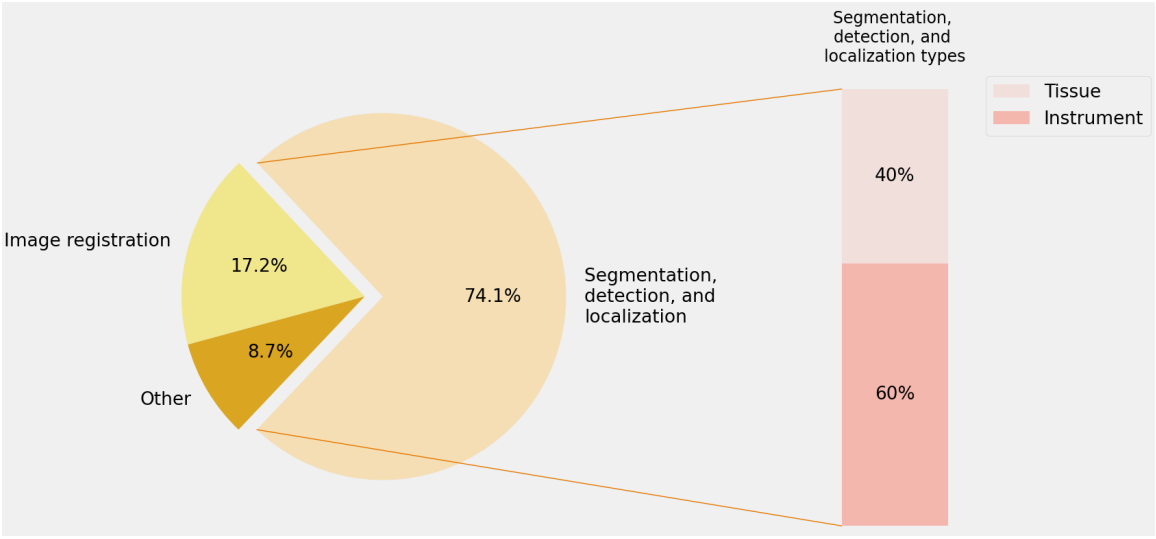

Figure 2.2: The methods were classified into three categories: 1. segmentation, detection, and localization, 2. image registration, and 3. other methods. Most methods perform segmentation, detection, and localization of medical instruments and target tissues. These methods can be further broken down into tissue and instrument segmentation, detection, and localization. The other methods include the classification of tissues, motion detection, etc.

tiple. Since typically different surgical procedures have different needs, the review for the developed techniques is conducted with respect to their clinical applications.

2.2.1 US-guided cardiac catheterization

Catheterization is common in various cardiac interventions, such as angioplasty and heart valve surgery. The catheter has a narrow tubular shape inserted into the patient's artery. The intraoperative X-ray is commonly acquired to localize the catheter. X-ray imaging has risks for interventionalists and patients due to its ionizing radiation. Given this fact, a safer choice, US-guided catheterization, is gaining popularity over intraoperative X-rays. However, locating the catheter in US images, especially near the heart chamber, is arduous. Computer-assisting image processing algorithms can automatically detect and localize the catheter in US images. Furthermore, they might segment the voxels showing the catheter with submillimeter precision. The segmentation task is challenging, and clinical requirements demand its adequate rapidness. Yang *et al.*, in several studies, showed that DL

Table 2.1: A summary of DL-based methods for heart catheterization is presented. The methods are mostly focused on catheter segmentation. The examined datasets are all private.

Reference	Task	Proposed approach	Dataset	Key metric and performance
Yang *et al.* [61]	Catheter localization	A CNN with binary pre-selection of candidate voxels, and applied a Frangi vesselness filter [72] with adaptive thresholding	3D *ex-vivo* porcine heart US	Hausdorff distance of $1.64 \pm 1.82 voxels$
Yang *et al.* [62]	Instrument localization	A modified multi-scale U-Net [73] with a hybrid loss consisting of a contextual loss and a class-balanced focal loss	3D *ex-vivo* porcine heart US and 3D US of *in-vivo* human heart during TAVI operations	Dice score (%) of 69.6 ± 10.9 for *ex-vivo* and 65.8 ± 9.2 for *in-vivo* data
Yang *et al.* [63]	Catheter localization	A 3D U-Net [73] with a cross-entropy focal loss	3D *ex-vivo* porcine heart US	Skeleton error of $1.28mm$
Yang *et al.* [64]	Catheter detection	An early fusion CNN and a late fusion CNN [74] with a weighted cross-entropy loss	3D *ex-vivo* porcine heart US	Position error of $1.7 voxels$
Yang *et al.* [65]	Instrument segmentation	Path-of-interest selection with fusion of a Pyramid-UNet [62] and a direction-fused U-Net which is based on a VGG16 encoder [75]	3D *ex-vivo* porcine heart US and 3D US of *in-vivo* human heart during TAVI operations	Dice score (%) of 70.5 ± 9.2 for *ex-vivo* and 66.5 ± 8.3 for *in-vivo* data
Yang *et al.* [66]	Instrument segmentation	Semi-supervised learning of a deep Q-network using a hybrid loss that consists of uncertainty and contextual constraints	3D *ex-vivo* porcine heart US and 3D US of *in-vivo* human heart during TAVI operations	Dice score (%) of 69.1 ± 7.3 for *ex-vivo* and 68.6 ± 7.9 for *in-vivo* data
Yang *et al.* [67]	Catheter segmentation	Weakly-supervised learning using a ResNet10 encoder [76] with the class attention maps-guided [77] pseudo-label generation	3D *ex-vivo* porcine heart frustum US	Dice score (%) of 65.4 ± 9.7
Yang *et al.* [70]	Catheter segmentation and localization	A direction-fused U-Net which is based on a VGG16 encoder [75]	3D *ex-vivo* porcine heart US	Dice score (%) of 67.7 ± 12.0
Min *et al.* [71]	Catheter segmentation	A VGG encoder [75] with pre-selection of candidate voxels, and applied a Frangi vesselness filter [72]	3D *ex-vivo* porcine heart frustum US	Dice score (%) of 67.3 ± 14

approaches could help the localization and detection of the catheter in US images [61–71]. They proposed methods to segment pixels/voxels into catheter and non-catheter classes. The methods were validated in several applications, such as Transcatheter Aortic Valve Implantation (TAVI). The methods are summarized in Table 2.1. The methods are primarily validated using private 3D *ex-vivo* animal and *in vivo* human datasets. In terms of instrument segmentation, these methods achieved Dice scores up to 70%.

2.2.2 US-guided brachytherapy

Brachytherapy is a procedure for treating certain kinds of cancers. In this procedure, radioactive small-size seeds are placed into the targeted region of the patient's body using

needles or a catheter. The radiation dose of seeds in brachytherapy is well-localized to the target and spares the adjacent tissues. Therefore, it has gained the attention of radiation therapists. For prostate brachytherapy, TRUS is commonly used to guide multiple medical instruments to the targeted region for the correct placement of seeds. Multi-needle localization, detection, and segmentation in US images can help the accurate insertion of radioactive seeds and possibly improve the treatment efficacy. The algorithms are ideally required to process US images in real-time. Robust techniques to handle noise and signal distortion are required for employment in clinical applications. Zhang *et al.*, in two different studies, proposed multi-needle localization methods using an attention U-Net [73] and a Region-based Convolutional Neural Network (R-CNN) ([78] and [79] respectively). They validated their method on 3D TRUS of patients who underwent High-Dose-Rate (HDR) brachytherapy. A CNN model was developed by Andersen *et al.* [80] to digitize needles in 3D TRUS of prostate HDR brachytherapy patients. Wang *et al.* [81] proposed a U-NET and an additional VGG16-based deep network to segment brachytherapy needles in 3D volumes reconstructed from 2D TRUS slices. Liu *et al.* [82] trained and tested a U-Net model to localize catheter in 3D reconstructed TRUS images taken during several prostate HDR brachytherapies.

Intraoperative prostate segmentation can accelerate the brachytherapy procedure, improve the patient's prognosis, and reduce costs. Girum *et al.* [83, 84] proposed DL approaches using a U-Net and a generative CNN respectively to segment the prostate in 3D reconstructed volumes from 2D TRUS slices. Orlando *et al.* [85] proposed a DL method using a modified U-Net for prostate segmentation on clinically diverse 3D TRUS images. Later, they developed two DL methods using a modified U-Net and a U-Net++ [86, 87], which were trained on 2D TRUS slices [88]. Nevertheless, the methods were tested on 3D TRUS volumetric images. A prostate segmentation method using multidirectional V-Net [89] was proposed by Lei *et al.* [90]. A clinical subject delineation method using residual neural networks for low-dose-rate brachytherapy was developed by Anas *et al.* [91]. The method was validated on 2D TRUS slices with manual segmentations available. Karimi *et al.* [92, 93] proposed a novel CNN architecture for prostate segmentation

in 2D TRUS images. Xiuxiu *et al.* [94] proposed a deep-attentional GAN-based method to improve the resolution of 3D TRUS images. Golshan *et al.* [95] proposed a modified LeNet architecture [96] for radioactive seeds segmentation in 3D TRUS images. Accurate implantation of radioactive seeds in the targeted region can improve the efficacy of brachytherapy.

Pre- and possibly intra-operative MRIs are acquired for operation planning, visualization and detection of medical instruments in prostate brachytherapy. MRIs generally have less noisy images and show better details of the target area. Image registration of intraoperative TRUS with the MRI can leverage the prostate and medical instruments segmentation. Chen *et al.* [97] proposed a DL approach using V-Net and U-Net architectures to segment and register the prostate in MR and TRUS. Zeng *et al.* [98] performed 3D non-rigid registration of MR-TRUS using convolutional and recurrent neural networks.

Brachytherapy is not confined to prostate cancer treatment. Rodgers *et al.* [99] proposed a DL-based method for needle localization in 3D transvaginal US images of interstitial gynecologic HDR brachytherapy. Sun *et al.* [100] generated pseudo-CT images from intraoperative US images of cervical cancer patients for brachytherapy. The DL methods in this section are summarized in Table 2.2. The Dice score, followed by the shaft and needle tip localization errors, are the key metrics for quantitative performance assessment. In general, the algorithms achieved sub-millimeter shaft and needle tip localization errors.

2.2.3 US-guided regional anesthesia

Needle-based regional anesthesia is conventionally used in operating rooms. It usually requires an experienced expert to deliver the anesthetic injection. US-guided regional anesthesia can help the anesthesiologist with the procedure. Detection and localization of the injection needle shaft and tip can be challenging. In 2D US scans, needle tips are occasionally out-of-plane or not visible. Processing US original RF data or reconstructed images is helpful for accurate and reliable procedures. The algorithms are ideally required to process US data in real-time. DL approaches can help with needle localization in US images [106]. Mwikirize *et al.* [107–109] developed CNNs in three studies to localize the needle tip in

Table 2.2: A summary of DL-based methods for US-guided brachytherapy is presented. The methods are mostly focused on target and instrument segmentation. Public datasets are marked with "*".

Reference	Task	Proposed Approach	Dataset	Key metrics and Performance
Zhang *et al.* [78]	Multi-needle localization	A deep supervised attention U-Net with a weighted total variation regularization	3D *in-vivo* TRUS of prostate HDR brachytherapy	Shaft localization error of $0.29 \pm 0.23mm$ and needle tip localization error of $0.44 \pm 0.83mm$
Zhang *et al.* [79]	Multi-needle localization	A reformulated large-margin mask R-CNN [101] that is combined with a density-based spatial clustering [102]	3D *in-vivo* TRUS of prostate HDR brachytherapy	Shaft localization error of $0.09 \pm 0.04mm$ and needle tip localization error of $0.33 \pm 0.36mm$
Andersen *et al.* [80]	Digitization of prostate brachytherapy needles	A 3D U-Net architecture [73] with a Dice loss	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy	Root-Mean-Square Deviation (RMSD) median (interquartile range) of $0.55 \quad (0.35 - 0.86)$
Wang *et al.* [81]	Needle segmentation	A U-Net followed by a VGG16 network with a categorical cross-entropy loss	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy	Resolution of needle trajectories of $0.66mm$ and $0.31mm$ in x and y direction respectively
Liu *et al.* [82]	Catheter localization	A U-Net architecture with a focal Tversky loss function [103]	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy	80% within $2mm$ catheter reconstructions
Girum *et al.* [83]	Prostate clinical target-volume boundary detection	A modified U-Net with an integrated regression model using global average pooling	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy	Dice score (%) of 88.0 ± 2.0
Girum *et al.* [84]	Prostate clinical target-volume segmentation	A CNN for prior-knowledge generator and a CNN for the segmentation	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy, 3D *in-vivo* postoperative CT scans of prostate HDR brachytherapy, and *2D *in-vivo* echocardiography images	Dice score (%) of 96.9 ± 0.9, 95.4 ± 0.9, and 96.3 ± 1.3 for TRUS, CT, and 2D echocardiography images respectively
Orlando *et al.* [85]	Prostate segmentation	A 2D modified U-Net with a Dice loss	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy and biopsy	A median (first quartile - third quartile) absolute Dice score (%) of $94.1 \quad (92.6 - 94.9)$
Orlando *et al.* [88]	Prostate segmentation	Trained U-Net and U-Net++ [86, 87] architectures separately using 2D slices	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy and biopsy	A median (first quartile - third quartile) absolute Dice score (%) of $94.1 \quad (92.6 - 94.9)$ and $94.0 \quad (92.2 - 95.1)$ for U-Net and U-Net++ respectively
Lei *et al.* [90]	Prostate segmentation	A multidirectional deeply supervised V-Net [89] with a hybrid loss that consists of a binary cross-entropy loss and a batch-based Dice loss	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate	Dice score (%) of 91.9 ± 2.8
Anas *et al.* [91]	Clinical target-volume delineation	CNNs based on ResNets [76] and dilated convolution at deeper layers	2D *in-vivo* TRUS of prostate brachytherapy patients	Dice score (%) of 93.67 ± 3.71
Karimi *et al.* [92, 93]	Clinical target-volume segmentation	Sparse subspace clustering [104] of features learned with a convolutional auto-encoder and a modified U-Net architecture	2D *in-vivo* TRUS of prostate brachytherapy patients	Dice score (%) of 93.9 ± 3.5
Xiuxiu *et al.* [94]	Improving US image resolution	Integrating a deeply supervised attention model into a generative adversarial network-based	3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate	Mean absolute error of 6.5 ± 0.5
Golshan *et al.* [95]	Brachytherapy seeds detection	A LeNet [96] architecture with a cross-entropy loss	3D *in-vivo* volumes reconstructed from 2D original TRUS RF data of prostate brachytherapy patients	Precision, recall, and F1-score (%) of 78.0 ± 8.0, 64.0 ± 10.0, and 70.0 ± 8.0 respectively
Chen *et al.* [97]	MR to TRUS image registration and prostate segmentation	Segmentation-based registration using two weakly-supervised 3D V-Nets [89] for segmentations and a 3D U-Net for the registrations	3D *in-vivo* T2w MRI and 3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy	Dice score (%) of 97.0 ± 0.0 and 87.0 ± 5.0 for segmented mask and manual contours respectively
Zeng *et al.* [98]	MR to TRUS prostate registration	A modified U-Net [73] and a bidirectional convolutinoal LSTM with a hybrid loss that consists of a bending energy loss and a Dice loss	3D *in-vivo* T2w MRI and 3D *in-vivo* TRUS volumes reconstructed from 2D slices of prostate HDR brachytherapy	Dice score (%) 90.0 ± 4.0
Rodgers *et al.* [99]	Needle localization	A 2D U-Net [120] for 2D data and randomized 3D Hough transforms [105] for 3D data	3D *in-vivo* transvaginal US (TVUS) volumes reconstructed from 2D slices of interstitial gynecological HDR brachytherapy, 2D US slices of phantom brachytherapy, and 2D US slices of ablation therapy	Median position difference (first quartile - third quartile) of $0.27 \quad (0.20 - 0.68)mm$ and $0.79 \quad (0.62 - 0.93)mm$ for 2D and 3D TVUS respectively
Sun *et al.* [100]	Pseudo-CT image synthesis from US	A part of VGG19 [75] network and a hybrid loss that consists of a content loss, a style loss, and a Dice loss	3D *in-vivo* CT scans and 3D *in-vivo* US volumes of cervical cancer patients, additional 3D CT scans of cervical cancer patients, and 3D US phantom data	T-test of structural similarity index between the ground-truth and synthesized CT with $t = 3.22$ and $t = 2.86$ for background and foreground regions respectively

real-time 2D US images. Gao *et al.* [110] proposed a needle segmentation method using a U-Net architecture. Pourtaherian *et al.* [111] proposed a needle tip detection method using orthogonal-plane CNNs. They validated their method on *ex-vivo* chicken breast 3D US images. Later, they developed a method for the localization of needle tips with submillimeter domain accuracy using dilated CNNs [112].

Nerve segmentation in images of US-guided regional anesthesia can help the practitioners with the process. Automatic non-learning methods using Kalman filters could rapidly perform nerve and artery segmentations [113]. Generally, these methods are computationally expensive and require intensive hyperparameter tuning. Esmistad *et al.* [114] proposed a technique using U-Net for musculocutaneous, median, ulnar, and radial nerve segmentation during axillary nerve block procedures. Baby *et al.* [115] developed a U-Net model to delineate the brachial plexus in 2D US images. A conditioned U-Net model (www.kaggle.com/harolddiaz1018/unet-cond) was trained by Diaz-Vargas *et al.* [116] to segment ulnar, median, femoral, and sciatic nerves in 2D US images. Maneas *et al.* [117] modified an established residual neural network to improve the axial and lateral resolution of tracked US images for needle localization. They trained their model on synthetic data, and the model was validated on a fetal sheep heart *in-vivo* data. The DL methods in this section are summarized in Table 2.3. The shaft and needle tip localization errors are the key metrics for quantitative performance assessment. The algorithms achieved submillimeter errors.

2.2.4 US-guided liver ablation

Image-guided microwave ablation (MWA) is a promising therapeutic percutaneous intervention that provides a high intralesional temperature. Real-time US imaging techniques can visualize the target for accurate lesion MWA and complete tumour eradication. However, the ablation region margin is not easily detectable in US images. While ablation region delineation is feasible using techniques such as ultrasound elastography [122], we focus our review on DL techniques. Unsupervised classification of target region tissues by leveraging an ML/DL-based method is a candidate approach. Zhang *et al.* [123] utilized the

Table 2.3: A summary of DL-based methods for US-guided regional anesthesia is presented. Anesthesia needle tip localization is the focus of the majority of works. Public datasets are marked with "*".

Reference	Task	Proposed Approach	Dataset	Key metrics and Performance
Mwikirize *et al.* [107]	Real-time needle detection	A region-based CNN [118] and a region-proposal CNN	2D *ex-vivo* US bovine and porcine tissues, and 2D US of bovine/porcine tissues overlaid on lumbosacral spine phantom	Shaft localization error of $0.82^\circ \pm 0.4^\circ$, and needle tip localization error of $0.23 \pm 0.05mm$
Mwikirize *et al.* [108]	Real-time needle tip localization	Needle enhancement followed by a CNN for needle tip classification, and a CNN regression network	2D *ex-vivo* US of bovine, porcine, and chicken tissues overlaid on lumbosacral spine phantom	Needle tip localization error of $0.55 \pm 0.07mm$
Mwikirize *et al.* [109]	Needle tip localization	A novel network that consists of convolutional layers and recurrent layers (CNN-LSTM) with a Mean Squared Error (MSE) loss	2D *ex-vivo* US of bovine, porcine, and chicken tissues overlaid on lumbosacral spine phantom	Needle tip localization error of $0.52 \pm 0.06mm$
Gao *et al.* [110]	Needle localization and enhancement	Beam steering followed by a modified U-Net for segmentation, and a needle fusion algorithm	2D *ex-vivo* US of bovine, porcine, and chicken tissues	Needle tip localization error of $0.29 \pm 0.02mm$
Pourtaherian *et al.* [111]	Needle detection	Two CNNs with shared and independent convolutional filters respectively using a categorical cross-entropy cost	3D *ex-vivo* US of a chicken breast	Precision 83% at 76% recall
Pourtaherian *et al.* [112]	Needle localization	CNNs with dilated convolutions and spatial pyramid pooling features	3D *ex-vivo* US of a porcine leg	Qualitative assessment
Esmistad *et al.* [114]	Nerve segmentation	A modified U-Net	2D *in-vivo* US of volunteers and patients undergoing axillary nerve block procedures	Precision of 88%, 63% 79%, 67%, and 44%, and recall of 0.81, 0.56, 0.71, 0.62, and 0.37 for blood vessel, musculocutaneous nerve, median nerve, ulnar nerve, and radial nerve respectively
Baby *et al.* [115]	Nerve segmentation	A modified U-Net	*2D *in-vivo* US of patients' brachial plexus	Dice score 71%
Diaz-Vargas *et al.* [116]	Peripheral nerve segmentation	A conditioned U-Net with a Dice loss	2D *in-vivo* US of patients' ulnar, median, femoral, and sciatic nerves	Dice score (%) of 70.0 ± 27.0
Maneas *et al.* [117]	Instrumented ultrasonic tracking	ResNet architecture [121] with a L1-loss	2D synthetic US RF data, and 2D *in-vivo* US of fetal sheep heart	Root-mean-square error of 0.15 ± 0.03 for the synthetic data, and signal energy concentration ration of 99.9% for the *in-vivo* data

Table 2.4: A summary of DL-based methods for US-guided liver ablation is presented. The examined datasets are all private.

Reference	Task	Proposed Approach	Dataset	Key metrics and Performance
Zhang *et al.* [123]	Thermal lesion detection	Matched pathology images to US RF data followed by training a CNN with two paths	2D *ex-vivo* US B-mode and RF data liver tissues	Dice score 86.88%
Wang *et al.* [124]	Thermal lesion detection	Image registration of RF data and optical images followed by training a CNN	2D *ex-vivo* US B-mode and RF data, and optical images of the porcine liver tissues	Receiver operating characteristic curve of 0.87
Kondo *et al.* [125]	Tumour motion detection	A VGG16 [75] followed by a U-Net architecture with a hybrid loss	2D US of liver phantom	Mean absolute error of $0.342mm/frame$
Arif *et al.* [126]	Needle detection	Image registration of needles in different time points and needle segmentation using a compressed V-Net [89]	3D *in-vivo* US of liver biopsy patients, and 3D US of puncturing phantoms	Mean absolute error of $1.00mm$ and 2.0° for needle position and orientation respectively

original US RF data matrices and trained a CNN network to delineate the ablation region in *ex-vivo* data of the porcine liver. Wang *et al.* [124] proposed a CNN-based method for ablation region detection and monitoring MWA. They performed image registration of US RF data and optical images to boost the accuracy of their method in terms of receiver operating characteristic curves. Kondo *et al.* [125] proposed an out-of-plane motion detection system using CNNs to track the liver tumour movement in ablation therapies.

Ablation needle detection and visualization can help interventionalists during the MWA procedure. Arif *et al.* [126] proposed a real-time bi-planar needle detection and visualization for liver 3D US images. Their method utilizes a U-Net architecture and specific post-processing to perform the needle detection. They execute the registration of images in different time frames acquired from liver phantom and ten patients. The DL methods in this section are summarized in Table 2.4. Dice score and mean absolute error are the key metrics for quantitative performance assessment.

2.2.5 US-guided brain glioma resection

US machines' portability and cost-effectiveness of US imaging contributed to the growing popularity of intraoperative US acquisition. The US machines and probes can be calibrated and synced with a neuronavigation system in operating rooms. Practitioners may execute an image registration of preoperative images and the intraoperative US to update the surgical plan. For instance, in brain glioma surgery, intraoperative US images can be registered

to the preoperative MRIs (or US images). Because after surgeons open the dura, the brain shifts up to $18mm$ due to several causes, including gravity, cerebrospinal fluid loss, drug administration, retraction, resection etc [127, 129]. This phenomenon is commonly named brain shift. Brain shift can make the preoperative planning invalid. Therefore, fast registration of preoperative and intraoperative data is crucial. Public datasets of BITE [127] and RESECT [129] impacted the development of DL-based methods for brain-shift correction. In the CuRIOUS2018 Challenge, which was held in conjunction with MICCAI 2018, the participating teams were asked to register preoperative MRI to intraoperative US images of RESECT dataset. The challenge results and participating teams' methods are summarized and compared in [128]. Canilini *et al.* [130] proposed a DL method using a CNN to segment sulci and falx cerebri in US images. Then, they used the segmentation masks to register intraoperative, preoperative and postoperative US images. The method was tested on BITE and RESECT datasets. Given the fact that these datasets provide manual homologous landmarks, Canilini *et al.* calculated mean Target Registration Error (mTRE) for the quantitative validation of their method. Later, they trained a U-Net architecture to generate segmentation masks of resection cavities [131]. They registered the US volumes using these masks.

Zeineldin *et al.* proposed DL-based methods with U-Net architectures in different studies to register preoperative MRI to intraoperative US images [132–134]. They employed MSE Loss for their model training in [132]. Later, they used MSE loss and NCC loss in a comparison study in [133] and NCC loss in [134]. Pirhadi *et al.* [135] employed a Siamese neural network [136] to perform landmark-based registration of pre-resection intraoperative US to post-resection intraoperative US scans.

Finding the precise boundaries of the tumour and its segmentation can assist the surgeons to optimize the resection boundary. Zeineldin *et al.* [137] employed U-Net and TransUNet networks [138] to segment brain tumours in US images. Carton *et al.* [139] proposed a DL-based method with a 3D U-Net architecture to segment the brain tumours of RESECT dataset intraoperative US images. In addition to lesion segmentation, classification of the lesion into different glioma grades or solitary brain metastases can be substantial because

Table 2.5: A summary of DL-based methods for US-guided brain glioma resection is presented. Most methods perform image registration for brain shift correction in BITE [127] and RESECT [129] datasets. Public datasets are marked with "*".

Reference	Task	Proposed Approach	Dataset	Key metrics and Performance
Canilini *et al.* [130]	Segmentation and registration of US volumes	Segmentation by a modified U-Net [73] and registration of generated masks	*3D *in-vivo* US volumes reconstructed from 2D slices of RESECT [129] and BITE [127] datasets	mTRE of $2.05 \pm 1.12mm$ for RESECT and $2.48 \pm 2.67mm$ for BITE dataset
Canilini *et al.* [131]	Resection cavity segmentation and registration of US volumes	Segmentation by a modified U-Net [73] and registration of generated masks	*3D *in-vivo* US volumes reconstructed from 2D slices of RESECT [129] and BITE [127] datasets	mTRE of $1.21 \pm 0.66mm$ for volumes before and after resection of RESECT, $1.22 \pm 1.20mm$ for volumes before and during resection of RESECT, and $2.38 \pm 2.78mm$ for BITE dataset
Zeineldin *et al.* [132]	MR to US registration	A U-Net architecture with a MSE loss	*3D *in-vivo* US volumes reconstructed from 2D slices and 3D T2-FLAIR MRI of RESECT [129]	Mean squared error of 85
Zeineldin *et al.* [133]	MR to US registration	Two U-Net architecture with MSE and NCC losses respectively	*3D *in-vivo* US volumes reconstructed from 2D slices and 3D T2-FLAIR MR of RESECT [129] and BITE datasets	mTRE of $0.84 \pm 0.16mm$ for RESECT and $1.47 \pm 0.61mm$ for BITE dataset
Zeineldin *et al.* [134]	MR to US registration	A U-Net architecture with a NCC loss	*3D *in-vivo* US volumes reconstructed from 2D slices and 3D T2-FLAIR MRI of RESECT [129] and BITE datasets	mTRE of $0.99 \pm 0.22mm$ for RESECT and $1.68 \pm 0.65mm$ for BITE dataset
Pirhadi *et al.* [135]	Landmark-based US volumes registration	A Siamese network [136] for detecting landmarks with a 2.5D approach [141]	*3D *in-vivo* US volumes reconstructed from 2D slices of RESECT [129] and BITE [127] datasets	mTRE of $1.22 \pm 0.46mm$ for volumes before and after resection of RESECT, $1.11 \pm 0.43mm$ for volumes before and during resection of RESECT, and $1.76 \pm 1.48mm$ for BITE dataset
Zeineldin *et al.* [137]	Brain tumour segmentation	U-Net [120] and TransUNet [138] architectures	*3D *in-vivo* US volumes reconstructed from 2D slices of RESECT [129] dataset	Dice scores (%) of 93.50 and 93.70 for U-Net and TransUNet respectively
Carton *et al.* [139]	Brain tumour segmentation	Three U-Net networks with Dice losses	*3D *in-vivo* US volumes reconstructed from 2D slices of RESECT [129] dataset	Median Dice score (%) of 72.00
Capeda *et al.* [140]	Glioblastoma and solitary brain metastases classification	Employed Inception V3 network from Orange software version 3.26 (University of Ljubljana, Slovenia)	2D *in-vivo* US images of supratentorial tumour patients who underwent craniotomy	Classification accuracy values of 0.79 to 0.94 for B-mode US and 0.84 to 0.97 for elastography data

the surgical procedures vary for each case. Cepeda *et al.* [140] proposed a DL approach to analyze the candidate lesions in patients who underwent craniotomy. They used B-mode and strain elastography images to correctly classify the lesions as glioblastoma or solitary brain metastases. The DL methods in this section are summarized in Table 2.5.

2.2.6 Other clinical applications

Sections III.A-E reviewed the DL approaches in widely studied clinical applications. This section reviews the clinical applications with a few DL-based approaches. Lee *et al.* [142] proposed a DL method to classify liver fibrosis. They utilized the data for patients who underwent liver resection surgery. Gillies *et al.* [143] employed a U-Net architecture with a Dice loss to detect general interventional tools in 2D US images. They utilized the datasets of prostate and interstitial gynecologic brachytherapy, liver, and kidney ablations. Wang *et al.* [144] proposed a deep attentive method for prostate segmentation. Hu *et al.* [145] developed an adversarial deformation regularization method for preoperative and procedural TRUS image registration. However, the developed methods of Wang *et al.* [144] and Hu *et al.* [145] have not been designed for a focused application, and they can be used for prostate brachytherapy or prostatectomy.

2.3 Discussion and future directions

Based on the literature included in the review, deep learning techniques have shown great promise to enhance the value of intra-operative US in surgical interventions. While segmentation, detection, and localization are the main techniques under development, these also need to be adapted to the application-specific needs and from the current state-of-the arts, we identified a few unmet clinical needs that could be addressed by DL methods in the future. In the literature of brachytherapy, most efforts in DL techniques were dedicated to prostate treatment, even though US-guided brachytherapy was also practised for lung cancer, breast cancer, anal cancer, and abdominal wall metastases. Similarly, deep learning approaches in US-guided ablation is primarily focused on liver while kidney and prostate ablation therapies still have potentials for further technical development. In US-guided tumour resection procedures, similar DL methods can be further adapted for lumpectomy, prostatectomy, tongue cancer resection, laparotomy, pancreatic cancer resection, and bladder cancer resection. Finally, although, US was investigated as an intraoperative imaging tool in orthopedic surgery procedures, extensive evaluation studies is still lacking. Cur-

rently, most focus in this domain has been given to developing accurate, robust, and fast bone segmentation [146, 147]. We believe a similar effort should be directed to evaluate US bone registration approaches [148].

Despite the excellent performance, deep learning techniques, including those reviewed in this article still have several drawbacks. First, most algorithms still require large well-annotated data to achieve good performance; second, the trained networks often have limited adaptability to new domains (e.g., images from different scanner types or setting). Lastly, most existing algorithms still lack of transparency to help verify the quality of the outcomes. Currently, the lack of large-scale well annotated databases, especially the public repositories poses a bottleneck in algorithm development and fair performance benchmarking, and this also partially contributed to the various under-explored clinical applications as mentioned earlier, besides their application-specific challenges. Compared with MRI and CT scans, clear structural delineations in US is more difficult due to the nature of the imaging principles, and often co-registered biopsy, MRI, and CT data may be required when it comes to confirmation of pathological tissue segmentation. As direct contact is needed, for endoscopic applications, image acquisition also demands elaborate setup using surgical navigation systems or robotic assistance. These further complicate the construction of relevant datasets besides the privacy concerns commonly associated with medical data sharing.

In current literature, convolutional neural networks, especially different variants of U-Net architecture [120] have dominated the reviewed methods. In many applications, to overcome the limited data, CNNs previously trained with other imaging modalties (e.g., natural images) were adapted to the application domain with transfer learning [149]. However, partially due to the lack of public data, general purpose DL algorithms that are more tolerating to scanner types and clinical applications still face major challenges. A few initiatives in MRI and CT DL registration and segmentation, such as the Learn2Reg MICCAI Challenge [150] and the Medical Segmentation Decathlon challenges [151] have attempted to help development these types of algorithms, but there is still a lack of similar endeavors in US. The more recent Vision Transformers (ViT) has shown better performance in learn-

ing long-range spatial dependencies than CNNs, which require more elaborate architecture design to model spatial context of the image [152]. Adoption of ViT and its variants may further improve the accuracy of existing and future DL-methods for intra-operative US.

Interpretability and trustworthiness of deep learning algorithms are crucial for the widespread adoption of the end products to the clinic. Conventional algorithms often have a goal-driven black-box design, and in this case, without careful verification, faulty automatic outputs can cause harms to the patients. The latest trend in explainable AI (XAI) intends to improve algorithm transparency through techniques, including spatial attention/activation visualization [153, 154], uncertainty estimation, and multi-task learning [155]. For various surgical applications, XAI methodologies can potentially further detect and explain problematic results from DL-based iUS processing that offer real-time feedback to improve the robustness and reliability of the algorithms, and thus the safety and efficiency of the surgery.

2.4 Summary

This literature review studied 58 DL-based approaches for US-guided heart catheterization, brachytherapy, regional anesthesia, liver ablation, and glioma resection. Near 74% of reviewed methods perform segmentation, detection, and localization of medical instruments and target tissues. Possible research directions for DL-based approaches were discussed in Section 2.3.

Chapter 3

Multi-modal 3D ultrasound and CT in image-guided spinal surgery: public database and new registration algorithms

---

Intraoperative US is non-invasive, renders real-time images of soft tissue, and has been used in IGS to help increase the accuracy of surgical interventions [156, 157]. Acquiring intraoperative US for IGS is beneficial in multiple types of spinal surgeries: for lumbar vertebrae spinal fusion surgery for DDD, for canine spinal cord surgery (mostly in the middle vertebrae) for IVDD, and for spinal surgeries that stabilize vertebral fractures [158, 159].

Prior to a spinal surgery, CT images are acquired pre-operatively to plan the intervention. To help guide the surgical procedure during surgery, intra-operative US images corresponding to the pre-operative CT images can help increase the accuracy of the surgery. Image registration in IGS is required because the intra-operative US must be spatially aligned with the pre-operative image so that the image fusion algorithm can be applied to integrate the information during the operation. Simulation of the US image from the pre-operative image can increase the accuracy of this multi-modal image registration [160].

Current US simulation techniques are divided into two categories: ray-based methods and wave-based methods [161, 162]. Ray-based methods are quicker in computation time but they do not provide realistic images whereas wave-based methods are more realistic but lack US specific artifacts such as shadowing [163]. Alternatively, wave-based methods, for instance Field II simulation software, can be used as a ground-truth for other simulation techniques at the expense of having a greater computation time [164].

Previous work has developed multi-modal image registration algorithms to align the CT-US and CT-simulated US images [38, 39]. However, the following shortcomings exist in the implementation and validation of their methods. First, the utilized datasets are not publicly available. Employing a publicly available data can provide researchers a valuable information to further compare their methods with the existing ones. Second, the implementation of single-scale methods restrict performing registration on larger initial misalignment. Third, by assuming that the rigid registration is sufficient to compensate the initial misalignment, given the 6 Degree Of Freedom (DOF), the resolution of images, and the machine precision, exhaustive image registrations may require to be performed for each range of initial misalignment. This can reveal valuable information about the robustness and applicability of method in real applications.

Following the success of our database of US and MR images [129], we created a database of CT scans and corresponding US images, and proposed a simple and realistic wave-based US simulation method to simulate the US images from the CT images. In the first part of the database, the US from The Cancer Genome Atlas Sarcoma of the three patients' lumbar vertebrae (TCGA-SARC) [165, 166] were simulated from their CT images. In the second part of the database, the CT scan, the spatially tracked US, and the simulated US of the canine thoracic and cervical vertebrae phantom are provided. In the third part of the database, the CT scan, the spatially tracked US, and the simulated US of the lamb lumbar vertebrae are provided. For each of the two latter datasets, we provided 16 landmarks of analogous structures in the CT and US images and performed fiducial registration to acquire a silver standard ground-truth of the registration.

This dataset is the first of its kind and the images can be used for the evaluation of

image registration techniques to improve image–guidance in spinal surgery. The simulated US images provide a gold standard ground–truth for the registration techniques while the spatially tracked US images give a more realistic view of the vertebrae. In addition, the simulated US images can be used as a guide to validate real–time US simulation methods that are used in some image registration techniques [167, 168], as well as preoperative simulation and planning [169].

In light of the above facts, we implemented two patch-based rigid registration algorithms: one based-on NCC and the other based-on the CR [40,41,170]. The barrier method with the logarithmic barrier function was used as the optimization technique [171]. Then, 6000 registration problems were created on the proposed datasets. By applying the methods on the set of registration problems, 12000 registrations were performed to align the images and compare the performance of methods. Our results show that both methods were able to register CT and US images of the proposed datasets successfully. This study intends to design an algorithm to perform inter-modal image registration [42].

3.1 Potential applications of proposed database

The database consists of three different sets. In each set, the CT scan of the subject and the corresponding simulated US image have been included. The US images are simulated from the CT scans using the Field II software and are similar to real US images. The simulated US image displays the shadowing artifact of US images below the bone surface. They also provide a gold standard ground-truth in alignment with the CT scans as it is difficult to obtain with *ex-vivo* and *in-vivo* data. Thus, CT scans and simulated US images are an ideal resource for evaluating multimodal image registration techniques. Medical image registration methods can be used in IGS systems in order to help surgeons obtain accurate intraoperative surgical information.

The simulated US images are perfectly aligned with the CT scans. One of the volumes can be displaced by a transformation. The motivation of using simulated US is that to use an image registration algorithm to find the transformation and align the images perfectly.

For datasets two and three, the spatially tracked intraoperative US images were acquired immediately after acquiring the CT scans from resulting 15 homologous which were selected. Thus, the provided landmarks provide a silver standard ground-truth in alignment for the evaluation of image registration techniques. Since the simulated US images are aligned with CT scans, the homologous landmarks can be used for monomodal image registration of simulated US images and intraoperative US images.

We proposed a novel US simulation method to simulate US images from the CT scans. The simulated images are realistic and display US artifacts such as shadowing. Therefore, the simulated US images can be used to validate real-time US simulation methods that are used in some image registration techniques, as well as preoperative simulation and planning. Real-time US simulation techniques can be used for more accurate preoperative planning, multimodal image registration, training of physicians and clinicians for PoCUS, and to enhance US transducer designs.

In addition to the US simulation and image registration, the database can be used for other image processing applications such as denoising, segmentation, and training of supervised learning methods. As this database is the first of its kind and the similar databases with these features are rare, online availability of this database give a rudimentary access of data to researchers, clinicians, and specialists.

3.2 Acquisition and validation methods

3.2.1 Subjects

The first dataset consists of the CT images provided by TCGA-SARC [165, 166]. Patients were imaged with the CT scanner (GE LightSpeed VCT) using the protocol 5.7 CAP STANDARD-3CC/SEC. The axial slices had a thickness of $5.00mm$ and an in-slice resolution of $0.74 \times 0.74mm^2$. We extracted the lumbar vertebrae of the three patients with ID numbers TCGA-QQ-ASV2, TCGA-QQ-ASVC, and TCGA-QQ-A8VG using the 3D Slicer software.

The second database consists of the canine thoracic and cervical vertebrae phantom

data. The phantom gel was created using a mixture of water, Knox unflavored gelatin, sugar-free Metamucil psyllium fiber supplement [172], and a Rubbermaid Premium Dry Food Storage Canister to store the mixture. The canine spine model was formed using vinyl tubing, a wooden skewer, rubber O-rings, and 10 vertebrae, namely the section of the canine cervical vertebrae (C6-C7) and a section of the thoracic vertebrae (T1-T8). Figure 3.1 shows the canine vertebrae before having it immersed into the gel.

Figure 3.1: Canine vertebrae, with the vinyl tubing inserted through the cavities and the rubber O-rings between each vertebrae, prior to complete immersion into the gel.

The third database consists of the lamb vertebrae phantom data. Lamb vertebrae are most similar to human vertebrae in both the lumbar and thoracic regions [173]. Herein, we acquired our data using the L1-L5 vertebrae for a lamb. To simulate the spinal surgery, we created a surgical cavity (hemilaminectomy) on the posterior side of the vertebrae, and we performed a dorsal midline incision and removed the soft tissue covering that area (Fig. 3.2a). A means for US imaging was created by immersing the lumbar vertebrae into a gel as using the methods described earlier (Fig. 3.2b). Table 3.1 shows the summary of the subjects.

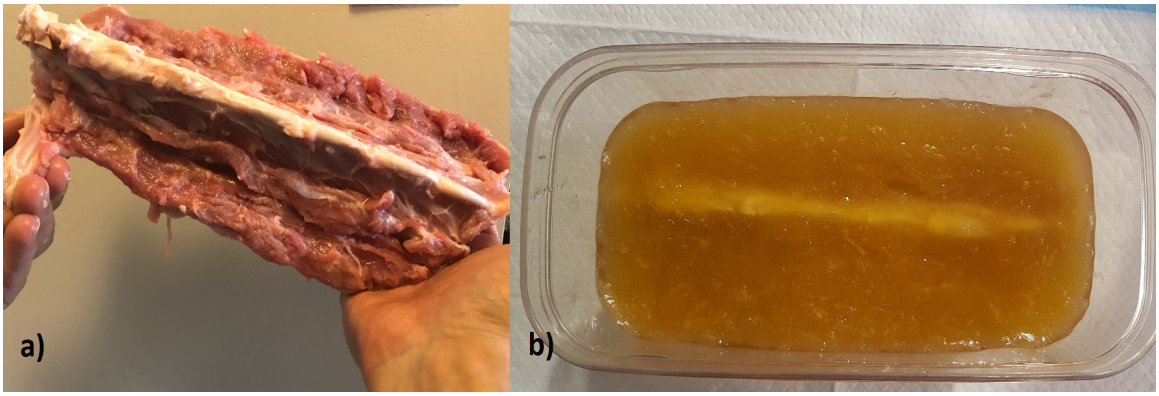

Figure 3.2: The *ex-vivo* phantom. a) Lamb lumbar vertebrae before complete immersion into the gel and after removing the tissues over the dorsal midline. b) The lamb lumbar vertebrae in the gel phantom.

Table 3.1: Subjects' information

Patient	Dataset	Vertebrae Included	CT Scan	Intra-operative US	Simulated US
TCGA-QQ-ASV2	1	L1-L5	Yes	No	Yes
TCGA-QQ-ASVC	1	L1-L5	Yes	No	Yes
TCGA-QQ-A8VG	1	L1-L5	Yes	No	Yes
Canine Phantom	2	C6-C7-T1-T2	Yes	Yes	Yes
Lamb Phantom	3	L2-L5	Yes	Yes	Yes

3.2.2 Simulation of US from CT

The Field II simulation software [174] was employed to simulate US images based on the CT scans. The simulated transducer assumes that the data was acquired with the patients lying in the prone position with the probe being perpendicular to the patients' back. The transducer consists of 192 elements with 64 active elements at a time having a frequency of 3.6 MHz and a propagation speed of 1540 m/s. The Field II simulation generated 50 raw RF lines of data from 10,000 scatterers. Each scatterer point is randomly located in a continuous space from the corresponding interpolated CT image where a Gaussian noise was added with a mean of zero and unit variance as its parameters.

After generating simulated RF signals, 2D B-mode US images were created from the

RF data, followed by reconstruction of 3D volumes. Furthermore, the volumetric US image was resampled to the resolution of the CT volumetric image and as a result, the US and CT images were aligned by construction.

The simulator does not simulate shadowing artifacts of US below the bone surface. Hence, the CT volumetric image was used as a template to correct the US image. Similar to [38], forward ray-tracing followed by backward ray-tracing were used to extract the bone surfaces from the CT image. In the forward (backward) ray-tracing, the transducer radiates the sound waves through the CT images from anterior to posterior direction (posterior to anterior direction) and when they reach tissues with intensities of T or larger, it identifies the tissue as the bone surface. Consequently, the rest of the image in the wave's direction would appear as a dark shadow. The value of T was set to 150 Hounsfield in [38] given the acquired datasets. By inspecting the bone surfaces in the CT volumes in our datasets, we found that $T = 270$ is the optimal value. Finally, the processed CT volumetric image was multiplied voxel-wise to the US volumetric image as a mask. Figure 3.3 shows a slice of the simulated US images in the first row and, the second row displays the slice of the simulated US image overlaid on their corresponding CT image. The texture of the simulated US images is similar to real US images because they show the bone surfaces as back-scatters from the US wave. It is important to note that there are shadows below the bone surfaces where the US wave could not penetrate the tissues. Inherently co-registered, the aligned CT and US images make a gold standard ground-truth to validate image registration algorithms.

3.2.3 CT scan imaging

The phantom CT scan was acquired at Concordia University's PERFORM Centre in Montreal, Canada using a GE Discovery PET/CT 690 (Waukesha, WI) scanner with a 7.4 90000133 L-Spine Survey Helical protocol. The CT image has $0.351 \times 0.351mm^2$ in-slice resolution and $0.625mm$ of slice thickness. The CT scan was executed while the patient remained in the prone position with the orientation of the head facing the machine. The CT scan was acquired for the canine phantom and the lamb lumbar vertebrae.

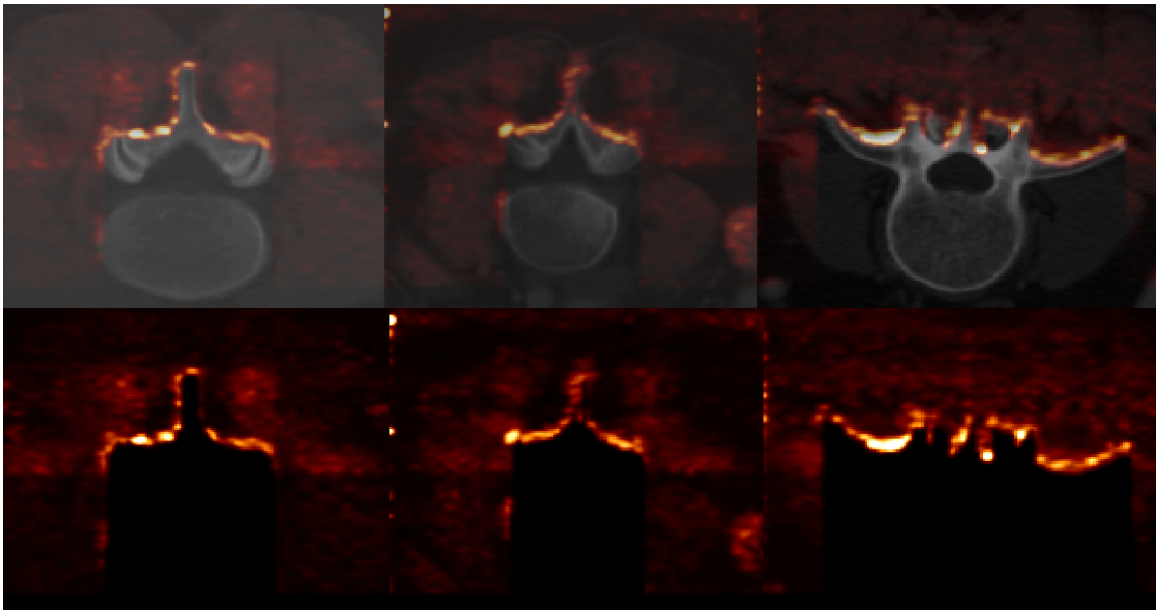

Figure 3.3: Axial view of a slice of overlaying of US images on their corresponding CT images for Patient TCGA-QQ-ASV2 (column 1), Patient TCGA-QQ-ASVC (column 2), and Patient TCGA-QQ-ASVG (column 3) in the first row and simulated US images in the second row. US probe is located at the top of the image.

3.2.4 Ultrasound imaging

Following the CT scan, we immediately acquired the spatially tracked US image of the canine and lamb phantoms to minimize potential deformations of the models. The US images were acquired with an Alpinion E-CUBE 12R ultrasound machine (Bothell, WA) at the PERFORM Centre. The SC1-4H curvilinear 2D phase array transducer was set to a frequency of 4.0 MHz and a depth of 10.0 cm. Prior to the data acquisition, the probe was calibrated by N.M. at the PERFORM Centre using the PLUS Toolkit [175]. The probe was calibrated according to the method elaborated in [176]. The RMSE of $0.2mm$ and $3.7mm$ has been achieved for the pivot and US probe calibration. As it was explained in [176], The RMSE does not reflect the true error of calibration. There are some situations where the accuracy of calibration improves while the RMSE increases. The best way to verify the calibration is the qualitative validation, so the calibrations were validated by N.M.. The US images were tracked with Northern Digital Inc. (NDI, Waterloo, ON) Polaris camera and NDI passive reflective markers. The US images were recorded with the Epiphan Systems

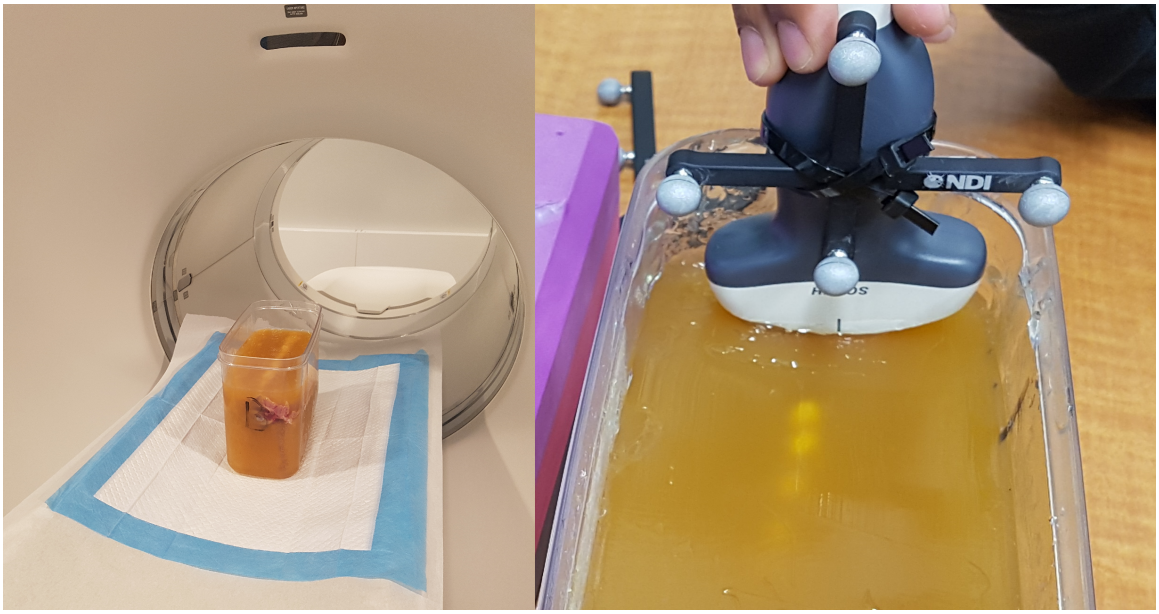

Figure 3.4: Imaging the lamb vertebrae phantom. Acquiring CT scan (left) and tracked US (right).

Inc. DVI2USB3.0 using the PLUS Toolkit [175], OpenIGTLink and the 3D Slicer as the acquisition software. Figure 3.4 shows the acquisition of the CT scan on the left side of the figure, and the acquisition of the spatially tracked US on the right of the figure.

3.2.5 Landmark selection for validation

For each of the canine and lamb phantom, two sets of 21 landmark pairs based on homologous structures between the CT and US images were selected. Each set of landmarks were independently selected by N. M. and C. B. using the 'register' software which is a part of the MINC Toolkit (https://bic-mni.github.io). The first 16 landmarks were used for the fiducial registration and the remaining landmarks for the evaluation of registration. Clear points of reference in each vertebra's surface structure such as inferior facet, superior facet, spinous process, transverse process were used to accurately select landmarks. We performed fiducial registration on each set of landmarks using affine transformations through the 3D Slicer software (https://www.slicer.org) and then resampled the US volumetric image to the CT volumetric image on each set of landmarks. We obtained Fiducial

Registration Errors (FREs) of $3.5mm$ and $3.8mm$ for the first set of landmarks (selected by N. M.) of the canine and lamb phantom respectively and FREs of $1.3mm$ and $0.9mm$ for the second set of landmarks (Selected by C. B.) of the canine and lamb phantom respectively. As a result of the registration, the US and CT images were aligned with a silver standard ground truth. In order to validate the registration, using the last 5 landmarks, the FRE of less than $0.1mm$ was achieved for the canine and lamb phantom case. The set of 21 landmark pairs with FREs of $1.3mm$ and $0.9mm$ were provided in the database for the canine and lamb phantom respectively. Figure 3.5, shows axial view of intra-operative US and simulated US of the canine phantom C8 vertebra in the first and second rows respectively. From the left column, the figure shows CT, US, and overlaid CT-US respectively.

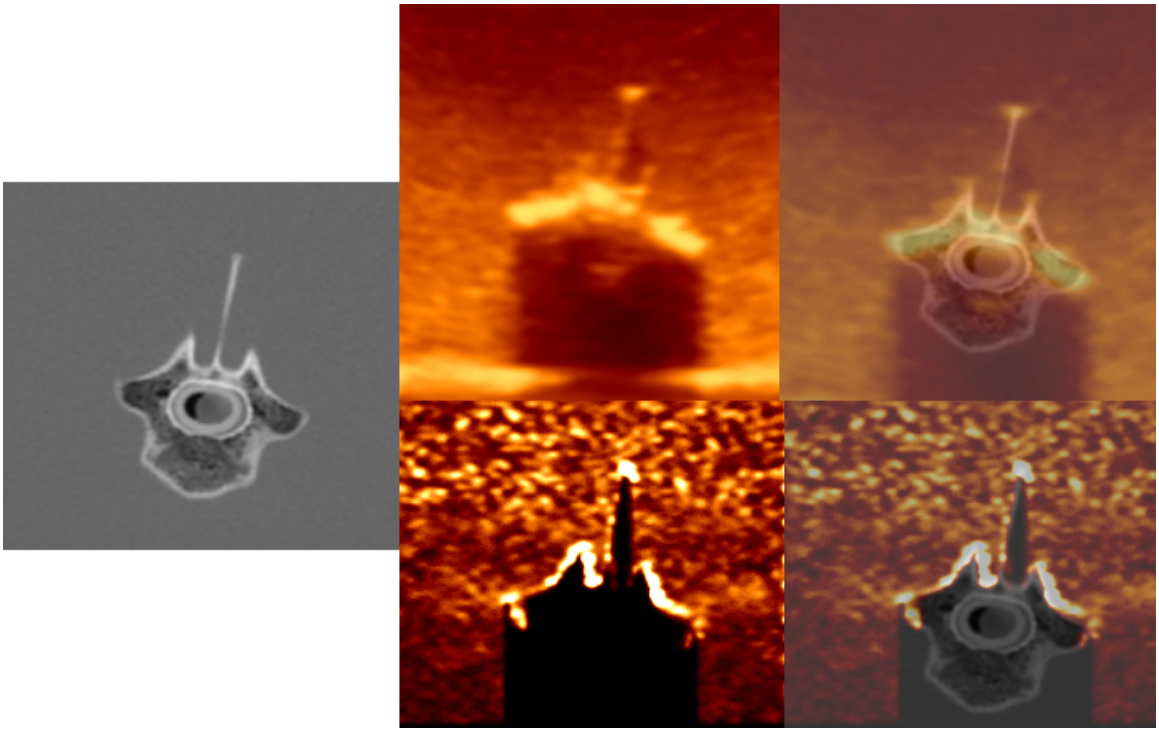

Figure 3.5: Axial view of intra-operative US and simulated US of the canine phantom C8 vertebra in the first and second rows respectively. From the left column, the figure shows CT, US, and overlaid CT-US respectively.

Figure 3.6 demonstrates the CT, US, and overlaid CT-US images of lumbar vertebrae the lamb in columns one, two, and three respectively. The rows show the sagittal view of intra-operative US and simulated US respectively. Table 3.2 shows the summary of

Table 3.2: Summary of landmark selection

Subject	Number of Landmarks	FRE for Rater 1	FRE for Rater 2
Canine Phantom	16	3.52mm	1.31mm
Lamb Phantom	16	3.82mm	0.93mm

landmark selection for canine phantom case and lamb lumbar vertebrae case.

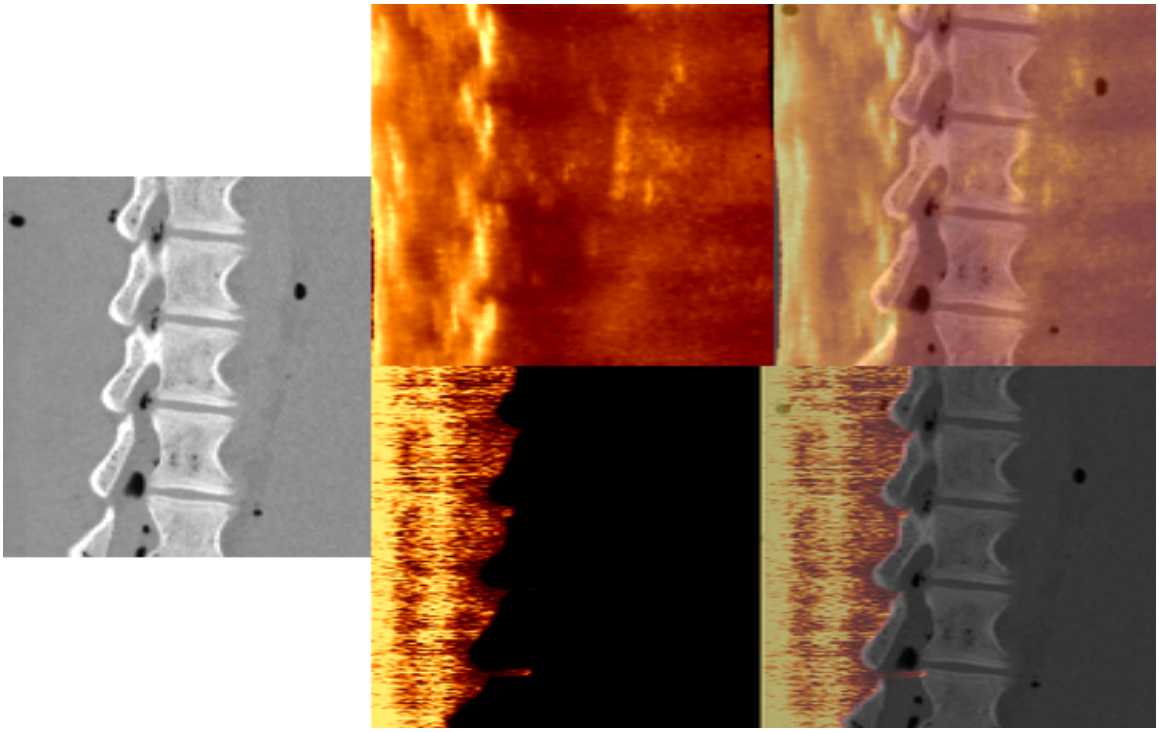

Figure 3.6: The sagittal view of the lamb lumbar vertebrae. The first row shows the intraoperative US and the second row shows the simulated US. Columns from the left to the right show CT, US, and overlaid US on the CT image respectively.

3.2.6 Data format and usage notes

All images are provided with NIFTI and MINC formats. The first dataset can be used immediately after loading NIFTI or MINC images and the CT scans and simulated US images are aligned with a gold-standard ground-truth. The second and third datasets (the canine phantom and the lamb phantom) contain the CT scan, the intraoperative US, and the simulated US. For the CT scans and intraoperative US, 21 landmarks in MNI tag files

are included so that they are aligned with a silver standard ground-truth and can also be used immediately after loading the NIFTI or MINC files. The CT scans and simulated US are aligned with a gold standard ground-truth. The data and the code are available at https://doi.org/10.5281/zenodo.2652540 and https://github.com/nimamasoumi/CT-US-Registration respectively. Potential application of the dataset is elaborated in Section S1.

3.2.7 Pre-processing of CT and US for registration

The CT and US images have different modalities and show different features of the rendered tissues. By inspecting the images, one might notice that in both modalities, the vertebral surface have rather high intensity compared to the other features in the images. Therefore by extracting the common features which are the bone surfaces in both volumes, the image registration will be facilitated.

The bone surface extraction in the images are based-on the method explained in [38] which is originally based- on [177]. Instead of applying the backward ray-tracing directly on the US images [38], it was applied on the gradient of US image from top to bottom in the axial view followed by a 3D median filter to remove some outliers. A slice of a pre-processed CT and US were demonstrated in Fig. 3.7 and Fig. 3.8 respectively. It is worth to mention that, the US pre-processing was only applied to the real US images and the simulated US images were utilized directly in the registration process.

3.2.8 Registration methods

Two different registration methods were implemented and later in this paper were compared. Both methods are rigid registration using interior-point optimization technique, namely the barrier method [171]. The formulation of transformation is provided in Section S2.

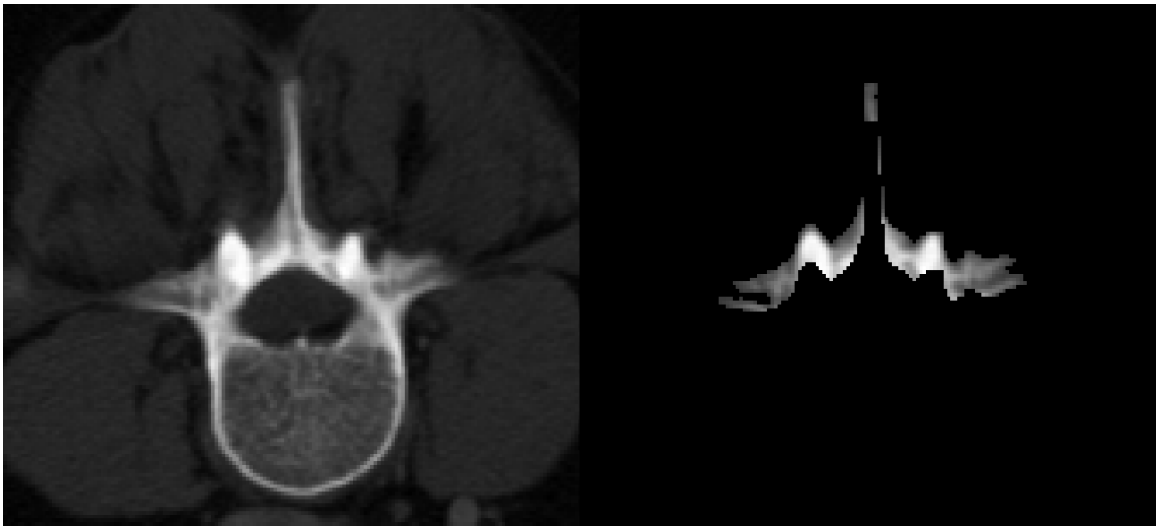

Figure 3.7: Axial view of a CT image (left) and its corresponding extracted bone surface (right).

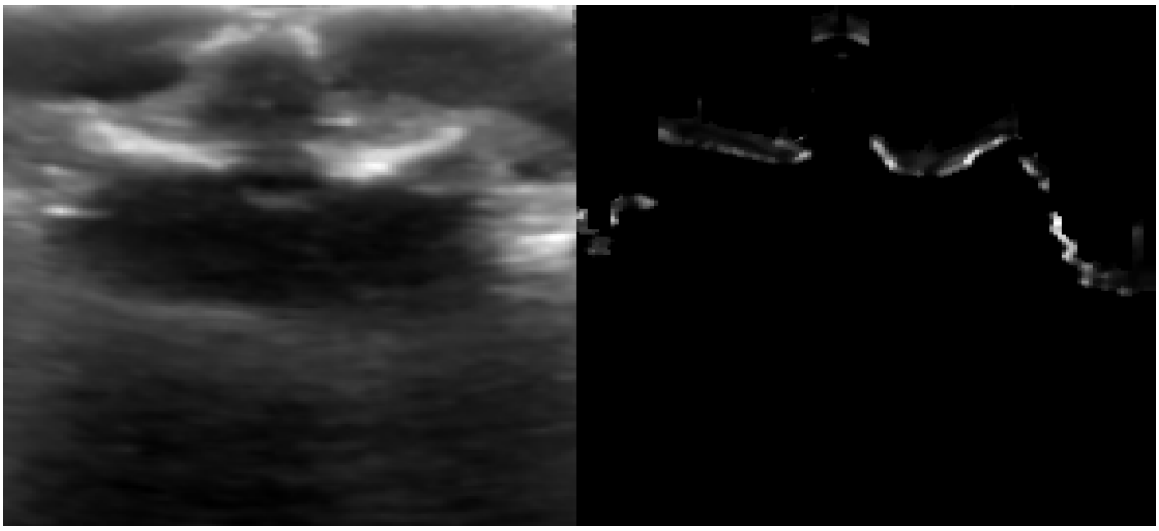

Figure 3.8: Axial view of an intra-operative US image (left) and its corresponding extracted bone surface (right).

Similarity metrics

Two different similarity metrics were employed to evaluate the dissimilarity of corresponding $7 \times 7 \times 7$ patches of images. The patches are selected randomly and the number of patches have a direct relationship on the area of bone surfaces in the US images. The first method is based-on the NCC. Given the fact that the registration is pairwise, if a patch of

the moving image is Y and a patch of the fixed image is X, the calculation of NCC over the N_P number of patches is as Eq. 3.1, 3.2:

$$D_{NCC}(Y,X) = \frac{1}{N_p}\sum_{i=1}^{N_p}(1 - |\psi(Y,X;\boldsymbol{\Omega_i})|) \tag{3.1}$$

$$\psi(Y,X) = \frac{\sum(Y.X)}{\sqrt{\sum Y^2}.\sqrt{\sum X^2}} \tag{3.2}$$

where $|\cdot|$ operator takes the absolute value. The second dissimilarity matrix is based on the Correlation Ratio and it was elaborated in [40]. The Eq. 3.3, 3.4, 3.5 are the CR equations as it was used in the optimization.

$$D_{CR}(Y,X) = \frac{1}{N_p}\sum_{i=1}^{N_p}(1 - \eta(Y|X;\boldsymbol{\Omega_i})) \tag{3.3}$$

$$1 - \eta(Y|X) = \frac{1}{N\sigma^2}\left(\sum_{t=1}^{N} Y_t^2 - \sum_{j=1}^{N_b} N_j\mu_j^2\right) \tag{3.4}$$

$$\mu_j = \frac{\sum_{t=1}^{N}\lambda_{t,j}Y_t}{N_j}, N_j = \sum_t \lambda_{t,j} \tag{3.5}$$

where $\lambda_{t,j}$ is the contribution of sample t to bin j in X, N is the total number of samples in a patch, and $\sigma^2 = Var[Y]$ is the variance of patch.

Transformation formulation

Since the salient features of the CT and US images are representing the bone surfaces and they are rigid objects, the rigid transformation was employed to estimate the misalignment of images using the optimization technique. The rigid transformation has six parameters $\theta_x, \theta_y, \theta_z, t_x, t_y, t_z$ which are the first three are the rotation along the x, y, and z axis respectively and the last three are the translation in the x, y, and z direction respectively. The parameters form a rigid transformation matrix with six degrees of freedom which commonly known as "*Euler angles*" parameterization of rigid transformation (Eq.3.6, 3.7)

$$T = \begin{pmatrix} c_y c_z & (s_x s_y c_z - c_x s_z) & (c_x s_y c_z + s_x s_z) & t_x \\ c_y s_z & (s_x s_y s_z + c_x c_z) & (c_x s_y s_z - s_x c_z) & t_y \\ -s_y & s_x c_y & c_x c_y & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.6}$$

$$s_x = \sin(\theta_x), c_x = \cos(\theta_x), s_y = \sin(\theta_y), c_y = \cos(\theta_y), s_z = \sin(\theta_z), c_z = \cos(\theta_z) \tag{3.7}$$

Optimization technique

The barrier method [171] was employed to find the optimal rigid transformation which minimizes the D_{NCC} and D_{CR} in Eq. 3.1 and Eq. 3.3 respectively. The natural logarithm was set as the barrier function using the multi-scale pyramid processing to estimate larger displacements. In the finest to most coarse resolution, maximum number of 15, 50, and 100 iterations were used (three levels).

3.3 Results

In the pre-processing, the bone surface of the CT images are extracted and they are already registered to the simulated US images as well as the real US images. In the following, the images are misaligned and using the registration methods, they were reregistered.

3.3.1 Synthetic data

Figure 3.9 shows the synthetic volume created by MATLAB. This volume was created to register to itself later in the experiments. Brighter part of the image has a range of intensities with additive Gaussian noise. The volume has general similarities to the bone surface extracted from the CT images and the US images. Registration of synthetic volume to itself is the simplified version of registering the bone surface extracted from the CT images to the US images.

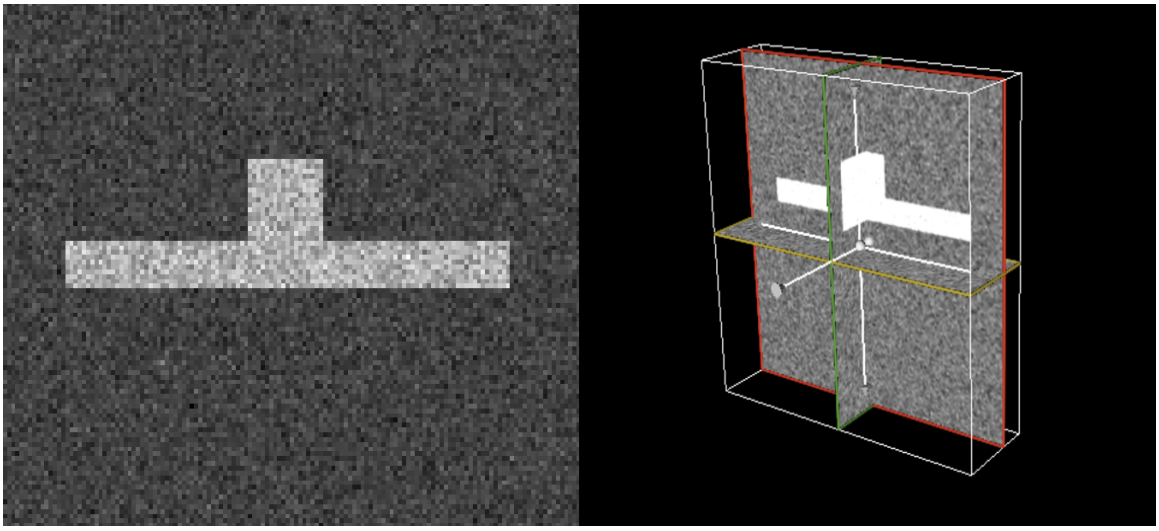

Figure 3.9: Axial view of the synthetic volume (left) and the 3D view (right).

3.3.2 Registration Problems

Given the fact that the CT images and the US images are pre-registered and aligned, in order to evaluate the registration methods described in Section 3.2.8, first the images are required to be misaligned. Therefore, 100 different registration problems were created for each misalignment intervals of $0-2$ voxels (vox), $2-4vox$, $4-6vox$, $6-8vox$, $8-10vox$, and $10-12vox$ which gives the total number of 600 registration problems for each pairs of images. Each misalignment problem was created by only rotation and the translation of images. The pairs which are selected to be misaligned and then registered are synthetic data to itself, CT to the simulation US of the subject TCGA-QQ-ASV2, CT to the simulation US of the subject TCGA-QQ-ASVC, CT to the simulation US of the subject TCGA-QQ-A8VG, CT to the simulation US of the Canine Phantom, CT to the intraoperative US of the subject the Canine Phantom, CT to the simulation US of the Lamb Phantom part 1, CT to the simulation US of the Lamb Phantom part 2, CT to the intraoperative US of the subject the Lamb Phantom part1, and CT to the intraoperative US of the subject the Lamb Phantom part 2. Note that the Lamb phantom volumes had almost twice number of slices than the other volumes and was divided into roughly two equal parts for the speed and the simplicity of registration problems.

3.3.3 Registration procedure

Having the pairs of images in the Section 3.3.2 (total ten pairs), the pre-processed CT images and the corresponding US pairs are passed to the misalignment unit. In our experiments, the pre-processed CT is fixed and the corresponding US pair was misaligned. Then the misaligned US pair is pre-processed (only for the intraoperative US) and then passed through the registration unit. The registration unit estimates a rigid transformation which aligns the images and passes the transformation to the validation unit. The validation unit compares the warping index before and after the registration.

3.3.4 Succes rate and box plots

The success rate is the number of registration problems which the methods could decrease the warping index for each pair and each error (misalignment) interval over the total number of registration for that pair and that interval (which is 100 for all intervals and pairs here). The following box plots are comparing the CR and NCC for each pair in registration problems where the methods could decrease the warping index. Figure 3.10 is the result of registration for the CT and intraoperative US of the Canine Phantom. Figure 3.11 shows the success rate for the corresponding pair. In these pairs, NCC outperformed CR in both decreasing the warping index and having better success rate. Figure 3.12 is the result of registration for the synthetic data to itself. Figure 3.13 shows the success rate for the corresponding pair. Both methods have high success rate in which CR had slightly better performance.

Figure 3.14 is the result of registration for the CT and simulated US of the Subject TCGA-QQ-ASVC. Figure 3.15 shows the success rate for the corresponding pair. The CR has slightly better performance in decreasing the warping index and generally better success rate.

Figure 3.16 is the result of registration for the CT and simulated US of the Subject TCGA-QQ-A8VG. Figure 3.17 shows the success rate for the corresponding pair. The CR has slighty better performance in decreasing the warping index and better success rate.

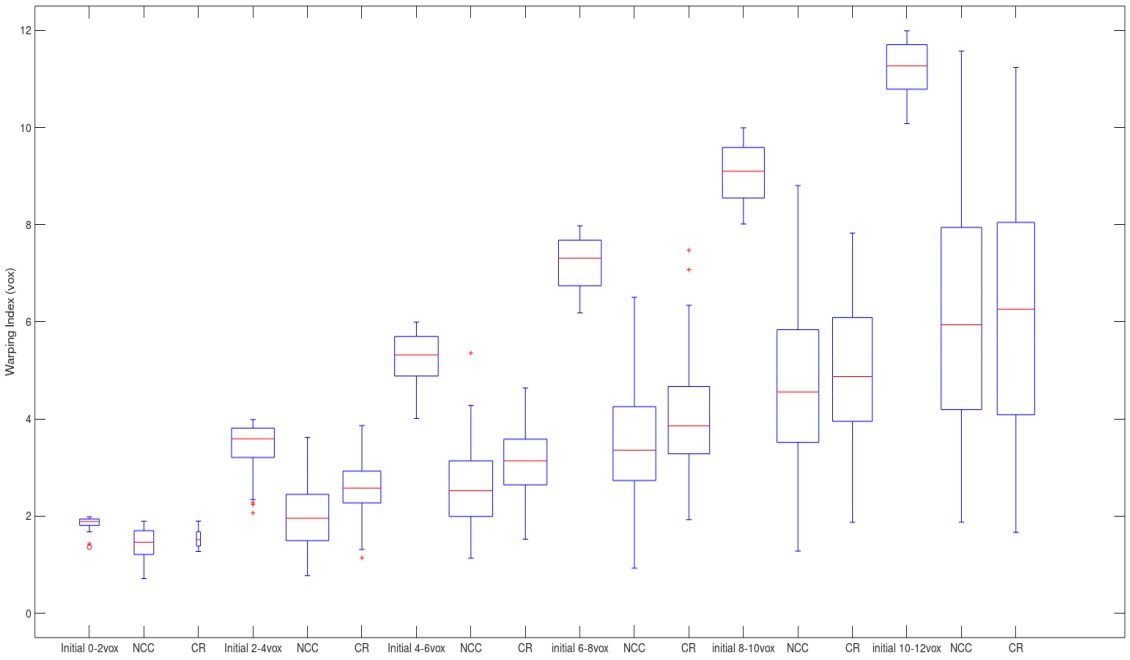

Figure 3.10: Comparison of NCC and CR methods in registration of the CT and intraoperative US of the Canine Phantom.

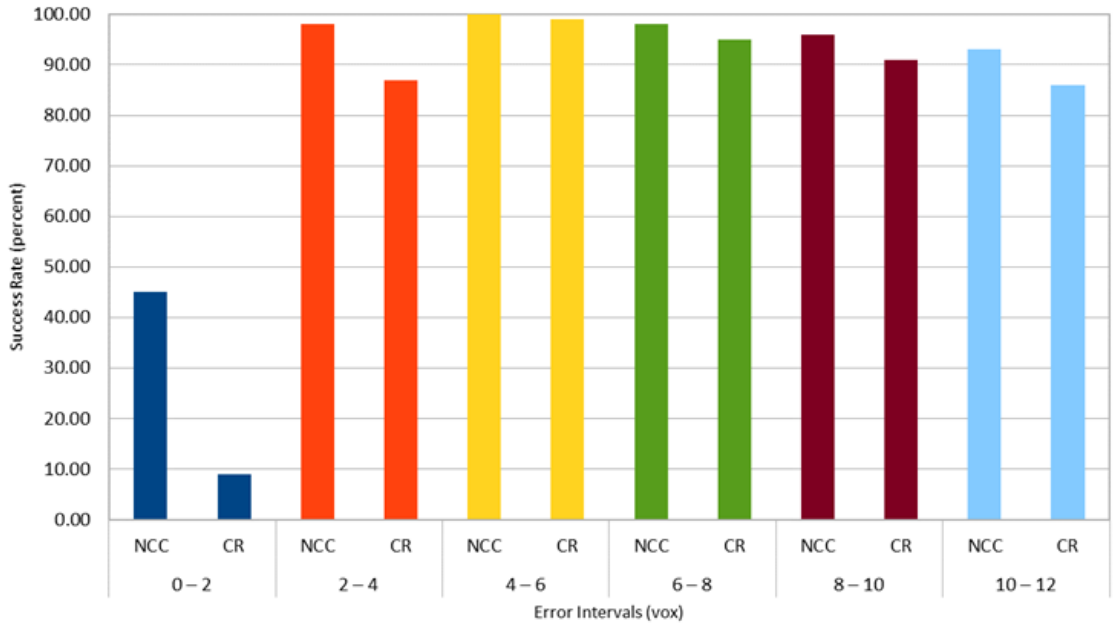

Figure 3.11: Success rate of the registration of CT and intraoperative US of the Canine Phantom for the NCC and CR.

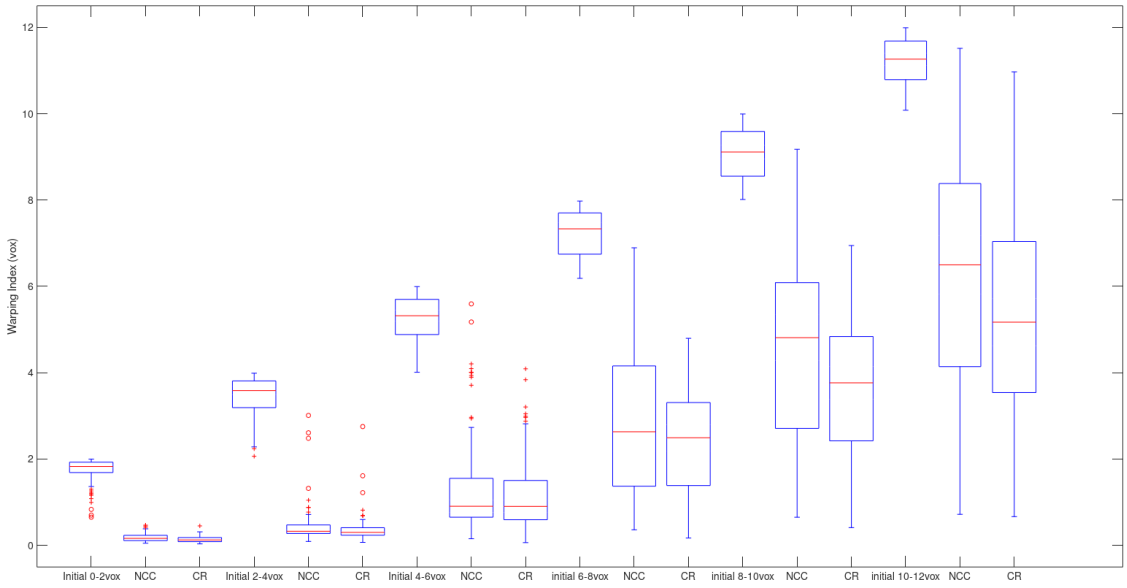

Figure 3.12: Comparison of NCC and CR methods in registration of the synthetic data to itself.

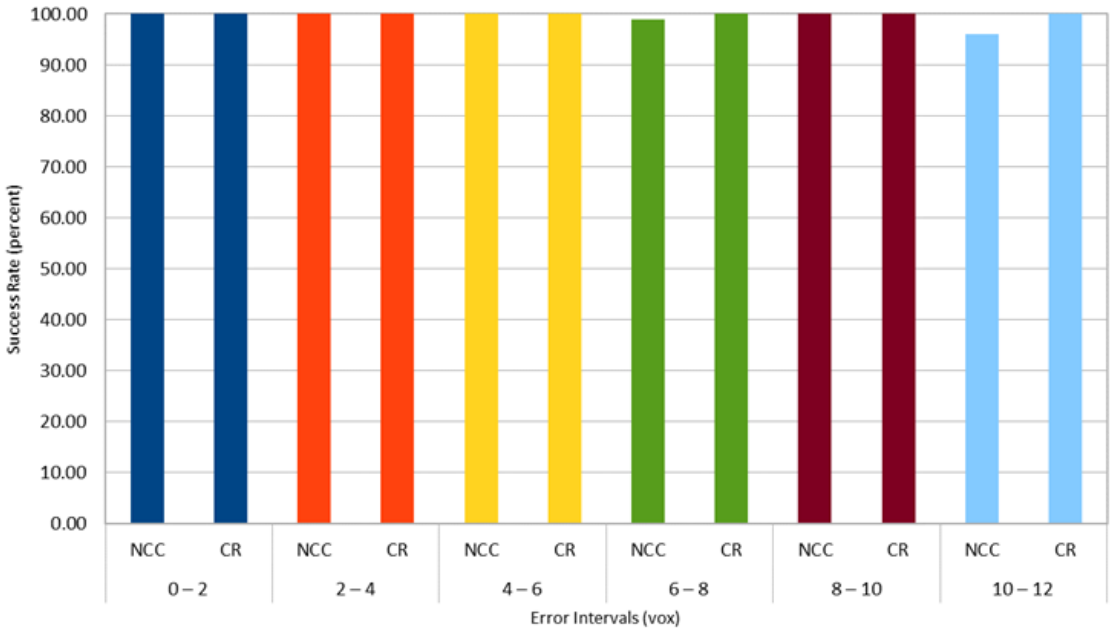

Figure 3.13: Success rate of the registration of synthetic data to itself for the NCC and CR.

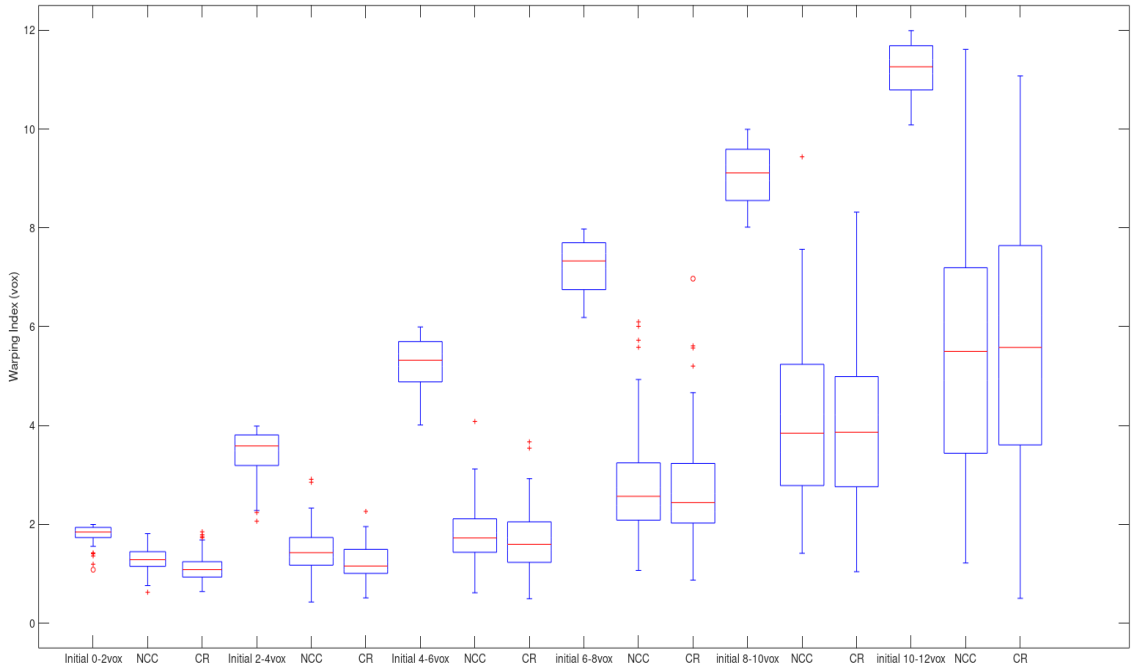

Figure 3.14: Comparison of NCC and CR methods in registration of the CT and simulated US of the Subject TCGA-QQ-ASVC.

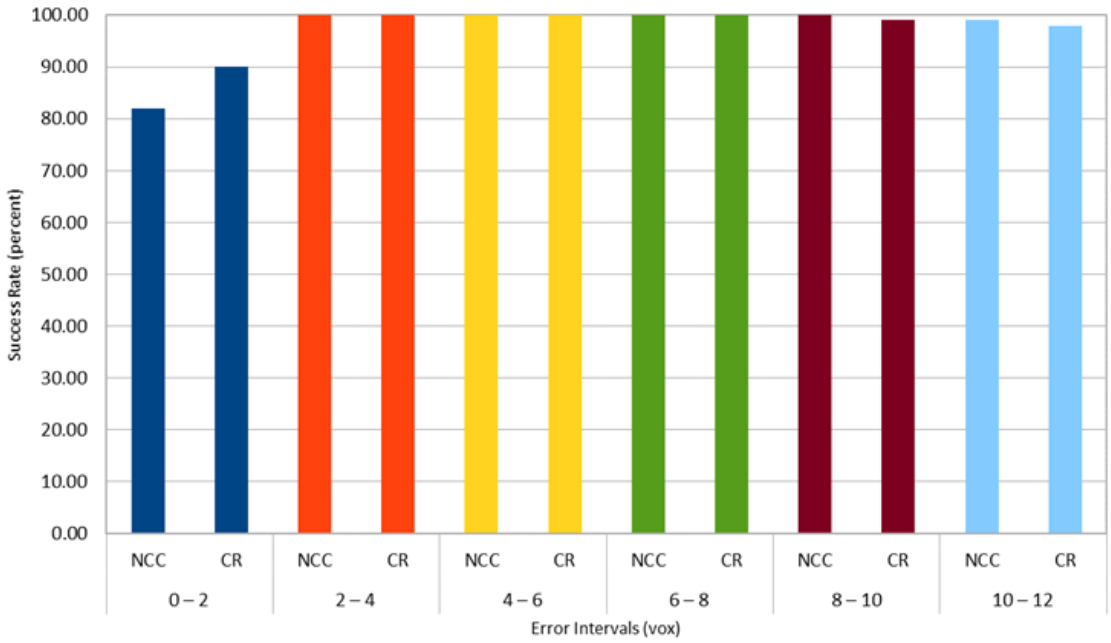

Figure 3.15: Success rate of the registration of CT and simulated US of the Subject TCGA-QQ-ASVC for the NCC and CR.

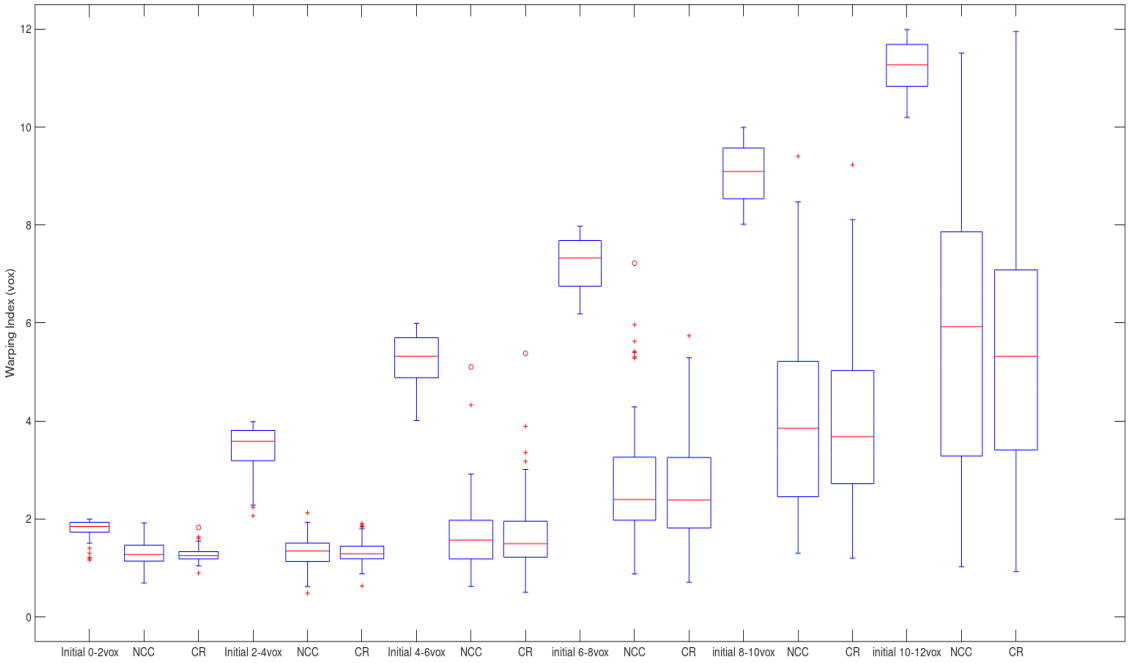

Figure 3.16: Comparison of NCC and CR methods in registration of the CT and simulated US of the Subject TCGA-QQ-ASVC.

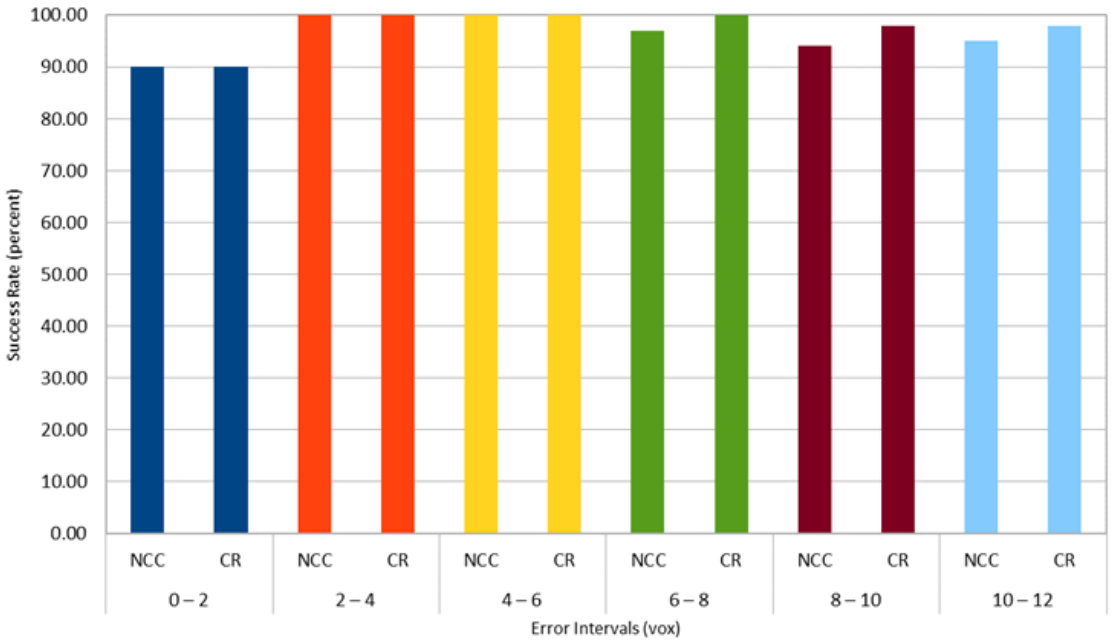

Figure 3.17: Success rate of the registration of CT and simulated US of the Subject TCGA-QQ-A8VG for the NCC and CR.

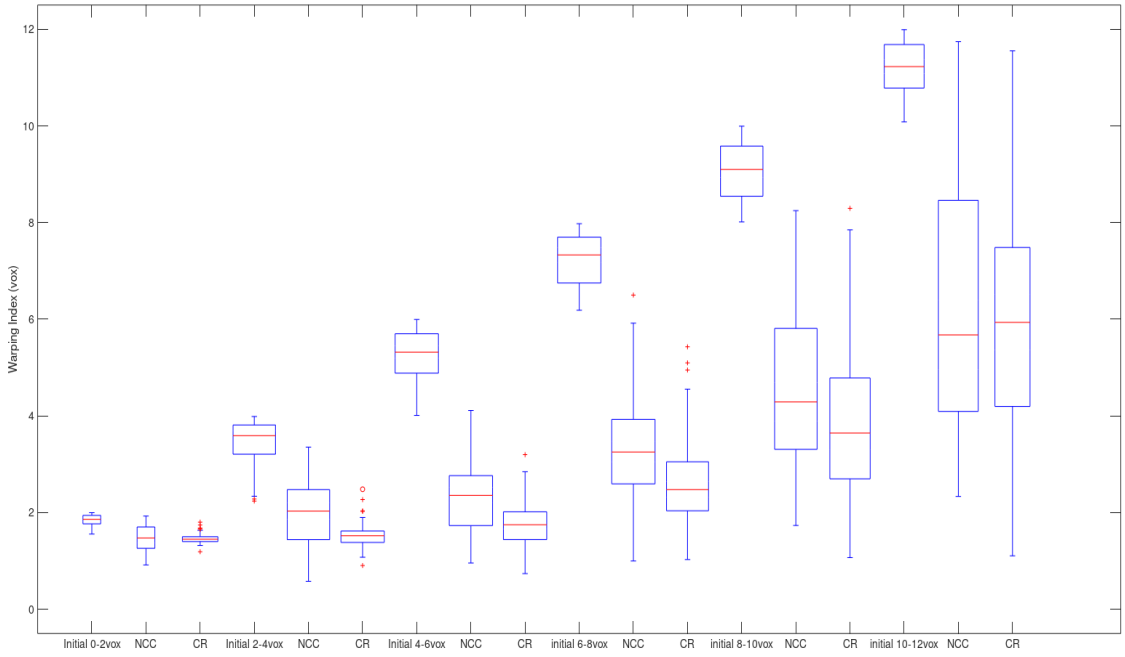

Figure 3.18: Comparison of NCC and CR methods in registration of the CT and simulated US of the Canine Phantom.

Figure 3.18 is the result of registration for the CT and simulated US of the Canine Phantom. Figure 3.19 shows the success rate for the corresponding pair. Except for the $10-12vox$ error interval, CR had a better performance in decreasing the warping index while for the last three intervals, NCC had better success rate and for the first three intervals, CR had better success rate.

Figure 3.20 is the result of registration for the CT and simulated US of the Lamb Phantom part 1. Figure 3.21 shows the success rate for the corresponding pair. The NCC method had slightly better performance in decreasing the warping index while the CR method had generally better success rate.

Figure 3.22 is the result of registration for the CT and simulated US of the Lamb Phantom part 2. Figure 3.23 shows the success rate for the corresponding pair. Except for the $10-12vox$ error interval, CR had a better performance in decreasing the warping index and it had a better success rate.

Figure 3.24 is the result of registration for the CT and intraoperative US of the Lamb Phantom part 1. Figure 3.25 shows the success rate for the corresponding pair. The NCC

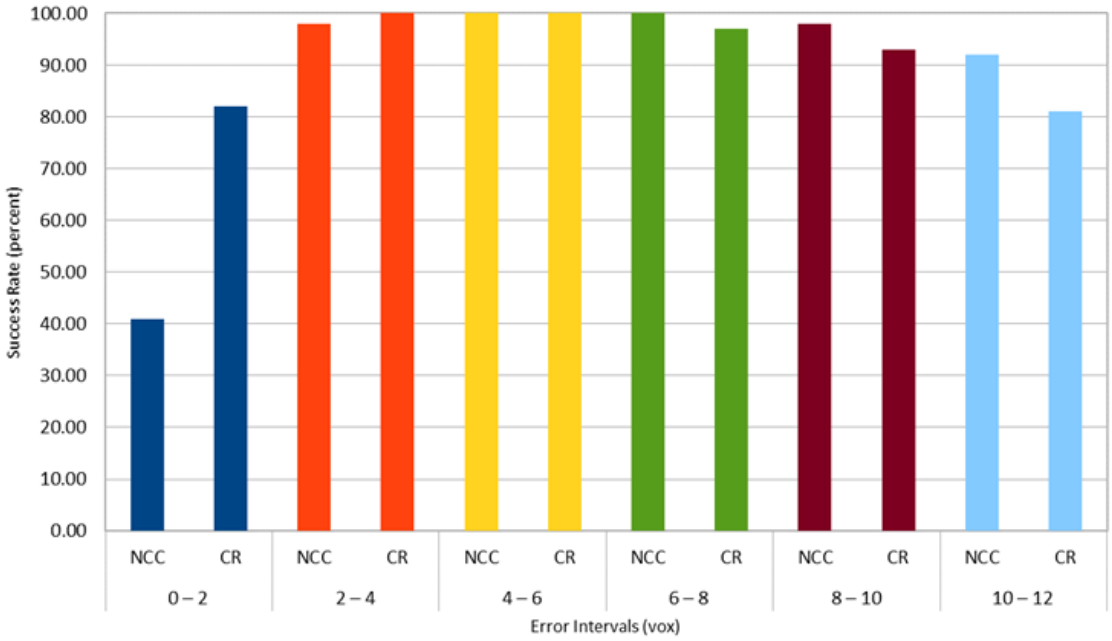

Figure 3.19: Success rate of the registration of CT and simulated US of the Canine Phantom for the NCC and CR.

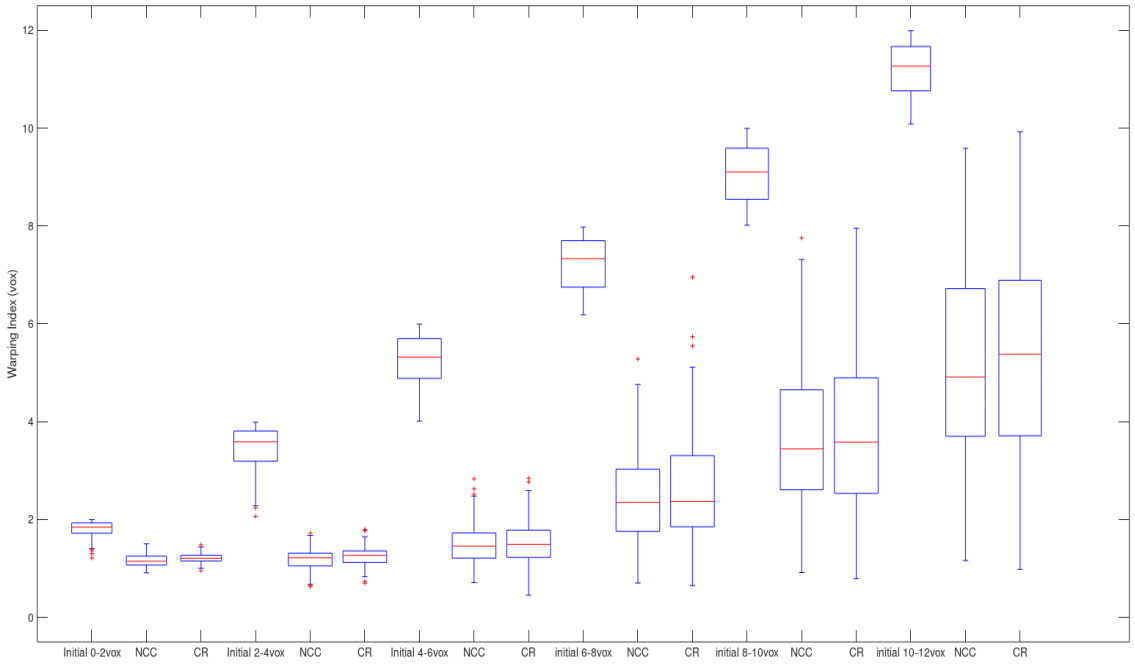

Figure 3.20: Comparison of NCC and CR methods in registration of the CT and simulated US of the Lamb Phantom part 1.

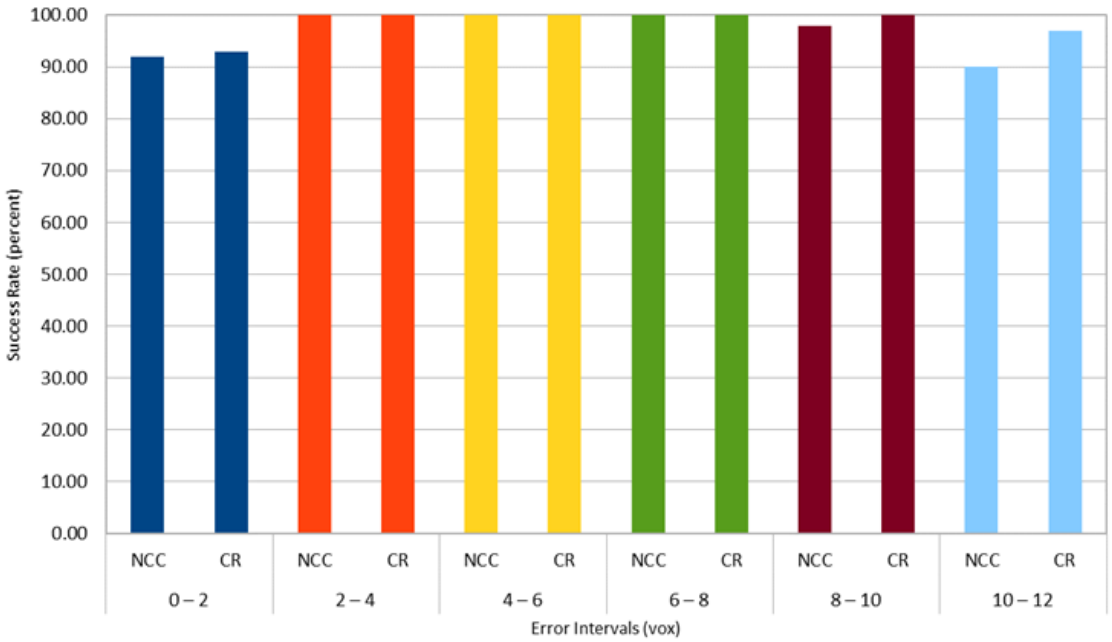

Figure 3.21: Success rate of the registration of CT and simulated US of the Lamb Phantom part 1 for the NCC and CR.

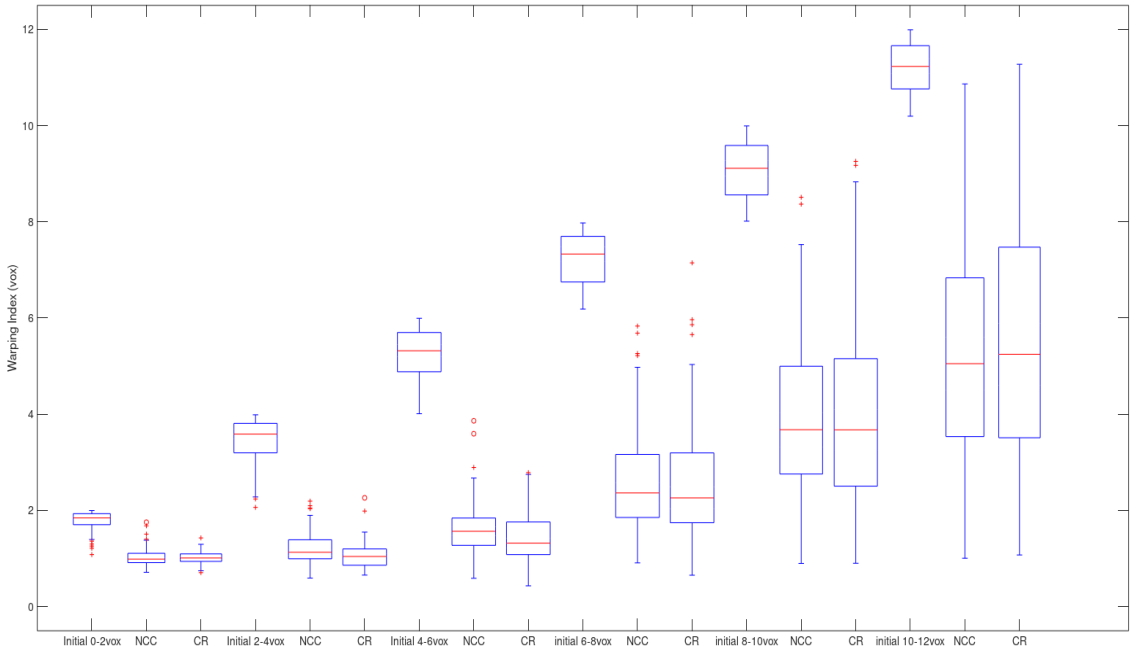

Figure 3.22: Comparison of NCC and CR methods in registration of the CT and simulated US of the Lamb Phantom part 2.

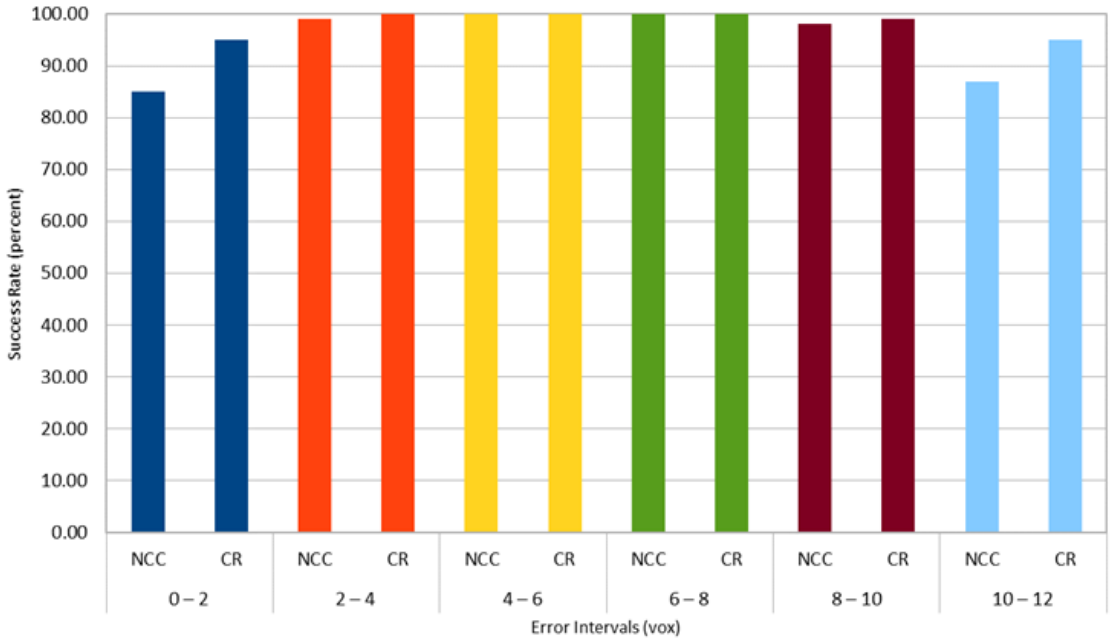

Figure 3.23: Success rate of the registration of CT and simulated US of the Lamb Phantom part 2 for the NCC and CR.

method outperformed the CR method in both decreasing the warping index and have better success rate.

Figure 3.26 is the result of registration for the CT and intraoperative US of the Lamb Phantom part 2. Figure 3.27 shows the success rate for the corresponding pair. The NCC method outperformed the CR method in both decreasing the warping index and have better success rate.

3.4 Discussions

In Section 3.2.2, a method was elaborated to produce simulated US images using the Field II package [174]. Since this package does not allow modeling the full reflection of the wave at the bone surface, we set the scatterer intensities below the bone surface to zero. This creates sharp edge artifacts at the bone surface, which can be considered as a limitation of this database.

The NCC registration method could show better performance in the registration prob-

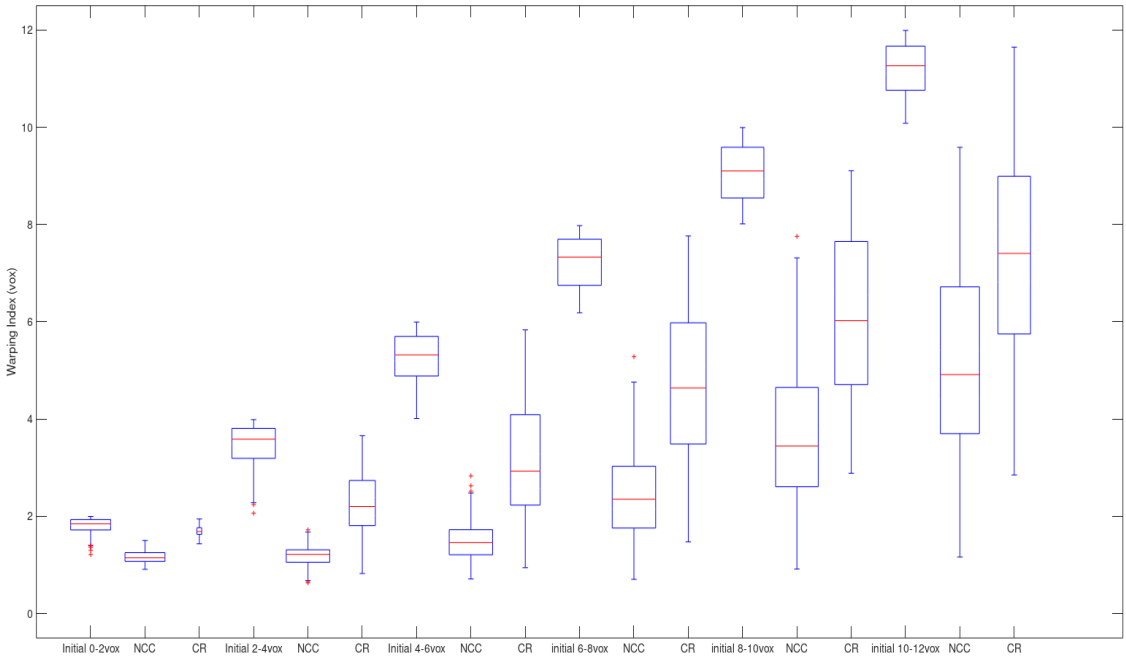

Figure 3.24: Comparison of NCC and CR methods in registration of the CT and intra-operative US of the Lamb Phantom part 1.

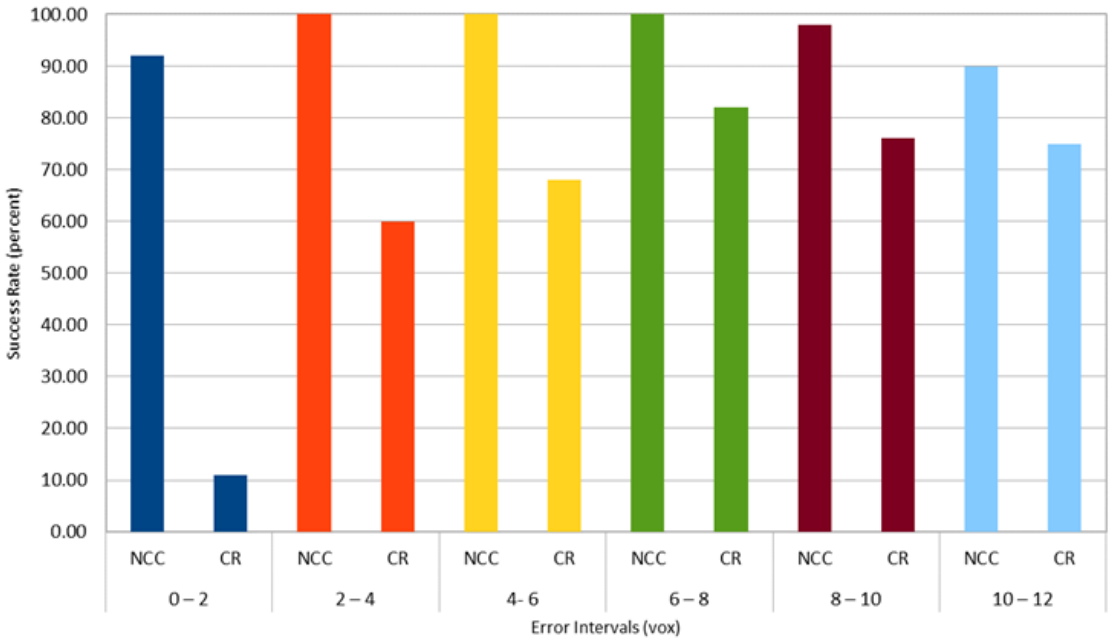

Figure 3.25: Success rate of the registration of CT and intraoperative US of the Lamb Phantom part 1 for the NCC and CR.

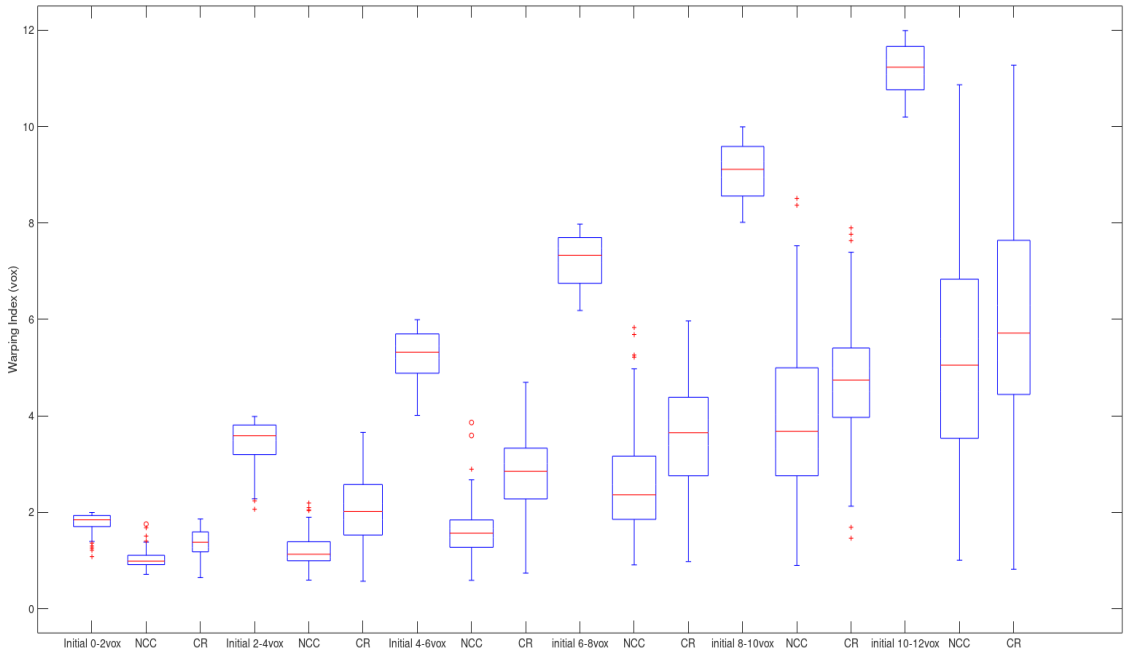

Figure 3.26: Comparison of NCC and CR methods in registration of the CT and intraoperative US of the Lamb Phantom part 2.

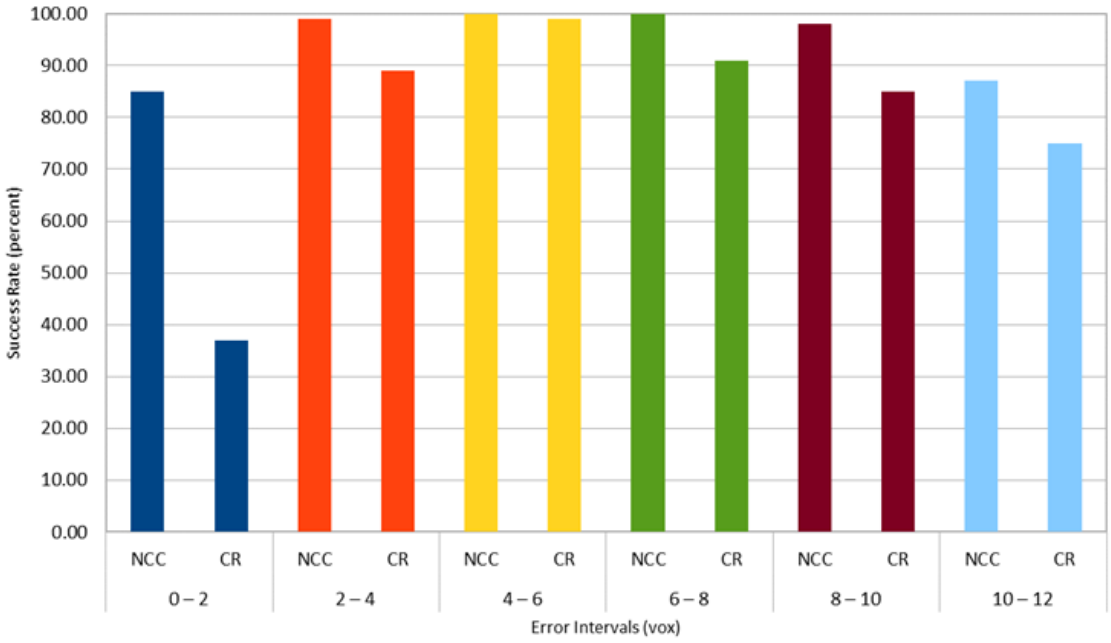

Figure 3.27: Success rate of the registration of CT and intraoperative US of the Lamb Phantom part 2 for the NCC and CR.

lems where real US image is involved while the CR generally had better performance where simulated US images are registered to the CT images. The registration algorithms were executed in a machine with 4 CPU workers and a machine with 12 CPU workers. Obviously the registrations that were implemented by the 12 CPU workers were 2 to 3 times faster. The NCC method was implemented in MATLAB (MathWorks, Inc., Natick, USA), and the CR implementation was in MATLAB and C++ to decrease the execution time of image binning.

In the small error intervals, especially $0 - 2vox$, compared to NCC, CR did not have good success in the alignment of CT and intra-operative US images. One reason is that both Canine Phantom and the Lamb Phantom have pre-existing initial misalignment even after fiducial registration. This may leave CR method with no improvement of the alignment which reduces the success rate.

3.5 Summary

In this chapter, a database of CT, intraoperative US, and simulated US were presented. This database consists of 3 datasets and in total 5 subjects. For each subject, the simulated US from CT was provided. The database provides a resource for evaluating image registration techniques. The simulated data have two applications. First, they provide the gold standard ground–truth which is difficult to obtain with ex–vivo and in–vivo data for validating US–CT registration methods. Second, the simulated US images can be used to validate real–time US simulation methods, since our database is simulated using the computationally expensive but physically realistic Field II package. Two patched-based rigid image registration methods were implemented to register the CT and US images after misaligning them. The methods are employing NCC and CR as the similarity metrics and the barrier method to optimize similarity of images. The results of registration show that the methods were successful in aligning the pre-processed CT and US images by decreasing the warping index. Given this fact, the proposed image registration techniques can be useful to potentially improve ultrasound-guided interventions of the spine.

Chapter 4

DiffeoRaptor: diffeomorphic inter-modal image registration using RaPTOR

Diffeomorphic image registration allows the computation of a smooth and invertible deformation field, and thus ensures that salient image features are not lost after image resampling with the obtained deformation fields. A key step in many clinical applications, diffeomorphic image registration can be employed in quantifying inter-subject variability of brain [179], studying Alzheimer's disease [180], statistical shape analysis [181], brain atlas construction [182], and estimation of tissue deformation for surgery [183].

Several studies have proposed diffeomorphic algorithms to perform intra-modal/contrast image registration. Beg *et al.* [3] implemented the LDDMM to register brain MRIs of Alzheimer's and Schizophrenia patients, but their computational cost was high. Later, many algorithms were proposed to make the computation more efficient. Vialard *et al.* [184] shortened the computational time by employing geodesic shooting to register 3D MRI scans of fetus brains. Zhang *et al.* [15] proposed FLASH to perform inter-subject registration of 3D brain MRIs. Similar to [184], they also employed geodesic shooting and improved the efficiency by performing the calculations in a band-limited space. Wu *et al.* [185] implemented Cross-Correlation (CC)-based LDDMM for fast brain image regis-

tration via GPU acceleration.

In general, performing diffeomorphic image registration with iterative optimization can be computationally expensive and time-consuming. Therefore, a number of DL-based algorithms were designed to tackle this problem [8]. In [187], the comparison with multiple registration tasks suggests that compared with DL-based techniques, classic registration methods still have good performance and can offer satisfactory speed with the option of parallel computing.

In the last decade, several groups have attempted to design inter-modal diffeomorphic image registration techniques in various applications. Mitra *et al.* [1] proposed an inter-modal diffeomorphic algorithm to register 2D transrectal ultrasound images to MR slices. Kutten *et al.* [2] implemented the MI-based LDDMM on a Hamiltonian framework to register CLARITY images. Reaungamornrat *et al.* [4] proposed a MIND Demons which is based on SyN [5], diffeomorphic Demons [6], and MIND features [7] to perform deformable MRI-CT registration for image-guided surgery. However, inter-modal image registration remains a challenging task in medical image registration. In general, the algorithms should show a certain degree of robustness against intensity inhomogeneities, noise, and image artifacts. Moreover, the algorithms should be time-efficient for real clinical applications. To address some of these requirements, Rivaz *et al.* [40] proposed RaPTOR to register 3D inter-modal images of the BITE database [127]. Later in [41], an affine version of RaPTOR was used to successfully register inter-modal images of RESECT [129] and BITE [127] databases. Recently in [42], a rigid version of RaPTOR was employed to register preoperative CT and intraoperative US images of lumbar vertebrae.

This study intends to design a diffeomorphic algorithm to perform intra- and inter-modal image registration [178]. In [40–42], it was shown that RaPTOR could successfully align images with different modalities. In [15], it was shown that FLASH could perform computationally efficient diffeomorphic registrations compared to vector momentum LDDMM [43]. However, RaPTOR and FLASH have the following drawbacks. First, RaPTOR uses B-spline as the transformation model which does not guarantee a smooth inverse transformation. Second, FLASH uses Sum-of-Squared Differences (SSD) that is

unable to directly measure the similarity between images of different modalities and contrasts [44]. Therefore, FLASH cannot be used to perform inter-modal/contrast image registration. Third, FLASH does not use multiresolution image pyramids to tackle larger deformations which is a standard approach in many inter-modal image registration methods. In [178], we proposed DiffeoRaptor, a novel algorithm to bring together the benefits of RaPTOR and FLASH while mitigating their drawbacks. We decided to build on RaPTOR to propose a novel diffeomorphic registration technique. Other excellent choices are normalized Gaussian fields (NGF) and MIND. FLASH framework was selected in favor of other diffeomorphic approaches. Because it is based on the well-established LDDMM framework. The performance of DiffeoRaptor was demonstrated in three applications, including 1) healthy individual MRI-to-template registration; 2) registration between AD and healthy brains, as well as brain scans at different stages of AD; 3) nonlinear registration of MR and CT abdominal data. The contributions of [178] are three-fold:

1. Proposing a diffeomorphic image registration framework using RaPTOR.

2. Devising inter-modal/contrast image registration with geodesic shooting in the bandlimited space of velocity fields.

3. Employing gradient descent (GD) with momentum to improve the convergence in contrast to classical GD optimization in FLASH and RaPTOR .

Our results show that DiffeoRaptor could achieve 1) better alignment of brain and abdominal images compared to Mattes MI+SyN, NiftyReg [188], and FLASH as assessed by Dice scores; 2) smoother deformation fields compared to Mattes MI+SyN and NiftyReg in the alignment of brain MR images, and 3) comparable computation time with FLASH while performing more challenging tasks.

4.1 Methodology

In this section, backgrounds of bandlimited space of velocity fields, bandlimited geodesic shooting, and formulation of RaPTOR metric are presented. Then, the formulation of Dif-

feoRaptor objective function is derived. Lastly, the optimization technique to minimize the objective function is detailed.

4.1.1 Space of bandlimited velocity fields

In pairwise diffeomorphic image registration, the reference image $X \in \Omega$ and the source image $Y \in \Omega$ are given. Ideally, the objective is to find a mapping $\phi \in \mathrm{Diff}(\Omega)$ such that $X \circ \phi \approx Y$ and $Y \circ \phi^{-1} = X$. Diffeomorphisms $\phi : \Omega \to \Omega$ are a smooth mapping that has a smooth inverse ϕ^{-1}. The tangent vector space at the identity $id \in \mathrm{Diff}(\Omega)$ over the space of diffeomorphisms is defined as $V = T_{\mathrm{id}}\mathrm{Diff}(\Omega)$. Given V, the space of bandlimited velocity fields $\widetilde{V}$ was constructed and proper Lie algebra in this space was defined in [15]. Time series $t \in [0, 1]$ of dffeomorphisms $\phi_t \in \mathrm{Diff}(\Omega)$ is created in the process of solving an ordinary differential equation (ODE). The time series of bandlimited velocity fields $\tilde{v}_t \in \widetilde{V}$ are related to ϕ_t^{-1} by Eq (4.1).

$$\frac{d\phi_t^{-1}}{dt} = -D\phi_t^{-1} \cdot \iota\left(\tilde{v}_t\right) \tag{4.1}$$

where D is the derivative operator and $\iota : \widetilde{V} \to V$ is the inverse Fourier transform from the bandlimited space to the space of dense velocity fields [15]. The geodesic shooting is the process of integrating the geodesic path of diffeomorphisms forward in time which is uniquely determined with the velocity $\tilde{v}_0$ in $t = 0$. The geodesic evolution equation in the discrete Fourier space is defined in Eq (4.2).

$$\frac{\partial \tilde{v}_t}{\partial t} = -\widetilde{K}\left[(\widetilde{D}\tilde{v})^T \star \tilde{m}_t + \widetilde{\Gamma}(\tilde{m}_t \otimes \tilde{v}_t)\right] \tag{4.2}$$

where K is the smoothing operator which is the inverse of the differential operator L. There is an in-depth discussion of possible choices of L in [3, 189, 190]. In this paper, it is set $L = (-\alpha\Delta + I)^c$ similar to [3, 15] where Δ is the Laplacian operator. $\widetilde{K}$ is the smoothing operator in the bandlimited space [15], $\star$ is the truncated auto-correlation, $\widetilde{\Gamma}$ is the discrete divergence, $\tilde{m}_t = \widetilde{L}\tilde{v}_t$ is the momentum, $\widetilde{L}$ is the representation of L in the frequency domain, $\otimes$ denotes the tensor product, and $\widetilde{D}$ is an operator that computes the

spatial gradient in the bandlimited Fourier space [15].

4.1.2 Geodesic shooting in the bandlimited space

By setting the geodesic shooting as the constraint of the cost function, it does not require calculating the velocity fields $\tilde{v}_t$ and diffeomorphisms ϕ_t in a dense time grid and it suffices to calculate the initial velocity $\tilde{v}_0 \in \widetilde{V}$. The cost function for FLASH was defined as Eq. (4.3).

$$E(\tilde{v}_0) = \frac{1}{2\sigma^2}\left\|Y \circ \phi_1^{-1} - X\right\|^2 + \langle \widetilde{L}\tilde{v}_0, \tilde{v}_0 \rangle, \quad \text{s.t. Eq. (4.2)} \tag{4.3}$$

where σ is the noise variance, $\|\cdot\|$ is the norm operator in the space Ω, $\widetilde{L}$ is the inverse of $\widetilde{K}$, and $\langle , \rangle$ is the inner-product in the space $\widetilde{V}$ [15]. Gradient of the energy function E can be calculated as in Eq. (4.4) for the minimization of cost.

$$\nabla_{\tilde{v}_1} E = \nu\left(- K\left(\frac{1}{\sigma^2}(Y \circ \phi_t^{-1} - X) \cdot \nabla(Y \circ \phi_1^{-1})\right)\right) \tag{4.4}$$

where $\nu : V \to \widetilde{V}$ is the projection mapping to the bandlimited space of velocity fields and K is the smoothing operator.

4.1.3 RaPTOR

One possible choice for the similarity metric is the CR [47]. For challenging inter-modal image registration tasks, calculation of CR needs to be robust and possibly time-efficient. RaPTOR is a dissimilarity metric that is based on CR [40] and addresses the shortcomings of CR [47]. RaPTOR and its derivative can be calculated as in Eq. (4.5). It calculates CR in local patches Θ. Instead of calculating the iso-sets of X, the histogram of X over N_b bins are calculated and then Parzen windowing was applied to make the bins continuous

and differentiable.

$$1-\eta(Y|X)=\frac{1}{N\sigma^2}\Bigg(\sum_{i=1}^{N}y_i^2-\sum_{j=1}^{N_b}N_j\mu_j^2\Bigg) \tag{4.5a}$$

$$\mu_j=\frac{\sum_{i=1}^{N}\lambda_{ij}y_i}{N_j},\ N_j=\sum_i\lambda_{ij} \tag{4.5b}$$

$$\mathrm{RaPTOR}(Y,X)=\Psi(Y,X)=\frac{1}{N_p}\sum_{i=1}^{N_p}(1-\eta(Y|X;\Theta_i)) \tag{4.5c}$$

$$\nabla_\varphi\Psi=\frac{\partial\Psi}{\partial\varphi}=\frac{\partial\phi}{\partial\varphi}\cdot\frac{\partial Y}{\partial\phi}\cdot\frac{\partial\Psi}{\partial Y} \tag{4.5d}$$

$$\begin{aligned}\frac{\partial(1-\eta)}{\partial y_i}=\frac{2}{N\sigma^2}\Bigg(y_i-\lambda_{i,j-1}\mu_{j-1}-\lambda_{ij}\mu_j-\\ \frac{1}{(N-1)\sigma^2}(y_i-\mu)\Bigg(\sum_{a=1}^{N}y_a^2-\sum_{c=1}^{N_b}N_c\mu_c^2\Bigg)\Bigg)\end{aligned} \tag{4.5e}$$

where N is the number of pixels in a image patch Θ_i, $\sigma^2=\mathrm{Var}[Y;\Theta_i]$ is the variance of a patch i in Y, y_i is the intensity of sample i in image Y, let j and $j-1$ be the closest bins to sample x_i (intensity of sample i in X) then according to its distance to these bins centers, λ_{ij} is the linear contribution of x_i to the bin j, N_p is the number of patches, φ is the parameter of transformation ϕ, and $\mu=E[Y]$ is the average value of Y. $\eta(Y|X)$ can measure the functional dependence between the input images. When there is no functional dependence $\eta(Y|X)=0$ and when $\eta(Y|X)=1$ there is a deterministic relationship between X and Y. Calculating gradient of RaPTOR analytically enables efficient minimization of the dissimilarity metric using gradient-based optimization and employing the outlier suppression technique elaborated in [40].

4.1.4 DiffeoRaptor

The energy function in Eq. (4.3) can be generalized to the form in Eq. (4.6).

$$E(\tilde{v}_0)=\mathrm{dist}\big(Y\circ\phi_1^{-1},X\big)+\langle\widetilde{L}\tilde{v}_0,\tilde{v}_0\rangle,\quad \text{s.t. Eq. (4.2)} \tag{4.6}$$

where $dist(,)$ is a normalized distance function or a dissimilarity function. DiffeoRaptor

is the cost function in the form of Eq. (4.6) with the RaPTOR defined in Eq. (4.5) as the dissimilarity function. So it takes the form in Eq. 4.7.

$$E(\tilde{v}_0) = \Psi\big(Y \circ \phi_1^{-1}, X\big) + \langle \widetilde{L}\tilde{v}_0, \tilde{v}_0 \rangle, \quad \text{s.t. Eq. (4.2)} \tag{4.7}$$

Eq. (4.4) is no longer valid for Eq. (4.7) and the gradient of cost function needs to be calculated for the optimization. A similar approach to [3] is taken to calculate $\partial_u \Psi$, the variation of cost in Eq. (4.7) with respect to the velocity $u = D\phi_1^{-1}$ which is obtained by taking the derivative of ϕ_1^{-1}.

Given the fact that we are working with image intensities in a grid according to Eq. (4.5), the variation of energy $\partial_u E$ takes the form $\partial_u E = \langle \nabla_u E, u \rangle_{V_g}$ and therefore $\partial_u \Psi = \langle \nabla_u \Psi, u \rangle_{V_g}$. The inner-product $\langle , \rangle_{V_g}$ calculation is over a finite grid (V_g is the space of velocities where the inner-product $\langle , \rangle_{V_g}$ is taken). To calculate the Gateaux derivative of cost in Eq. (4.7), one is required to derive $\partial_u \Psi$ first as in Eq. (4.8).

$$\partial_u \Psi = \langle \frac{\partial \Psi}{\partial Y} \cdot \nabla(Y \circ \phi_1^{-1}), u \rangle_{V_g} \tag{4.8}$$

Detailed derivation of Eq. (4.8) is presented in the following. The variation of RaPTOR can be calculated using the Eq. 4.9 by perturbing the diffeomorphism ϕ_1^{-1}. The diffeomorphism is perturbed by the velocity $u = D\phi_1^{-1}$ which is obtained by taking the derivative of ϕ_1^{-1}.

$$
\begin{aligned}
\partial_u \Psi =& \frac{1}{N_p}\sum_{k=1}^{N_p} \partial_u (1-\eta(Y\circ\phi_1^{-1}|X;\Theta_k)) = \\
& \frac{1}{N_p}\sum_{k=1}^{N_p} \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(\eta(Y\circ\phi_1^{-1}|X;\Theta_k)-\eta(Y\circ(\phi_1^{-1}+\epsilon u)|X;\Theta_k)\Big) = \\
& \frac{1}{N_p}\sum_{k=1}^{N_p} \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Bigg(\frac{1}{N\sigma^2}\Big(\sum_{i=1}^{N} y_i^2 - \sum_{c=1}^{N_b} N_c\mu_c^2\Big) \\
& -\frac{N-1}{N\sum_{i=1}^{N}(z_i(\epsilon)-\mu)^2}\Big(\sum_{i=1}^{N}(z_i(\epsilon))^2 - \frac{1}{N_j}\sum_{i=1}^{N}\lambda_{ij}(z_i(\epsilon))^2 \\
& -\frac{1}{N_{j-1}}\sum_{i=1}^{N}\lambda_{ij-1}(z_i(\epsilon))^2 + \frac{1}{N_j}\sum_{i=1}^{N}\lambda_{ij}y_i^2 + \frac{1}{N_{j-1}}\sum_{i=1}^{N}\lambda_{ij-1}y_i^2\Big)\Bigg) = \\
& \Bigg(\frac{1}{N_p}\sum_{k=1}^{N_p}\frac{2}{N\sigma^2}\Big(\sum_{i=1}^{N} y_i h_i u_i - \sum_{i=1}^{N}\lambda_{i,j-1}\mu_{j-1}h_i u_i - \\
& \sum_{i=1}^{N}\lambda_{ij}\mu_j h_i u_i - \frac{1}{(N-1)\sigma^2}\Big(\sum_{i=1}^{N}(y_i-\mu)h_i\Big)\Big(\sum_{i=1}^{N} y_i^2 - \sum_{c=1}^{N_b} N_c\mu_c^2\Big)\Big)\Bigg) \\
=& \langle \frac{\partial\Psi}{\partial Y}\cdot\nabla(Y\circ\phi_1^{-1}), u\rangle_{V_g}
\end{aligned}
\tag{4.9}
$$

where $z(\epsilon) = Y\circ(\phi_1^{-1}+\epsilon u)$, $h = \nabla(Y\circ\phi_1^{-1})$, and $\nabla(Y\circ\phi_1^{-1})u = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(Y\circ(\phi_1^{-1}+\epsilon u) - Y\circ\phi_1^{-1}\Big)$ is used in the fourth equality. The term $\frac{\partial\Psi}{\partial Y}$ is a vector with N elements. Eq. 4.10 shows element ith of the vector using the notation $\{\frac{\partial\Psi}{\partial Y}\}_{i=1}^{N}$.

$$
\begin{aligned}
\left\{\frac{\partial\Psi}{\partial Y}\right\}_{i=1}^{N} = \frac{1}{N_p}\sum_{k=1}^{N_p}\frac{2}{N\sigma^2}\Bigg(& y_i - \lambda_{i,j-1}\mu_{j-1} - \lambda_{ij}\mu_j \\
& - \frac{1}{(N-1)\sigma^2}\Big(y_i-\mu\Big)\Big(\sum_{a=1}^{N} y_a^2 - \sum_{c=1}^{N_b} N_c\mu_c^2\Big)\Bigg)
\end{aligned}
\tag{4.10}
$$

Eq. (4.8) indicates that $\nabla_u\Psi = \frac{\partial\Psi}{\partial Y}\cdot\nabla(Y\circ\phi_1^{-1})$ which is known and can be calculated using the Eq. (4.5e). By similar calculations to [3] and [15], the gradient of cost can be written as Eq. (4.11).

$$\nabla_{\tilde{v}_1} E = \nu\bigg(- K\bigg(\frac{\partial \Psi}{\partial Y} \cdot \nabla(Y \circ \phi_1^{-1})\bigg)\bigg) \tag{4.11}$$

the gradient in Eq. (4.11) for the velocity in $t = 1$ can be used to find the gradient $\nabla_{\tilde{v}_0} E$ in $t = 0$ with the reduced adjoint Jacobi field in bandlimited velocity fields elaborated in [15]. This process is called backward integration. To minimize the cost in Eq. 4.7, forward integration of Eq. 4.2 is used to find the velocity in $t = 1$. Then $\nabla_{\tilde{v}_0} E$ is used in GD with momentum optimization to update the velocity. Finally, Eq. 4.1 is used to calculate diffeomorphisms. Since similar process was used in [15] to calculate diffeomorphisms, the diffeomorphic registration is guaranteed. The employment of multi-resolution pyramid, gradient descent with momentum, and implementation details of DiffeoRaptor can be found in the following sections.

4.1.5 Mutli-resolution pyramid and gradient descent with momentum

For DiffeoRaptor, the GD with momentum was used to break the zig-zag pattern of GD and reduce the chance of trapping in a local minimum. Thus, the convergence for DiffeoRaptor could be possibly better than FLASH and RaPTOR. The update equation in iteration i for GD with momentum is outlined in Eq. 4.12. In Eq. 4.12, p_i is the momentum parameter, and $(\tilde{v}_0)_i$ is the velocity field in iteration i. Setting $p_i = \frac{0.1}{1-(0.9)^{i+1}}$ produces best results for us.

$$(\tilde{v}_0)_{i+1} = (\tilde{v}_0)_i - \epsilon \nabla_{\tilde{v}_0} E + p_i((\tilde{v}_0)_i - (\tilde{v}_0)_{i-1}) \tag{4.12}$$

Multi-resolution image pyramid is employed often in image registration techniques to improve convergence of the algorithm. However, FLASH does not implement this technique. In DiffeoRaptor, it was implemented for up to three levels of image pyramids to produce larger deformations. Fig. 4.1 shows image energy term (Ψ), total energy (E), and MSE between the template and source image over the iterations for three image pyramid levels. In this test case, there were maximum ten iterations needed for the convergence. Since RaPTOR and followingly DiffeoRaptor use random patches to calculate Ψ, the quan-

tities over the iterations are not strictly decreasing. Moreover, for the Level 0, MSE is increasing after the 5th iteration which does not necessarily mean that the image alignment has been deteriorated.

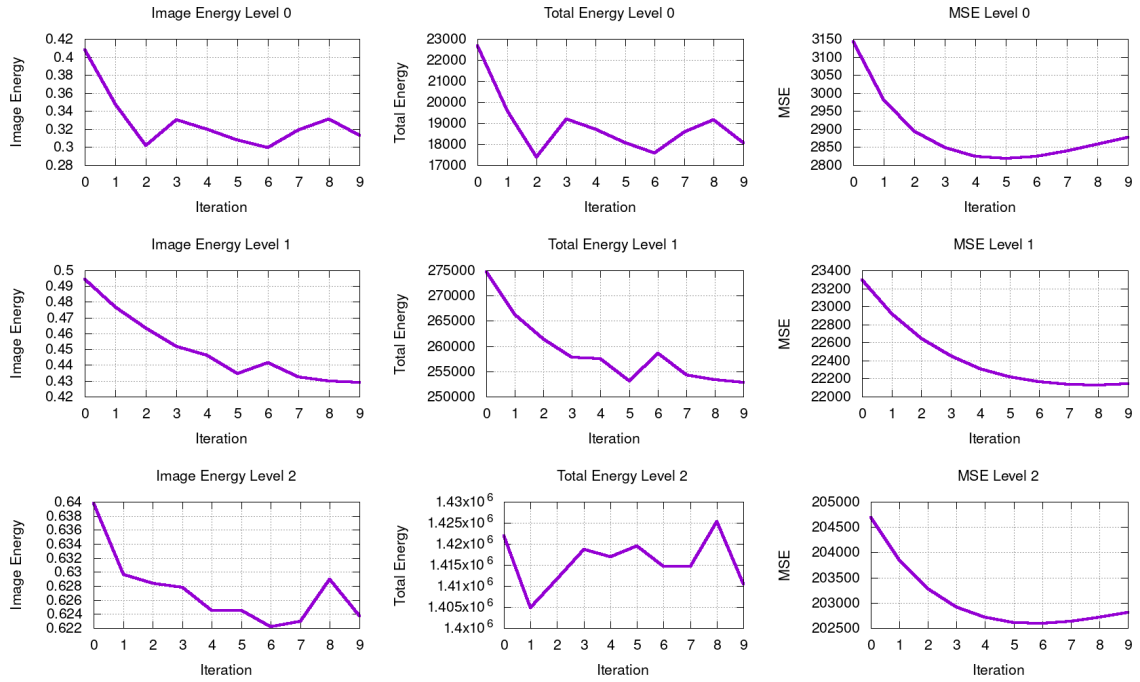

Figure 4.1: Image energy term (Ψ), tatal energy (E), and MSE between the template and source image over the iterations for three image pyramid levels. In this test case which is a T1 image to a template registration, maximum five iterations were needed for the convergence,

4.1.6 Implementation Details

Parameter settings for DiffeoRaptor are similar to what is elaborated in RaPTOR [40] and FLASH [3]. The patch size of $7 \times 7 \times 7$ and number of patches of 3000, 2000, and 200 for the finest to the coarse level respectively, produce the best result according to the following parameter search experiment. In this experiment, six subjects of IXI dataset are selected which are not used (hold-out validation set) in the image registrations experiments of Section 3.2. $\binom{6}{2} = 15$ image registrations are performed by selecting two distinct subjects each time. Table 4.1 shows five different parameter setups and mean Dice score of overlapping

regions. Setup 1 has better mean Dice score than the other setups. Setup 1 and 5 have similar mean Dice scores, however, the image registrations using Setup 5 takes more execution time due to the selection of more patches.

Table 4.1: Mean Dice score evaluation of 15 T1-T1 inter-subject registrations of hold-out validation set of IXI data for DiffeoRaptor in overlapping regions of brain tissues and sixteen subcortical structures for 5 parameter setups. Note that the number of patches are for the finest to the coarse level respectively.

Setup	Patch size	Number of patches	Dice - brain tissue	Dice - subcortical structure
1	$7 \times 7 \times 7$	$3000, 2000, 200$	0.70 ± 0.01	0.83 ± 0.02
2	$5 \times 5 \times 5$	$3000, 2000, 200$	0.70 ± 0.01	0.82 ± 0.02
3	$9 \times 9 \times 9$	$3000, 2000, 200$	0.70 ± 0.01	0.82 ± 0.02
4	$7 \times 7 \times 7$	$1500, 1000, 100$	0.69 ± 0.01	0.82 ± 0.02
5	$7 \times 7 \times 7$	$5000, 3000, 400$	0.70 ± 0.01	0.83 ± 0.02

Given the discussion on the size of bandlimited space in [15], the bandlimited space size is set to $16 \times 16 \times 16$. However, the maximum number of iterations are smaller for each level of image pyramids compared to FLASH's maximum number of iterations in the full resolution. Note that the registration in the full resolution has the greatest impact on the overall computational time. Typically in our experiments, the number of iterations are fewer than 30 for each level of image pyramids.

Similar to Rivaz *et al*. [40], by using a test dataset, we noticed that setting Y as the moving image and X as the fixed image produced the best results for us. Note that in all the registrations, X is the reference T1w image and Y can be T1w, T2w, PDw, or CT.

4.2 Experiments and results

DiffeoRaptor was validated on three public datasets: IXI[1], OASIS3 [49], and The Cancer Imaging Archive (TCIA) MR-CT abdominal data [165]. It is compared against Mattes MI+SyN, which is available in Advanced Normalization Tools (ANTs) [191] and NiftyReg [188] (using the normalized mutual information (NMI) as the similarity metric), as well

[1] http://brain-development.org/ixi-dataset

Table 4.2: Abbreviation of subcortical structures which were automatically labelled in the segmentation of brain volumes using volBrain [193].

Subcortical Structure	Abbreviation
Left/Right Ventricle	LV/RV
Left/Right Caudate	LC/RC
Left/Right Putamen	LP/RP
Left/Right Thalamus	LT/RP
Left/Right Globus Pallidus	LGP/RGP
Left/Right Hippocampus	LH/RH
Left/Right Amygdala	LA/RA
Left/Right Accumbens	LAC/RAC

as in several tasks with FLASH. Dice scores of overlapping regions are used as evaluation metrics. The default parameters for NiftyReg with the GD optimization produced the best results for us. Mattes MI+SyN is a diffeomorphic algorithm which uses Mattes MI as the similarity metric and models the deformation fields with SyN, and is suitable for inter-modal/contrast image registration. The parameters for Mattes MI+SyN were tuned such that it produced the optimal results. The number of bins for MI was set to 32 and the gradient step, the update field variance, and total field variance were set to 0.5, 3, and 0.5 for SyN respectively.

4.2.1 Pre-processing of brain MRI

Brain MR images of the IXI and OASIS3 datasets, were first skull-stripped using nonlocal intracranial cavity extraction [192]. For each case, the extracted brain was carefully inspected. Then, two types of segmentations were generated for each volume using the volBrain algorithm [193] so that Dice scores can be used to evaluate registration accuracy. Here, in the first one, brain tissues are classified into Cerebrospinal Fluid (CSF), Gray Matter (GM), and White Matter (WM). The second type of segmentation consists of 16 subcortical structures which are abbreviated in Table 4.2. Lastly, the volumes were affinely registered using ANTs with Mattes MI as the metric (see Fig. 4.2).

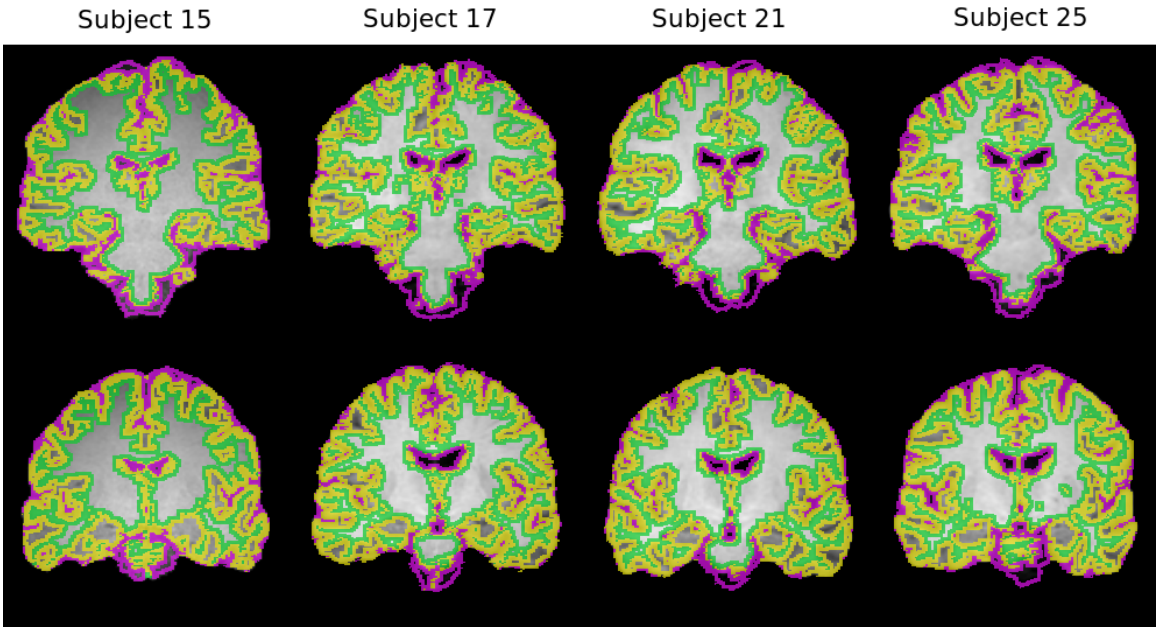

Figure 4.2: Coronal view of two slices (rows) of four different IXI dataset subjects (columns). The images are overlaid by the segmentation of CSF, GM, and WM. The large variability of structures across subjects require a deformable registration.

4.2.2 IXI Dataset: inter-subject registration

Twenty young adult subjects ($age < 30yo$) of the IXI dataset were selected randomly. Given the fact that the IXI datatset offers T1w, T2w, and PDw for each subject, three different tasks were designed, including T1-T1, T1-T2, and T1-PD registrations. T1w MRI scans of three subjects (Subject 15, 17, and 21) were randomly selected as the reference volume and the rest are set as the source volumes for inter-subject registration (in total

Table 4.3: Dice score (mean±std) evaluation of T1-T1, T1-T2, and T1-PD registrations of IXI dataset for DiffeoRaptor, Mattes MI+SyN, FLASH, and NiftyReg in overlapping regions of brain tissues and sixteen subcortical structures.

Task and Evaluation Region	Affine only	DiffeoRaptor	Mattes MI+SyN	FLASH	NiftyReg
T1-T1 Brain Tissues	0.62 ± 0.03	0.72 ± 0.04	0.72 ± 0.04	0.64 ± 0.04	0.67 ± 0.03
T1-T1 Subcortical Structures	0.67 ± 0.06	0.78 ± 0.03	0.78 ± 0.04	0.67 ± 0.06	0.74 ± 0.05
T1-T2 Brain Tissues	0.62 ± 0.03	0.67 ± 0.04	0.67 ± 0.04	0.64 ± 0.03	0.64 ± 0.04
T1-T2 Subcortical Structures	0.67 ± 0.06	0.71 ± 0.06	0.71 ± 0.06	0.66 ± 0.05	0.65 ± 0.05
T1-PD Brain Tissues	0.62 ± 0.03	0.67 ± 0.04	0.67 ± 0.04	0.64 ± 0.03	0.64 ± 0.04
T1-PD Subcortical Structures	0.67 ± 0.06	0.71 ± 0.05	0.69 ± 0.06	0.67 ± 0.06	0.68 ± 0.06

Table 4.4: Dice score (mean±std) evaluation of ICBM152-T1, ICBM152-T2, and ICBM152-PD registrations of IXI dataset for DiffeoRaptor, Mattes MI+SyN, FLASH, and NiftyReg in overlapping regions of brain tissues and sixteen subcortical structures.

Task and Evaluation Regions	Affine only	DiffeoRaptor	Mattes MI+SyN	FLASH	NiftyReg
ICBM152-T1 Brain Tissues	0.62 ± 0.04	0.68 ± 0.04	0.70 ± 0.04	0.63 ± 0.04	0.71 ± 0.05
ICBM152-T1 Subcortical Structures	0.70 ± 0.06	0.80 ± 0.02	0.78 ± 0.04	0.71 ± 0.04	0.74 ± 0.07
ICBM152-T2 Brain Tissues	0.62 ± 0.04	0.65 ± 0.05	0.67 ± 0.06	0.62 ± 0.04	0.65 ± 0.05
ICBM152-T2 Subcortical Structures	0.70 ± 0.06	0.76 ± 0.04	0.76 ± 0.08	$0.67 \pm 0.06)$	0.73 ± 0.08
ICBM152-PD Brain Tissues	0.62 ± 0.04	0.66 ± 0.04	0.66 ± 0.05	0.63 ± 0.04	0.64 ± 0.05
ICBM152-PD Subcortical Structures	0.70 ± 0.06	0.75 ± 0.05	0.73 ± 0.07	0.70 ± 0.05	0.73 ± 0.07

$3 \times 19 = 57$ registrations). The results of Dice score evaluation are summarized in Table 4.3, which shows that DiffeoRaptor, Mattes MI+SyN, and NiftyReg could successfully align volumes in each task whereas FLASH underperformed in terms of Dice scores in intra-contrast tasks and failed in inter-contrast tasks as expected. It can also be seen that DiffeoRaptor in general did better than Mattes MI+SyN.

4.2.3 IXI dataset: subject-to-template registration

Given the IXI subjects in Section 4.2.2, the volumes are set as the source volumes and they were registered to the T1w ICBM152 template [194]. Here, the template is set as the reference volume and similar tasks were performed as Section 4.2.2 for subject-to-template registration. The results are summarized in Table 4.4, which shows that DiffeoRaptor, Mattes MI+SyN, and NiftyReg could successfully align volumes in each task while DiffeoRaptor in general did better than Mattes MI+SyN and NiftyReg in alignment of subcortical structures.

Figure 4.3 demonstrates two coronal views of registration results. The subcortical structures are shown in the figure as colored outlines. DiffeoRaptor show better alignment of slices and anatomical structures compared to other methods. The cerebrum shape with DiffeoRaptor registration looks closer to the ICBM152 template than other methods.

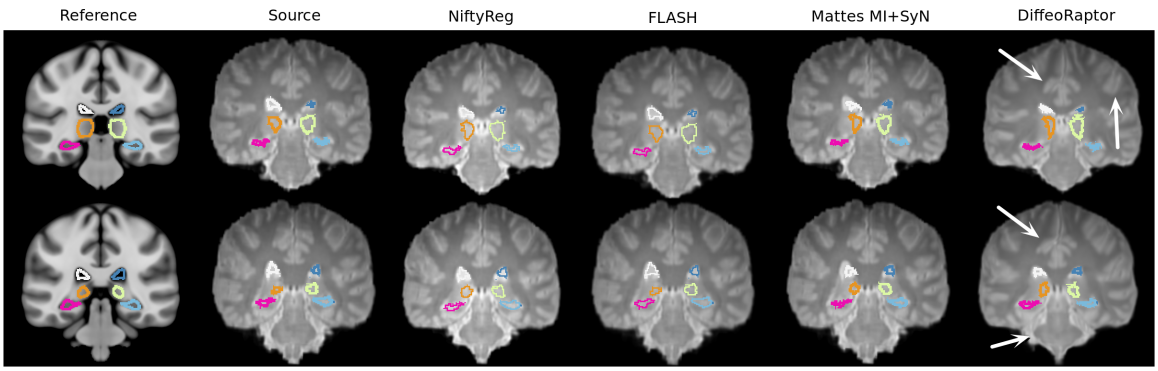

Figure 4.3: From the left to right: coronal slices of the ICBM152 (reference volume), the PDw source volume of the IXI dataset, result of NiftyReg, FLASH, Mattes MI+SyN, and DiffeoRaptor respectively. Rows show different coronal views. Subcortical structural segmentations are shown in colored contours. Arrows are pointing to the regions where the image alignments are more visible.

Table 4.5: Dice score evaluation (mean±std) of T1-T2 inter-subject registration of IXI data with OASIS3 data for DiffeoRaptor, Mattes MI+SyN, and NiftyReg in overlapping regions of brain tissues and sixteen subcortical structures.

Task and Evaluation Regions	Affine only	DiffeoRaptor	Mattes MI+SyN	NiftyReg
T1-T2 Brain Tissues Mean	0.51 ± 0.13	0.61 ± 0.07	0.54 ± 0.16	0.56 ± 0.16
T1-T2 Subcortical Structures	0.56 ± 0.17	0.71 ± 0.11	0.63 ± 0.21	0.59 ± 0.18

4.2.4 OASIS3 dataset: intra- and inter-subject registration

The OASIS3 dataset consists of subjects intended for investigating AD [49]. Twenty AD patients from this dataset were randomly selected with matching T1w and T2w MRIs. In the first sub-task, intra-contrast intra-subject registration was performed for brain scans obtained at different stages of AD progression, where the T1w volume at the baseline was set as the reference and the T1w image from the latest session ($>$ 6 months apart) with visible atrophy was registered to the reference. This sub-task represents the need in neuroimage analysis for tracking disease-related anatomical changes.

The results of intra-subject registration of OASIS3 data were summarized in Table 4.6. In Table 4.6, it can be seen that for the first task (T1-T1 intra-subject registration), Diffeo-Raptor, Mattes MI+SyN, and NiftyReg improved the Dice score by 0.01. The reason is that the affine registration aligned the images such that there was less room for improvement.

Table 4.6: Mean Dice score evaluation of T1-T1 intra-subject registration of OASIS3 data for DiffeoRaptor, Mattes MI+SyN, and NiftyReg in overlapping regions of brain tissues and sixteen subcortical structures.

Task and Evaluation Regions	Affine only	DiffeoRaptor	Mattes MI+SyN	NiftyReg
T1-T1 Brain Tissues	0.83 ± 0.02	0.84 ± 0.01	0.83 ± 0.01	0.84 ± 0.02
T1-T1 Subcortical Structures	0.87 ± 0.02	0.88 ± 0.01	0.88 ± 0.01	0.88 ± 0.01

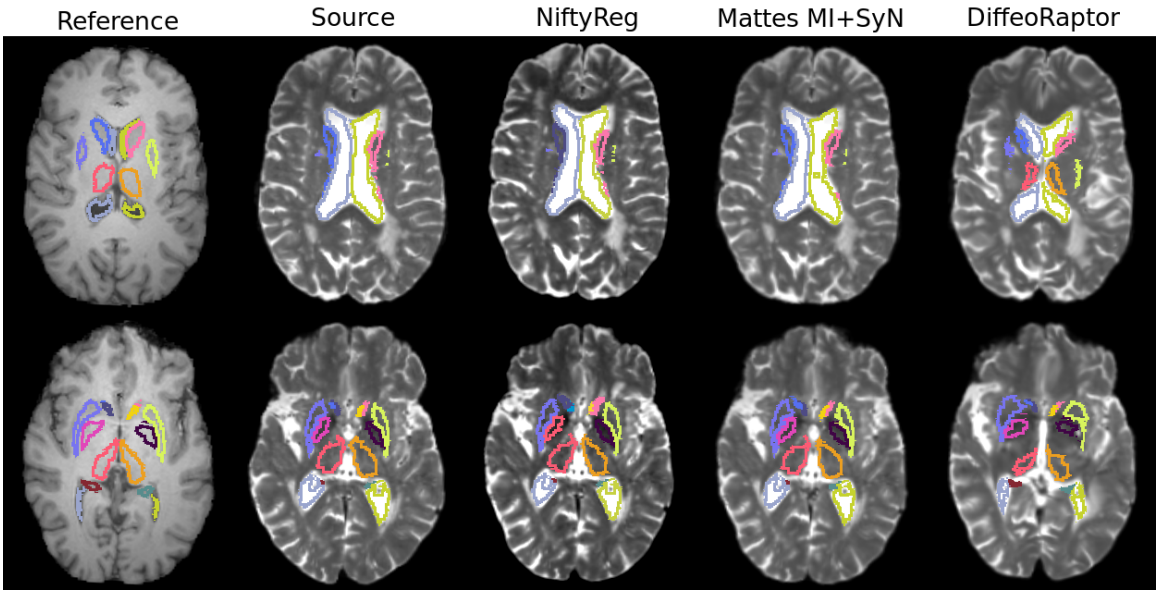

Figure 4.4: From the left to right: axial slices of the T1w reference volume from the IXI dataset, the T2w MRI source volume of the OASIS3 dataset, result of NiftyReg, Mattes MI+SyN, and DiffeoRaptor respectively. Rows show different axial views. Subcortical segmentations are shown in colored contours.

In the second sub-task, T1w MRIs of four young healthy adults of the IXI datatset in Section 4.2.2 were used as the references and the T2w MRI scans of the latest session for each subject from the OASIS3 dataset were set as the source volumes, resulting in $4 \times 20 = 80$ registrations. This way, we defined a more challenging, inter-contrast, inter-subject, and inter-dataset task to better compare DiffeoRaptor with Mattes MI+SyN and NiftyReg.The results of T1-T2 registrations are summarized in Table 4, where DiffeoRaptor outperformed Mattes MI+SyN and NiftyReg. Note that FLASH was not included in these experiments because it continuously failed to perform inter-contrast registration. In Fig. 4.4, it can be seen that DiffeoRaptor has improved the alignment of subcortical structures and ventricles

Table 4.7: Dice score (mean±std) evaluation of MR-CT intra-subject registration for TCIA abdominal data using DiffeoRaptor, RaPTOR [40], Mattes MI+SyN, and NiftyReg.

Evaluation Regions	DiffeoRaptor	RaPTOR	Mattes MI+SyN	NiftyReg
Liver	0.81 ± 0.07	0.81 ± 0.09	0.80 ± 0.10	0.79 ± 0.11
Spleen	0.71 ± 0.13	0.71 ± 0.16	0.69 ± 0.10	0.71 ± 0.13
Left Kidney	0.70 ± 0.15	0.71 ± 0.15	0.68 ± 0.17	0.65 ± 0.22
Right Kidney	0.70 ± 0.19	0.69 ± 0.19	0.67 ± 0.17	0.65 ± 0.22
Average	0.78 ± 0.10	0.77 ± 0.10	0.77 ± 0.11	0.76 ± 0.13

better than Mattes MI+SyN and NiftyReg.

4.2.5 TCIA abdominal MR-CT intra-subject registration

The TCIA dataset contains eight subjects. Each subject has a T1w MRI scan and CT scan (with deformation) of the abdomens. The manual segmentations of the liver, spleen, left kidney, and right kidney are provided by the Learn2Reg organizers (https://learn2reg.grand-challenge.org). By setting the MRI scan for each subject as the reference volume, CT scans were aligned to perform intra-subject registrations. The deformable registration for MR-CT of these subjects are required because the images were taken in different time points, with different modalities, and misalignments due to patient movement, respiration, and etc. The results are summarized in Table 4.7.

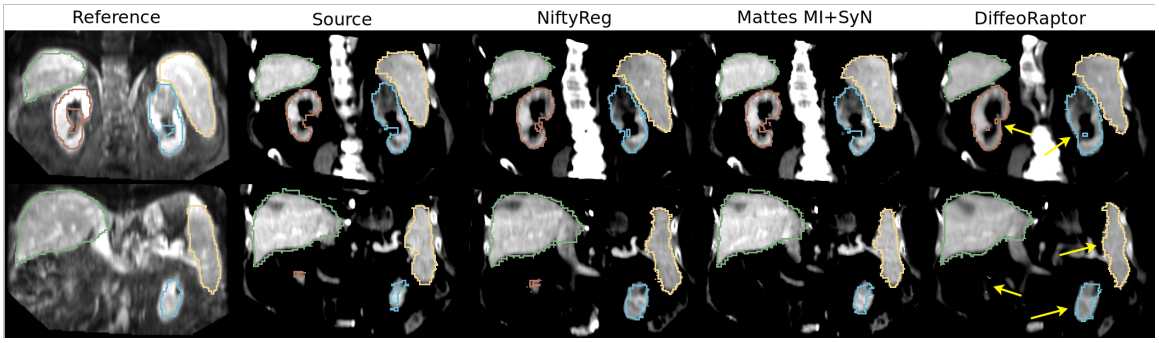

Figure 4.5: From left to right: coronal slices of Subject 7's MRI (reference volume), the corresponding CT source volume, results of NiftyReg, Mattes MI+SyN, DiffeoRaptor, and NiftyReg respectively. Rows show different slices of volumes. Segmentations of key organs are shown with colored contours. Arrows are pointing to the regions where the image alignments are more visible.

Table 4.8: Dice scores (mean±std) of cumulative results for DiffeoRaptor, Mattes MI+SyN, and NiftyReg in overlapping subcortical structures. The p-values from ANOVA are shown for each anatomical structure.

Evaluation Regions	DiffeoRaptor	Mattes MI+SyN	NiftyReg	p-value
LV	0.65 ± 0.14	0.60 ± 0.18	0.58 ± 0.18	1.21×10^{-5}
RV	0.64 ± 0.13	0.59 ± 0.17	0.57 ± 0.16	1.60×10^{-6}
LC	0.73 ± 0.11	0.70 ± 0.17	0.66 ± 0.16	4.30×10^{-6}
RC	0.72 ± 0.11	0.68 ± 0.19	0.64 ± 0.17	1.15×10^{-5}
LP	0.79 ± 0.08	0.75 ± 0.14	0.73 ± 0.14	4.90×10^{-7}
RP	0.78 ± 0.08	0.73 ± 0.15	0.70 ± 0.16	2.33×10^{-11}
LT	0.80 ± 0.09	0.78 ± 0.16	0.73 ± 0.15	1.58×10^{-7}
RT	0.78 ± 0.09	0.77 ± 0.15	0.71 ± 0.15	6.56×10^{-8}
LGP	0.70 ± 0.10	0.65 ± 0.14	0.60 ± 0.17	1.04×10^{-13}
RGP	0.68 ± 0.10	0.64 ± 0.14	0.57 ± 0.17	2.44×10^{-15}
LH	0.69 ± 0.09	0.65 ± 0.14	0.64 ± 0.14	7.31×10^{-5}
RH	0.73 ± 0.09	0.69 ± 0.14	0.67 ± 0.14	8.56×10^{-8}
LA	0.60 ± 0.15	0.55 ± 0.18	0.53 ± 0.18	3.37×10^{-5}
RA	0.60 ± 0.14	0.57 ± 0.17	0.53 ± 0.17	7.06×10^{-6}
LAC	0.50 ± 0.17	0.43 ± 0.21	0.37 ± 0.22	1.22×10^{-11}
RAC	$0.46 \pm (0.18$	0.45 ± 0.20	0.36 ± 0.21	7.70×10^{-9}
Average	0.72 ± 0.08	0.68 ± 0.14	0.65 ± 0.13	1.14×10^{-8}

Given the fact that the initial affine registration achieved mean Dice score of 0.72 ± 0.10, Table 4.7 shows DiffeoRaptor, RaPTOR [40], Mattes MI+SyN, and NiftyReg could successfully improve the image alignment. Besides, DiffeoRaptor outperformed Mattes MI+SyN and NiftyReg in alignment of all the targeted regions. Note that two subjects didn't have the segmentation of the right kidney and thus they were excluded from the Mean Dice calculation of Table 4.7. In Fig. 4.5, it can be seen that compared to the affine registration, Mattes MI+SyN and DiffeoRaptor show improvement in alignment of segmented organs. However, DiffeoRaptor shows better alignment of organs compared to Mattes MI+SyN and NiftyReg.

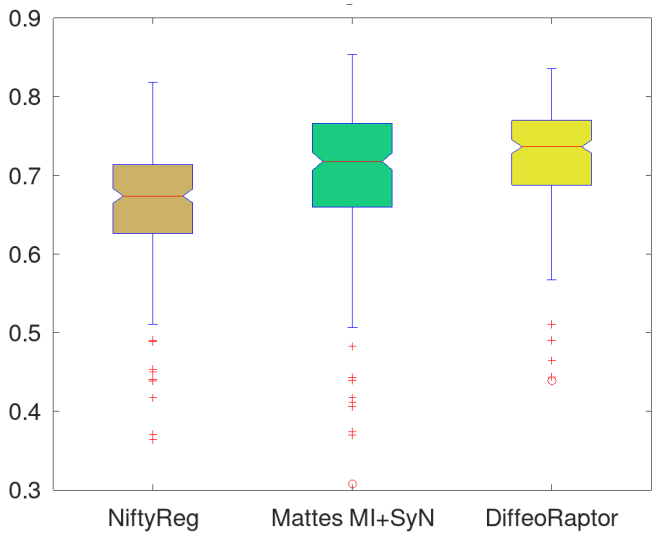

Figure 4.6: The box plots of average Dice score for the total of 291 brain image registrations. DiffeoRaptor has a higher mean and lower std with fewer outliers.

Table 4.9: Post-hoc multiple comparison (Tukey-Kramer) tests of DiffeoRaptor against Mattes MI+SyN and NiftyReg for the average Dice in overlapping subcortical structures. DiffeoRaptor results are better than Mattes MI+SyN and NiftyReg ($p < 0.05$).

Methods	p-value
DiffeoRaptor vs Mattes MI+SyN	1.92×10^{-2}
DiffeoRaptor vs NiftyReg	3.62×10^{-6}

4.2.6 Cumulative results

Given the inter-contrast registration results (total 291) in Section 4.2.2, 4.2.3, and 4.2.4 for brain structures, the mean Dice scores and the associated p-values from comparing the three methods using the one-way analysis of variance (ANOVA) were listed for the sixteen subcortical structures in Table 4.8. Furthermore, post-hoc multiple comparison (Tukey-Kramer) tests were performed to reveal the performance of the methods (Table 4.9). With the statistical tests, we confirm that DiffeoRaptor outperforms the rest in terms of Dice scores for aligning each subcortical region, as well as the mean Dice score ($p < 0.05$). It is worth mentioning that the average mean Dice is 0.63 ± 0.12 for the affine registration. To better visualize the results for the last row of Table 4.8, the box plots of average Dice scores over all evaluation regions are demonstrated in Fig. 4.6.

4.2.7 Deformation smoothness analysis

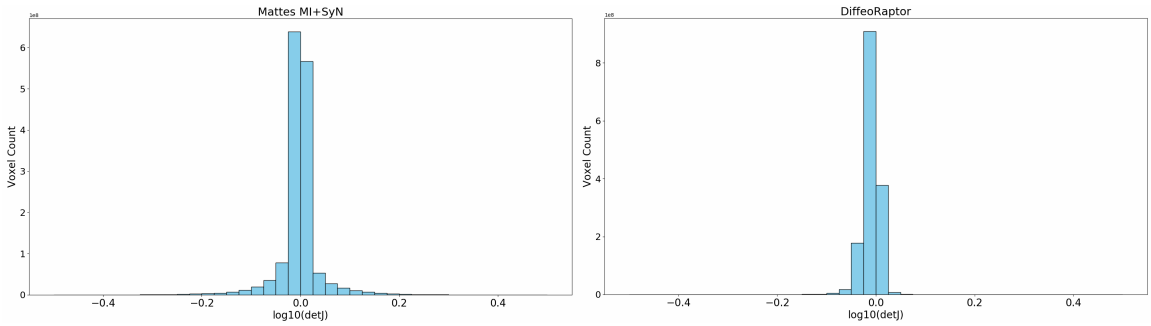

Figure 4.7: The logarithm of determinant of Jacobian $log_{10}(det(J))$ was calculated for each voxel of the deformation field. Then they were accumulated in bins for Mattes MI+SyN and DiffeoRaptor.

With the cumulative results in Section 4.2.6, in each registration, the determinant of Jacobian J was calculated for each voxel of the deformation field. Figure 4.7 shows $log_{10}(det(J_\phi))$ for each voxel that they were accumulated in bins for DiffeoRaptor and Mattes MI+SyN. For example, the bin centred at the origin means no deformation, bins with the negative centres show contraction, and bins with positive centre show expansion. The further the bin from the centre, the more deformation the bin represents. From the experiments, we observed that the number of non-zero samples is similar across DiffeoRaptor, Mattes MI+SyN, and NiftyReg. However, DiffeoRaptor has fewer samples far from the central bin (Fig. 4.7).

Table 4.10: Comparison of deformation smoothness of DiffeoRaptor, Mattes MI+SyN, and NiftyReg by the determinant of Jacobians of deformation fields (J_ϕ).

	Method	$\log_{10}(\det(J_\phi)) \neq 0$	$\sum \lvert\log_{10}(\det J_\phi)\rvert$	$\det(J_\phi) \leq 0$
(i)	Mattes MI+SyN	1.49×10^8	2.61×10^7	9038
(ii)	NiftyReg	1.49×10^8	3.62×10^7	47048
(iii)	DiffeoRaptor	1.49×10^8	1.95×10^7	0
	Difference of (i) and (iii)	9038	6.58×10^6	9038
	Difference of (ii) and (iii)	47048	1.67×10^7	47048

To further validate the smoothness of deformation fields, the quantity $\sum |\log_{10}(det(J))|$ was calculated where the summation is over all the samples available and the results are

summarized in Table. 4.10. It can be seen that the summation values in Mattes MI+SyN and NiftyReg are higher than DiffeoRaptor. This suggests that DiffeoRaptor generates smoother deformations while having generally better performance. The other way to compare the deformation fields is counting the number of voxels in which $\det(J_\phi) \leq 0$. If $\det(J_\phi) \leq 0$, it means the mapping of that voxel was not diffeomorphic. Last column of Table. 4.10 shows that Mattes MI+SyN and NiftyReg had many voxels with this property, whereas DiffeoRaptor did not have any voxels in which $\det(J_\phi) \leq 0$.

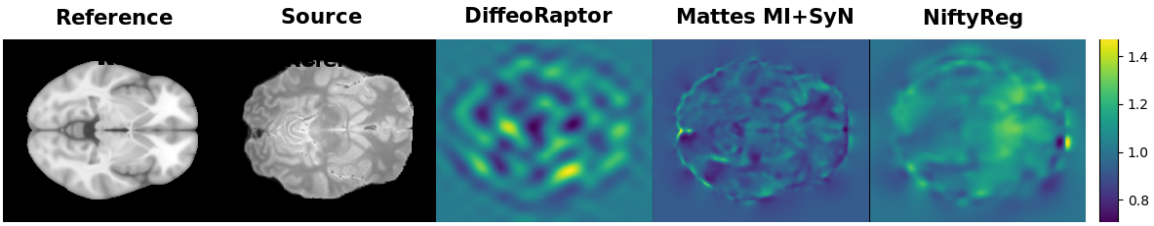

Figure 4.8: The visualization of Jacobian determinants for DiffeoRaptor, Mattes MI+SyN, and NiftyReg is presented. DiffeoRaptor shows smoother deformation compared to other methods.

Figure 4.8 shows the visualization of Jacobian determinants for DiffeoRaptor, Mattes MI+SyN, and NiftyReg. The registration is performed between ICBM152 T1w and Subject 15 of IXI dataset (PDw). For this registration, the number of negative and zero Jacobian determinants ($\det(J_\phi) \leq 0$) are zero for all the cases. However, the value of $\sum |\log_{10}(detJ_\phi)|$ for DiffeoRaptor, Mattes MI+SyN, and NiftyReg are 12954.1, 204088.1, and 360751.5 respectively. This indicates that the deformation by DiffeoRaptor is smoother than the other methods. DiffeoRaptor ensures diffeomorphic transformations, while SyN and NiftyReg do not guarantee them. So, DiffeoRaptor has the advantage of generating diffeomorphisms according to the experiments, which conforms with the theory.

Table. 4.11 presents the comparison of DiffeoRaptor, RaPTOR [40], Mattes MI+SyN, and NiftyReg in terms of deformation smoothness for TCIA abdominal MR-CT intra-subject registration. The quantity $\sum |\log_{10}(det(J_\phi))|$ is calculated where the summation is over all the samples available. For the ablation study, DiffeoRaptor suggests smoother deformation compared to RaPTOR [40]. NiftyReg generated 76118 samples with $\det(J_\phi) \leq 0$ which means the deformations are not diffeomorphic. While DiffeoRaptor outperformed

Table 4.11: Comparison of deformation smoothness of DiffeoRaptor, RaPTOR [40], Mattes MI+SyN, and NiftyReg by the determinant of Jacobians of deformation fields (J_ϕ) for TCIA abdominal MR-CT intra-subject registration.

	Method	$\log_{10}(\det(J_\phi)) \neq 0$	$\sum \lvert\log_{10}(\det J_\phi)\rvert$	$\det(J_\phi) \leq 0$
(i)	Mattes MI+SyN	4.71×10^7	2.39×10^6	0
(ii)	NiftyReg	4.71×10^7	1.55×10^6	76118
(iii)	RaPTOR	5.11×10^7	4.61×10^7	0
(iv)	DiffeoRaptor	5.11×10^7	3.75×10^7	0
	Difference of (i) and (iv)	3.95×10^6	3.51×10^7	0
	Difference of (ii) and (iv)	4.01×10^6	3.59×10^7	76118
	Difference of (iii) and (iv)	0	8.62×10^6	0

other methods in terms of average Dice, Mattes MI+SyN generated smoother deformations compared to other methods.

4.3 Limitations

In the work of Rohlfing *et al*. [195], it was shown that although similarity metric and automatically generated tissue overlaps are commonly used for validation of image registration accuracy, they can be unreliable. They investigated two possible solutions for this problem: 1) utilization of localized labels and 2) using manually generated segmentations. For our experiments, the registration accuracy is validated on 16 subcortical regions. Our validation using subcortical regions showed the agreement across the results on tissue overlaps, qualitative validation, and our similarity metric (RaPTOR). Similar approaches were used for validation of image registration in most of the current registration techniques (e.g. VoxelMorph [186]). However, we admit that these validation approaches (tissue overlap and similarity metrics) have their limitations. When resources are available, validation using an established manually labelled dataset and multi-modal ground truths (e.g. addition of matching anatomical landmarks) will form better assessments.

The segmentations of 16 subcortical structures are generated by the automated algorithm (volBrain [193]). The algorithm uses non-local means patch-based segmentation and only affine registration to align the target MRI and the atlas library. The algo-

rithm was compared against the widely used brain tissue/structure segmentation software, FreeSurfer [196] and FIRST [197] in in terms of Dice coefficients on a manually labelled dataset [193]. The average Dice scores of 0.91 ± 0.05, 0.73 ± 0.11, and 0.77 ± 0.10 were reported for volBrain, FreeSurfer, and FIRST, respectively [193], showing excellent performance of volBrain. Worth to mention that most of the current learning-based registration works (e.g. VoxelMorph [186]) use automatic segmentation, such as FreeSurfer for validation purposes due to limited resources for manually labeled datasets.

In Section 3.3, the T1w ICBM152 template was used in the automated algorithm volBrain to generate tissue maps and segmentation of subcortical structures. Due to partial volume effects from averaging several subjects, the generated segmentations may not be as accurate as expected. This could possibly be mitigated by manual segmentations of individual subject's brain MRI that contributed towards the T1w ICBM152, and then thresholding the averaged tissue maps. Unfortunately, this is not feasible due to limitation of resources and public availability of the data.

The Task 1 training dataset of Learn2Reg challenge is used in our MR-CT registration in Section 3.5. However, the leader-board only shows the result for the test dataset, which can be downloaded from the challenge website[2] but it does not include the associated image segmentations for validation purposes. Therefore, we decided to compare our results with those official participants of the Learn2reg challenge, whose verified results are published in the associated challenge summary paper [187]. In the report, the top three methods for Task 1 achieved the average Dice score of 0.78, 0.76, and 0.75 respectively. Given our results in Table 5 of the manuscript, the average Dice score of 0.78 was achieved. Therefore, DiffeoRaptor potentially has a similar performance to the top methods that participated in the challenge.

[2]https://learn2reg.grand-challenge.org/Learn2Reg2021/

4.4 Discussions

When RaPTOR is employed as the similarity metric, it may require additional parameter-tuning. This motivates more advanced optimization technique rather than the classical GD to minimize the cost function. This was shown and explored in [41] and [42]. For DiffeoRaptor, the parameter settings were mostly the default values from RaPTOR and FLASH as elaborated previsouly. However, for the cases where affine registration fails to perform good initial alignments, we should be careful in choosing the step size for the gradient update and the maximum number of iterations. The average computational times were calculated for DiffeoRaptor and FLASH on a single core of a 6 core Linux Mint system for 10 T1-T1 brain MRI registrations with the image size of $176 \times 256 \times 256$ voxels. The mean computational time per registration of DiffeoRaptor ($384.50 \pm 0.01s$) is comparable to that of FLASH ($416.14 \pm 0.01s$).

4.5 Summary

We present DiffeoRaptor, a diffeomorphic inter-modal/contrast image registration algorithm based on RaPTOR and geodesic shooting in bandlimited space. The algorithm is validated on several different applications. Compared with FLASH, Mattes MI+SyN, and NiftyReg, it achieves comparable or better results. In addition, DiffeoRaptor offers smoother deformation fields than Mattes MI+SyN and NiftyReg.

Chapter 5

DLCR: Deep Learning-based deformable image registration using Correlation Ratio

---

MRI has been used extensively in clinical applications such as non-invasive diagnosis of soft tissue diseases, surgical procedures, and quantification of neurodegenerative disorders such as PD and AD. RF pulse sequences are used to acquire MRI volumetric images. Different pulse sequences can produce a variety of contrasts. T1w contrast is used more than other MRI contrasts in specific applications such as voxel-based morphometry and characterization of brain anatomy because brain anatomical structures and tissues are interpretable in T1w images [199, 200]. Inter/intra-subject registration of T1w scans is a key component in various clinical studies. A standard practice to perform image registration starts with resampling the images to a common coordinate space. Then, an affine image registration algorithm is employed for rough alignments. Next, fine alignment of images can be achieved by performing a deformable image registration —an essential component of many neuroimaging pipelines [201–203]. Therefore, designing and implementing a time-efficient deformable image registration algorithm can be substantial.

Time-efficient image registration methods can reduce the waiting time for surgeons during operations and facilitate the exploitation of large datasets for researchers. Several

traditional methods have been proposed to achieve these objectives. Avants *et al.* [5] proposed the Symmetric Normalization (SyN) method. They validated their method on MRI subjects with frontotemporal dementia. Modat *et al.* [188] implemented a fast image registration to maximize normalized mutual information using GPUs. The method is called NiftyReg and could successfully perform fast registration of T1w images for AD patients. Zhang *et al.* [15] and Wu *et al.* [185] proposed fast diffeomorphic methods based on the well-known LDDMM [3] to register T1w images. Traditional image alignment methods can still offer a good speed for pairwise registrations when parallel computing and GPUs are employed. However, analyzing large datasets using these methods can take time and effort.

Recently, DL-based image registration methods could achieve promising results [8–10]. Yang *et al.* [11] proposed QuickSilver, based on LDDMM, to perform atlas-to-subject and inter-subject registration. Balakrishnan *et al.* [12] proposed VoxelMorph to perform inter-subject registration of brain MRIs. They trained, validated, and tested their network on several datasets and compared their method with SyN and NiftyReg. Mok *et al.* [13] proposed a symmetric diffeomorphic algorithm, and the method was evaluated on atlas-based registration T1w images of the OASIS dataset [50]. Wang *et al.* [14] proposed a fast method based on FLASH [15] to register 3D brain MRIs. They named their method DeepFLASH, and they demonstrated that the method could outperform FLASH, Quicksilver, and VoxelMorph in terms of Dice scores of overlapping brain structures and time efficiency. In a recent study, Zhang *et al.* [16] proposed a novel dual transform network to register brain MRIs. However, training a model to align images that require the estimation of large deformations is a challenging task and is still an open research area.

The aforementioned DL-based image registration methods were designed to estimate a deformation field for inter/intra-subject alignment of brain images. Nevertheless, these methods may fail to estimate a correct displacement field when larger deformations are required. One such example is the challenging task of image registration involving AD patients' data. Because the ventricles in AD patients are often enlarged, and the brain is deformed compared to a normal healthy subject. The motivation of this work is to propose

a DL-based image registration which is efficacious in challenging scenarios. This can be realized by carefully designing the loss function for the model's training to estimate large deformations.

Some recent unsupervised and semi-supervised DL-based methods perform challenging MR image registrations. Huang *et al.* [45] proposed an unsupervised inter-contrast MR image registration. They used an MI-based loss function to estimate large deformations between the subjects with stroke. Hoffmann *et al.* [46] proposed SynthMorph, a contrast-invariant image registration method that can successfully align brain MR images with different contrasts.

CR [47, 48] has been used successfully in traditional inter-modal and inter-contrast registrations. Rivaz *et al.* [40] proposed a time-efficient method to calculate CR. The method is called RaPTOR, and it was used for the automatic registration of MR and US of the BITE dataset [127]. Later, RaPTOR was employed in other methods to do inter-modal registration of MR-US of RESECT dataset [41, 129] and MR-CT lumbar vertebrae [42]. In a recent study, a diffeomorphic algorithm of RaPTOR, namely DiffeoRaptor, was employed to perform inter-contrast registration of MR images and inter-modal MR-CT registration of abdominal data [178]. To our knowledge, RaPTOR has not been used in DL-based image registration. However, employing RaPTOR in a loss function can be computationally expensive.

In this chapter, Deep Learning-based deformable image registration using Correlation Ratio (DLCR) is proposed. Semi-supervised training is used to minimize a loss function based on CR and Dice loss. A model with U-Net architecture [73] is trained to estimate a displacement field for 3D pairwise inter-subject registration of OASIS1 and OASIS3 [49] datasets. The proposed method is compared with the learning-based VoxelMorph using Mean Square Error loss function (VM-MSE) and two traditional methods, Symmetric image Normalization using Mattes Mutual Information metric (Mattes MI+SyN) and NiftyReg. The method is evaluated using the Dice scores of overlapping subcortical brain structures and tissues. Given the Dice scores, the method shows comparable improvement in image alignments with other methods. The contribution of this work is proposing a

new loss function to perform T1w-T1w image registration efficiently. The loss function is based on CR. In RaPTOR [40], Parzen windowing was used to calculate its derivative analytically. Instead, we use Kernel Density Estimation (KDE) using the derivative of logistic regression function [204] to make the loss function differentiable. This study intends to design an algorithm to perform intra-modal image registration [198].

5.1 Methodology

In this section, a background of CR is presented. Then, calculation of CR, DLCR, and implementation details of DLCR are elaborated.

5.1.1 Correlation Ratio (CR)

We choose the sample space, a d-dimensional torus with periodic boundary conditions $\Omega = \mathbb{R}^d/\mathbb{Z}^d$. Let $\mathcal{A}$ be a σ-algebra of Ω. Given two images, $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$, to be registered, which are measurable functions with respect to $\mathcal{A}$. Since the images are randomly selected, they can be viewed as random variables. In general, the function $\psi : \Omega \to \Omega$, which maps X to Y, may not exist. That means the images may not be functionally dependent. Given the probability function $\mathcal{P} : \mathcal{A} \to [0, 1]$, the expected value of X defined in the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is given by the Lebesque integral $E(X) = \int_\Omega X \mathrm{d}\mathcal{P}$. According to this definition, ψ can be selected such that it minimizes $\mathrm{Var}\big(Y - \psi\big(X\big)\big)$ where $Var(X) = \int_\Omega (X - E(X))^2 \mathrm{d}\mathcal{P}$ indicates the variance. Therefore, finding the optimal function $\psi^* : \Omega \to \Omega$ can be formulated as in the equation below:

$$\psi^* = \arg\min_{\psi} \mathrm{Var}\big[Y - \psi\big(X\big)\big] \tag{5.1}$$

It is known that Eq. 5.1 can be uniquely minimized, and the solution will be the conditional expectation of Y, given X [205], i.e. $E\big(Y|X\big) = \psi^*\big(X\big)$. This fact motivates the definition of CR as in Eq. 5.2:

$$\eta(Y|X) = \frac{\mathrm{Var}[E[Y|X]]}{\mathrm{Var}[Y]} = 1 - \frac{\mathrm{Var}[Y - E[Y|X]]}{\mathrm{Var}[Y]} \tag{5.2}$$

where $\eta(Y|X)$ varies between 0 and 1. It approaches 1 when there is a strong functional dependency between X and Y and it is 0 when there is no functional relationship.

Previous works proposed different methods to calculate CR. A nonparametric mapping was suggested by Roche *et al.* [47,48] to calculate CR over the iso-sets of X. RaPTOR [40] calculates CR over local patches. Instead of iso-sets of X, histograms with Parzen windowing technique were used to calculate the derivative of RaPTOR analytically.

5.1.2 DLCR

Given the definitions in Section 5.1.1, let X be the fixed image and Y be the image registered to X. The formulation allows the images to be 2D or higher dimensions; in our case, the images are 3D. We assume the images are affinely registered as part of the preprocessing. It is intended to estimate a deformation field $\phi : \Omega \to \Omega$ such that the deformed moving image $Y \circ \phi$ is similar to X. The U-Net architecture of Fig 5.1 is employed to derive the deformation field. Other neural network architectures may be applicable; the U-Net architecture was successfully employed and tested in several DL-based image registrations.

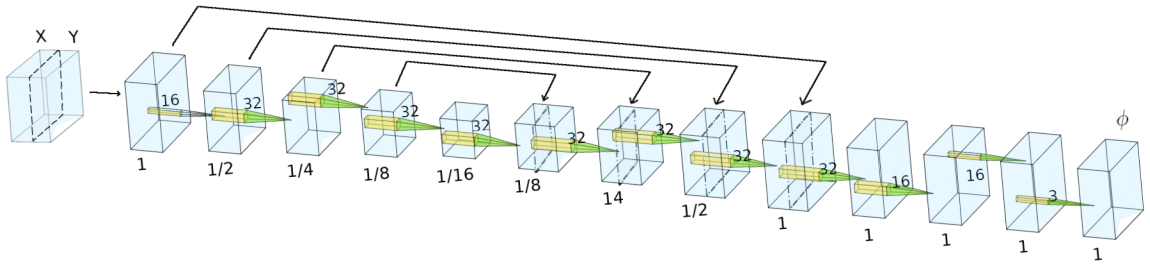

Figure 5.1: The employed U-Net architecture to estimate the deformation field ϕ is presented. Image volumes are shown in blue cubes. Besides each volume, the resolution compared to inputs X and Y is indicated. The convolution size is shown near the filters with yellow color. Arrows show the duplication of volumes.

Let $\mathcal{F}(\Omega)$ be the space of smooth deformation fields associated with Ω and $\phi \in \mathcal{F}(\Omega)$. The tangent velocity space at the identity $\mathrm{id} \in \mathcal{F}(\Omega)$ is denoted by $V = T_{\mathrm{id}}\mathcal{F}(\Omega)$. Then,

the U-Net model $\mathcal{M}$ with parameters θ approximates the velocity $u \in V$ given the input images with the equation $u = \mathcal{M}_\theta(X, Y)$. The time-dependent deformation fields $\varphi_t(\omega) = \varphi(t, \omega), \quad \omega \in \Omega$ are defined such that $\varphi_0(\omega) = \varphi(0, \omega) = \mathrm{id}$ and $\varphi_1(\omega) = \varphi(1, \omega) = \phi(\omega)$. The smooth deformation field ϕ associated with u can be obtained by solving the ODE $\frac{d\varphi_t}{dt} = v_t \circ \varphi_t$, where the velocity at time $t = 1$ is $v_1 = u$. We integrate the velocity field v_t and use the Euler method to solve the ODE given the boundary conditions $\varphi_0 = \mathrm{id}$ and $v_1 = u$. The tripartite loss function ($\mathcal{L}_{ss}$) of the equation below is employed for semi-supervised learning of the U-Net model.

$$\mathcal{L}_{ss}(X, Y, X_{seg}, Y_{seg}, \phi) = \mathcal{L}_{sim}(X, Y \circ \phi) + \gamma \mathcal{L}_{seg}(X_{seg}, Y_{seg} \circ \phi) + \lambda \mathcal{L}_{reg}(\phi) \tag{5.3}$$

where $\mathcal{L}_{sim}(X, Y \circ \phi) = 1 - \eta(Y \circ \phi | X)$ is the image dissimilarity metric that is built on CR, $\mathcal{L}_{seg}(X_{seg}, Y_{seg} \circ \phi)$ is Dice loss which X_{seg} and $Y_{seg} \circ \phi$ are labelled images of X and $Y \circ \phi$, respectively, $\mathcal{L}_{reg}(\phi)$ is the regularization loss, and λ and γ are weight constants. Minimizing the loss $\mathcal{L}_{ss}$ can obtain optimal θ^* parameters; in other words, $\theta^* = \arg\min_\theta \mathcal{L}_{ss}$. The calculation of each term of the loss function is elaborated in the following.

Similar to [40], we perform nonparametric estimation of $E(Y|X)$ and bining of X values to calculate CR (Eq.5.4).

$$1 - \eta(Y|X) = \frac{1}{N\sigma^2}\left(\sum_{i=1}^{N} y_i^2 - \sum_{j=1}^{N_b} N_j \mu_j^2\right) \tag{5.4a}$$

$$\mu_j = \frac{\sum_{i=1}^{N} \lambda_{ij} y_i}{N_j}, N_j = \sum_i \lambda_{ij} \tag{5.4b}$$

where N is the number of voxels in image Y, $\mathrm{Var}(Y) = \sigma^2$ is the variance of Y, y_i is the intensity of voxel i in image Y, N_b is the number of bins, and λ_{ij} is the contribution of x_i (the intensity of voxel i in image X) to the bin j. Instead of Parzen windowing with linear kernel, KDE using the derivative of logistic regression function is utilized. This makes the derivative of CR available for optimization purposes using the automatic differentiation in

PyTorch [206]. A similar approach to Avi-Aharon *et al.* [204] is taken to calculate the value of the j^{th} bin in a differentiable histogram (see the equation below).

$$\lambda_{ij} = \varsigma\left(\frac{x_i - c_j + L/2}{W}\right) - \varsigma\left(\frac{x_i - c_j - L/2}{W}\right) \tag{5.5}$$

where $\varsigma(z) = \frac{1}{1+e^{-z}} \quad z \in \mathbb{R}$ is the logistic regression function, $L = \frac{1}{N_b}$ is the bin width, $c_j = -1 + L(j + \frac{1}{2})$ are the bin centres, and $W = \frac{L}{2.5}$. In fig. 5.2, the histogram of a subject's MRI is shown with the KDE differentiable curve using Eq. 5.5.

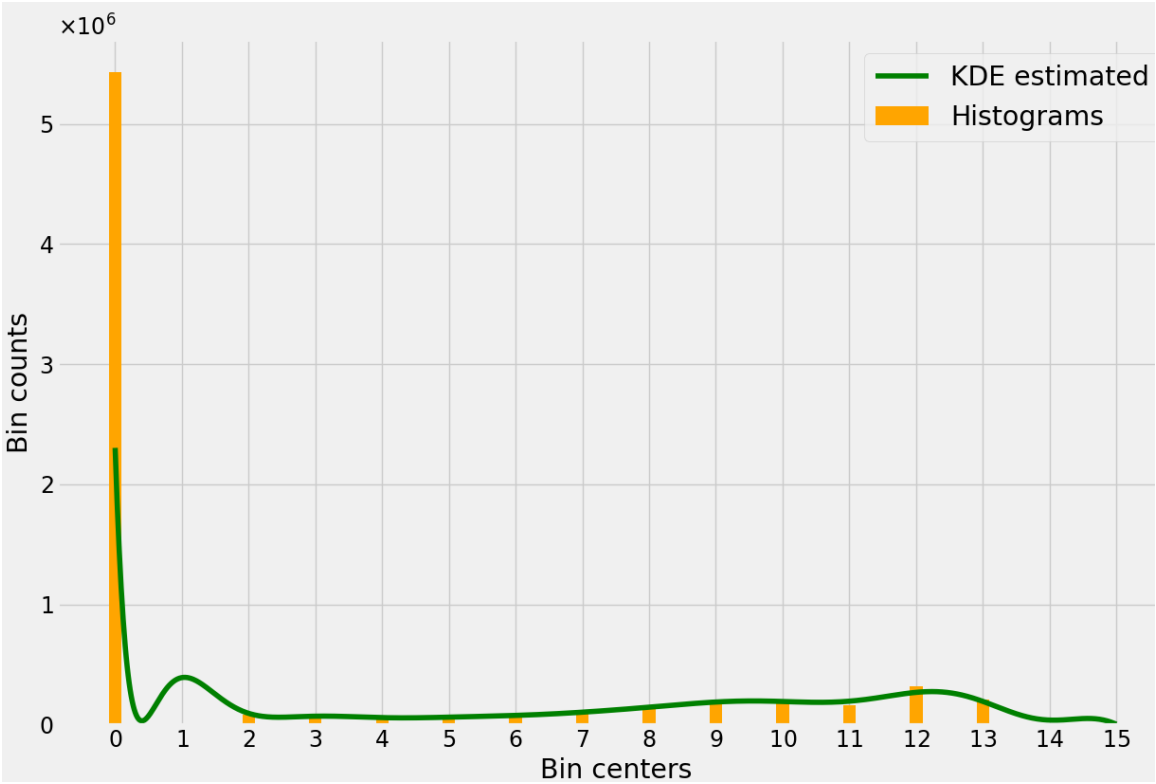

Figure 5.2: The histogram of a subject's MRI (orange) is shown with the KDE differentiable curve using the derivative of the logistic regression function (green).

In labelled images, X_{seg} and Y_{seg}, the anatomical structures are segmented and annotated into K integer values. The image background is commonly annotated with 0. Let $\mathcal{X}_k(X_{seg})$ and $\mathcal{Y}_k(Y_{seg} \circ \phi)$ denote the set of voxels that labelled with integer $k \in [1, K]$ in X_{seg} and $Y_{seg} \circ \phi$, respectively. The Dice score for structure k and the Dice loss for the

entire labelled images are calculated by Eq. 5.6a and Eq. 5.6a, respectively.

$$\mathrm{Dice}\big(\mathcal{X}_k(X_{seg}), \mathcal{Y}_k(Y_{seg} \circ \phi)\big) = 2 \cdot \frac{|\mathcal{X}_k(X_{seg}) \cap (\mathcal{Y}_k(Y_{seg}) \circ \phi)|}{|\mathcal{X}_k(X_{seg})| + |\mathcal{Y}_k(Y_{seg} \circ \phi)|} \tag{5.6a}$$

$$\mathcal{L}_{seg}(X_{seg}, Y_{seg} \circ \phi) = -\frac{1}{K} \sum_{k=1}^{K} \mathrm{Dice}\big(\mathcal{X}_k(X_{seg}), \mathcal{Y}_k(Y_{seg} \circ \phi)\big) \tag{5.6b}$$

where $|\mathcal{X}_k(X_{seg})|$ is the cardinal number of the set $\mathcal{X}_k(X_{seg})$. $\mathcal{L}_{seg}$ is near -1 when X_{seg} and Y_{seg} are similar and 0 when the labelled images are completely different.

There are several applicable formulations of the regularization term $\mathcal{L}_{reg}(\phi)$ in Eq. 5.3. In [207], a mathematical proof was established to demonstrate that a regularization term of the form of Sobolev norm of velocity field over the smooth deformation fields can ensure diffeomorphic image registration. Later this theory was validated in the well-established LDDMM methods [3,15]. A similar approach was employed in [12] to approximate diffeomorphic image registration. Analogously, we formulate the regularization term in Eq. 5.7 as a squared Sobolev norm of the velocity fields $||u \circ \phi||_V^2$ on the space $V = \{u : u \in V\}$.

$$\mathcal{L}_{reg}(\phi) = ||u \circ \phi||_V^2 = \int_{\omega \in \Omega} ||D(u \circ \phi)||_2^2 \mathrm{d}\omega \tag{5.7}$$

where $D(u \circ \phi)$ is the derivative operator and $||D(u \circ \phi)||_2^2 = \langle D(u \circ \phi), D(u \circ \phi) \rangle_{\mathcal{W}}$ is the squared L^2-norm of the vector $D(u \circ \phi)$ which takes value on the tangent space $\mathcal{W} = TV$ and inner-product $\langle , \rangle_{\mathcal{W}}$ operator is defined in the space $\mathcal{W}$. In practice, the vector $D(u \circ \phi)$ is a Jacobian matrix and the inner product is the Frobenius norm $\langle D(u \circ \phi), D(u \circ \phi) \rangle_W = ||D(u \circ \phi)||_F^2$.

5.2 Experiments

The experimental setup and the experiments to validate our method is presented in this section. Then, the comparison of DLCR with other methods is demonstrated.

5.2.1 Experimental Setup

This section includes details about datasets, evaluation metrics, preprocessing steps, and parameter setups.

Datasets

T1w MRI from subjects of three public datasets is employed for training, validation, and testing. Twenty-five subjects of the IXI dataset http://brain-development.org/ixi-dataset, twenty subjects of the OASIS3 dataset [49], and twenty subjects of the OASIS1 [50] are randomly selected. The IXI subjects' MRIs are collected from normal healthy subjects. OASIS1 dataset MRI data are collected from young, middle=aged, nondemented and demented older adults. OASIS3 consists of subjects with longitudinal MRIs collected in different visits to investigate AD progression. The data for their latest visit, which shows visible atrophies, are chosen for our experiments. The IXI subjects and fourteen OASIS3 subjects are used for training. The permutation of these subjects gives $(25 + 14) \times (25 + 14 - 1) = 1482$ images for pairwise registration. The permutation of the remaining six subjects of OASIS3 gives $6 \times (6 - 1) = 30$ image pairs, and the twenty subjects of OASIS1 provide $20 \times (20 - 1) = 380$ image pairs for validation and testing, respectively.

Evaluation metric

The Dice similarity metric for overlapping regions is used to measure the alignment of registered images. Dice scores for brain tissues, cortical, and subcortical structures are calculated according to Eq. 5.6a.

Preprocessing

First, the images were skull-stripped using nonlocal intracranial cavity extraction [192]. Then, they were resampled to $1 \times 1 \times 1mm^3$. The images were resized to $176 \times 208 \times 176$. The image sizes were fixed for convenience throughout the experiments. Then, the images

were transformed to the MNI Talairach-like brain-based coordinate space [208] and affinely registered to the ICBM152 template [194]. We carefully inspected the images to ensure linear alignments were successful. For each subject of IXI and OASIS3 datasets, two automatic segmentations were generated by the online tool volBrain [193]. In the first one, the segmentation of brain tissues classifies the cerebrum into CSF, GM, and WM. The second one annotates sixteen subcortical structures of the brain. For the OASIS1 dataset, manual segmentations were outlined by highly trained neuroanatomical technicians using Neuromorphemtrics' software [209]. Neuroanatomical delineation of each subject's MRI was carried out at Neuromophometrics, Inc. (http://neuromorphometrics.com), using the methods described in Caviness *et al.* [210].

Implementation details

The method is implemented using PyTorch [206] and built on the VM framework [12]. Adam optimizer [211] is used with a learning rate of 10^{-4}. The implementation allows for mini-batch stochastic optimization. In our experiments, one image volume per batch produced the best results. The weight parameters were set to $\lambda = 1$ and $\gamma = 1$ in Eq. 5.3. The method is trained with 6000 epochs. For each epoch, the loss of $\mathcal{L}_{ss}$ and its parts $\mathcal{L}_{sim}$, $\mathcal{L}_{seg}$, and $\mathcal{L}_{reg}$ are demonstrated in Fig. 5.3. $\mathcal{L}_{ss}$, $\mathcal{L}_{sim}$, and $\mathcal{L}_{seg}$ generally decrease with epochs.

Baseline methods

DLCR is compared with Mattes MI+SyN, NiftyReg, and VM-MSE. Mattes MI+SyN and NiftyReg are traditional image registration methods, and VM-MSE is a DL-based method. Mattes MI+SyN is a diffeomorphic image registration method available in Advanced Normalization Tools [191]. The method uses Mattes MI as the similarity metric and models the deformation fields with SyN [5]. The number of bins for Mattes MI was set to 32, and the gradient step, the update field variance, and the total field variance were set to 0.5, 3, and 0.5 for SyN, respectively. NiftyReg was employed with normalized mutual information as the similarity metric, cubic b-splines control points for modelling free-form deforma-

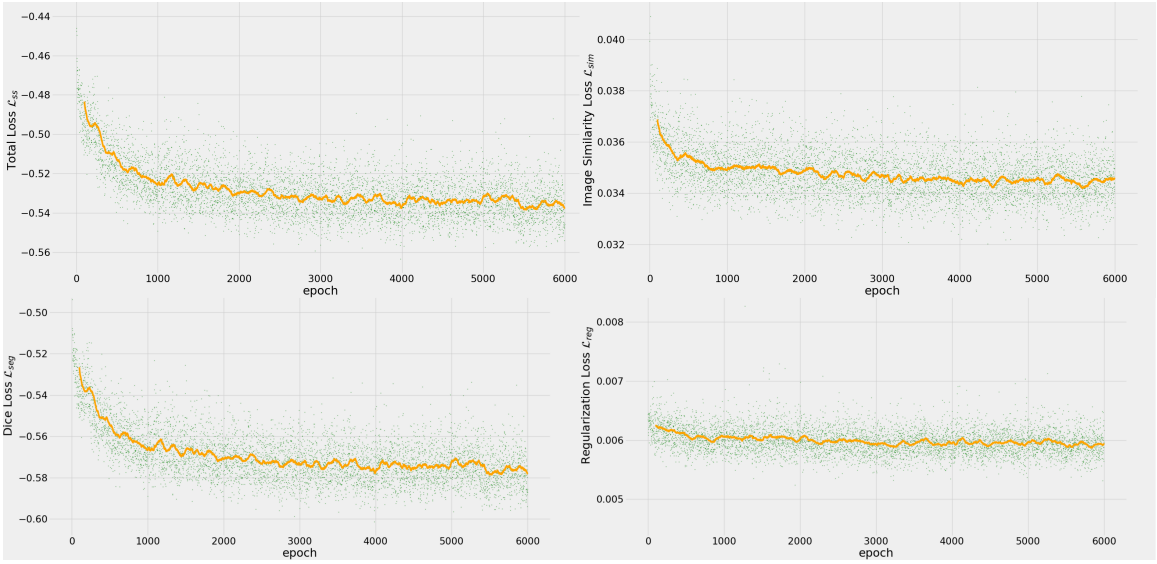

Figure 5.3: The loss $\mathcal{L}_{ss}$ (top left) and its parts $\mathcal{L}_{sim}$ (top right), $\mathcal{L}_{seg}$ (below left), and $\mathcal{L}_{reg}$ (below right) are demonstrated for each epoch. Loss function values are shown in green points, and the moving average of these data is plotted in orange.

tion fields, and conjugate gradient descent optimization with a maximum of 150 iterations. The default parameters of the software package were used. The baseline methods' parameters were tuned to produce optimal results. Mattes MI+SyN and NiftyReg were run on a Linux Mint system six CPU core. Parallel computing in the software packages was utilized to boost the computation time. VM-MSE was trained with a similar parameters setup described in Section 5.2.1. The batch size of four-volume images was used because it produced better results for us.

5.2.2 Inter-subject registrations of the OASIS3 dataset

Given six subjects of the OASIS3 dataset, $6 \times (6 - 1) = 30$ image pairs are available for registration. These sets of registration cases are used for validation and tuning the hyperparameters. The set defines a rather challenging task because the patients' ventricles are enlarged compared to healthy patients. The alignment quality is validated over sixteen subcortical structures and brain tissues using the Dice score of overlapping regions. The results are summarized in Table. 5.1. It can be seen that our method outperformed the DL-based VM-MSE method. However, traditional methods Mattes MI+SyN and NiftyReg

achieved better performance.

Table 5.1: Mean Dice score evaluation of the OASIS3 dataset for DLCR, Mattes MI+SyN, NiftyReg, and VM-MSE in overlapping regions of brain tissues and sixteen subcortical structures are presented. Note that standard deviation (std) is also calculated for each case.

Task and Evaluation Region	Affine only	DLCR	Mattes MI+SyN	NiftyReg	VM-MSE
T1-T1 Brain Tissues	0.51 ± 0.07	0.62 ± 0.05	0.67 ± 0.04	0.72 ± 0.04	0.53 ± 0.07
T1-T1 Subcortical Structures	0.60 ± 0.10	0.72 ± 0.08	0.77 ± 0.07	0.77 ± 0.07	0.63 ± 0.08

In Fig. 5.4 , the registered images using different methods are presented. It can be seen that our method shows comparable alignment of image with the other methods.

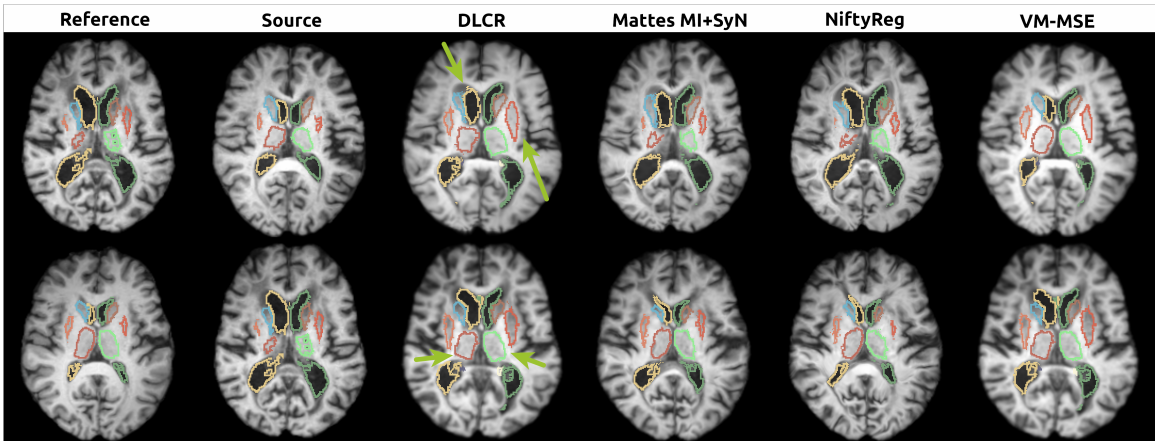

Figure 5.4: From left to right: axial slices of the reference volume, the source volume, and the performances of DLCR, Mattes MI+SyN, and NiftyReg are shown, respectively. Rows show the same slice of different subjects. Subcortical structural segmentations are shown in colored contours. The arrows for DLCR results show visible improvements in image alignments.

5.2.3 Inter-subject registrations of the OASIS1 dataset with manual segmentations

Given twenty subjects of the OASIS1 dataset, $20 \times (20 - 1) = 380$ image pairs are available for registration. The alignment quality is validated over 229 cortical and subcortical structures using the Dice score of overlapping regions. Manual segmentations are available for these subjects. Rholling *et al.* [195] showed that using automatic segmentations and a

similarity metric to evaluate the registration quality can be inaccurate in some cases. They demonstrated that using manual segmentations can mitigate this problem. Therefore, the evaluation of registration quality using manual segmentations can further validate the effectiveness of our method. Similar to Section. 5.2.3, the method is compared with Mattes MI+SyN, NiftyReg, and VM-MSE. The results are summarized in Table. 5.2. It can be seen that our method has a similar performance compared to VM-MSE. However, Mattes MI+SyN and NiftyReg achieved better Dice scores. In Fig. 5.5, the registered images using different methods are presented. It can be seen that DLCR could improve the alignment of images. The average test-time of each registration for DLCR is $2.57s$ on the CPU of a Ubuntu 20.04 LTS with four CPUs of AMD EPYC 7502 32-Core Processor, and $64.8ms$ on a NVIDIA A100 SXM4 GPU with 40GB RAM.

Table 5.2: Mean Dice score evaluation of the OASIS1 dataset for DLCR, Mattes MI+SyN, NiftyReg, and VM-MSE in overlapping regions of 229 cortical and subcortical structures is presented. Note that standard deviation (std) is also calculated for each case.

Task and Evaluation Region	Affine only	DLCR	Mattes MI+SyN	NiftyReg	VM-MSE
T1-T1 Cortical and Subcortical Structures	0.24 ± 0.22	0.29 ± 0.26	0.32 ± 0.29	0.33 ± 0.29	0.29 ± 0.26

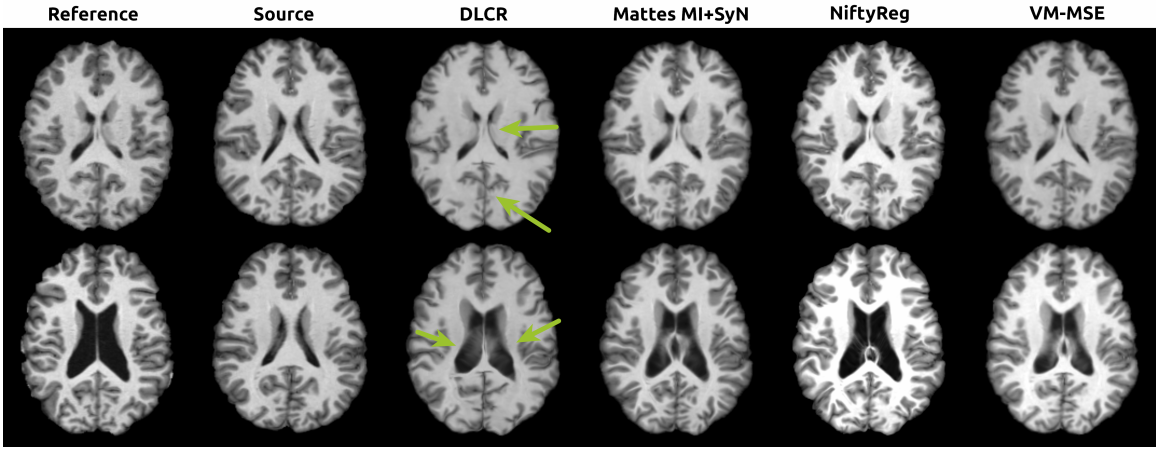

Figure 5.5: From left to right: axial slices of the reference volume, the source volume, result of DLCR, Mattes MI+SyN, and NiftyReg are shown, respectively. Rows show the same slice of different subjects. The arrows for DLCR results show visible improvements in image alignments.

5.3 Discussions

Given the training data, the model is trained to perform T1w-T1w registrations. The validation and testing of the model are also done for T1w-T1w image pairs. Trained weights of the model can be used for inter-contrast T1w-T2w and T1w-PDw registration. Therefore, more data are required to be added to training, validation, and test sets. Adding an ablation study of the method can complement the experiments to better show the difference between the DL-based and baseline methods. The deformation smoothness analysis of the DL-based and traditional methods can be helpful. The analysis purports to quantitatively show the difference between comparing methods and diffeomorphic image registration. Diffeomorphic image registration is considered superior to non-diffeomorphic ones because they generate smooth and invertible deformation fields (diffeomorphisms) that can be employed in various applications such as quantifying inter-subject variability of the brain [179] and brain atlas construction [181].

5.4 Summary

A model is trained to perform intra-contrast MR T1w-T1w registrations. The training is conducted on IXI and the OASIS3 datasets. It is validated and tested on OASIS3 and OASIS1 datasets, respectively. The proposed image registration method, i.e. DLCR, is compared with Mattes MI+SyN, NiftyReg, and VM-MSE. DLCR showed improvement in image alignments, could outperform the comparing DL-based method VM-MSE given the qualitative validation and achieved better Dice scores of overlapping cortical and subcortical structures.

Chapter 6

Conclusion

Section 6.1 summarizes the methods elaborated in the previous chapters. Given the presented methods, Section 6.2 illustrates the future direction of this thesis.

6.1 Concluding remarks

Chapter 1 gave a necessary introduction to image registration. Chapter 2 provided a literature review of DL approaches in US-guided surgery. Given the methods' validation datasets, we studied the techniques in six clinical applications: cardiac catheterization, brachytherapy, regional anesthesia, liver ablation therapy, and brain glioma resection. The methods were summarized, and we identified the unmet clinical needs.

Inter-contrast/modal medical image registration is a challenging problem, and the algorithms are application specific. Some registration problems involving US imaging are ill-posed. Previous chapters showcased novel techniques to register MR, CT, and US images. Chapter 3 presented a public database of US and CT lumbar vertebrae scans. The image acquisition and the potential application of the database were provided. The database consists of simulated and *ex-vivo* data. Besides, two rigid image registration methods were presented: one based on CR and the other based on NCC. The methods were validated and compared in exhaustive CT-US registration problems.

Chapter 4 elaborated on a diffeomorphic image registration approach. The method

was named DiffeoRaptor, and it was validated using inter-contrast brain MRIs. Furthermore, the method could successfully register abdominal CT-MR images. The deformation smoothness analysis showed that DiffeoRaptor produced smoother deformation fields than the state-of-the-art methods.

Chapter 5 presented a DL-based image registration approach based on CR. A model was trained in a semi-supervised manner given the generated brain segmentations. The method showed improvement in image alignments in terms of Dice score, compared with three other methods.

6.2 Future direction

Validation of registration methods in *in-vivo* datasets is crucial to show the potential of their applicability in clinics. The presented database in Chapter 2 can be expanded by adding *in-vivo* data. Public datasets of *in-vivo* CT and US scans of vertebrae are rare, and these datasets can help validate methods in US-guided surgeries. However, such data collection requires ethical approval, planning image acquisitions, and post-processing. The presented CT and US datasets can be employed to validate CT and US image segmentation techniques. Segmentation of US images is challenging, and these datasets provide a valuable resource for developing such techniques. Chapter 2 elaborated on two rigid image registration methods. Nevertheless, the datasets have the potential to perform experiments using affine and deformable image registrations.

Chapter 3 presented the validation of DiffeoRaptor [178] in inter-modal registration of abdominal CT-MR. Image registration of RESECT and BITE datasets using DifeoRaptor can be investigated. The performance of DiffeoRaptor in those datasets can be compared with RaPTOR [40] and ARENA [41]. DiffeoRaptor, RaPTOR, and ARENA use an identical similarity metric but employ diffeomorphic, B-spline, and affine transformations, respectively. This study may shed some insights into the choice of transformation model when the validation datasets and similarity metric is unchanged. The performance of DiffeoRaptor in image alignments and deformation smoothness should be compared with

learning-based methods. Several DL-based methods estimate smooth deformation fields; however, they do not guarantee diffeomorphic registrations. Building a DL-based method on DiffeoRaptor is another possible future direction. Since DiffeoRaptor produces smooth deformation fields, this DL-based method may show great potential in atlas building and brain inter-variability studies.

The need for time-efficient methods to analyze large datasets increases as the size of datasets grows. DL showed good performance and time efficiency in analyzing these datasets. In Chapter 5, DLCR was presented to validate intra-contrast T1w MR images. The method has the potential to be validated for inter-contrast/modal registration problems. A similar validation of DLCR to DiffeoRaptor in Chapter 4 can be interesting since DLCR reduced implementation time. The accuracy and robustness of DLCR in image alignment can be improved by expanding the training set and optimizing the training strategy. It would be interesting to train the method in a fully unsupervised manner, experiment with different loss functions, and explore more network architectures. Inter-modal registration using ViTs is an emerging area of research that can be explored and built on the proposed methods in this thesis.

References

[1] J. Mitra, Z. Kato, R. Martí, A. Oliver, X. Lladó, D. Sidibé, S. Ghose, J. C. Vilanova, J. Comet, and F. Meriaudeau, "A spline-based non-linear diffeomorphism for multimodal prostate registration", In *Medical image analysis*, 2012, 16(6), pp.1259-1279.

[2] K. S. Kutten, N. Charon, M. I. Miller, J. T. Ratnanather, J. Mateisky, A. D. Baden, "A Large Deformation Diffeomorphic Approach to Registration of CLARITY Images via Mutual Information", In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*. 2017, Lecture Notes in Computer Science, vol 10433. Springer, Cham.

[3] M. Beg, M. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms", In *International Journal of Computer Vision*, 2005, 61(2), 139–157.

[4] S. Reaungamornrat, T. D. Silva, A. Uneri, S. Vogt, G. Kleinszig, A. J. Khanna, J. P. Wolinsky, J. L. Prince, and J. H. Siewerdsen, "MIND Demons: Symmetric Diffeomorphic Deformable Registration of MR and CT for Image-Guided Spine Surgery", In *IEEE Transactions on Medical Imaging,* vol. 35, no. 11, pp. 2413-2424, Nov. 2016.

[5] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling

of elderly and neurodegenerative brain", In *Medical image analysis*, 2008, 12(1), pp.26-41.

[6] T. Vercauteren, X. Pennec, A. Perchant and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration", In *Neuroimage*, vol. 45, no. 1 Suppl, pp. S61-72, 2009.

[7] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel, "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration", In *Medical Image Analysis*, vol. 16, no. 7, pp. 1423-1435, 2012.

[8] X. Chen, A. Diaz-Pinto, N. Ravikumar, and A.F. Frangi, (2021) Deep learning in medical image registration. Prog Biomed Eng 3:012003

[9] H. R. Boveiri, R. Khayami, R. Javidan, and A. Mehdizadeh, "Medical image registration using deep neural networks: A comprehensive review", In *Computers and Electrical Engineering*, 2020, 87, p.106767.

[10] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: a review", In *Physics in Medicine and Biology*, 2020, 65(20), p.20TR01.

[11] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration–a deep learning approach", In *NeuroImage*, 2017, 158, pp.378-396.

[12] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag and A. V. Dalca, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," in IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1788-1800, Aug. 2019.

[13] T. C. Mok and A. Chung, "Fast symmetric diffeomorphic image registration with convolutional neural networks", In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, (pp. 4644-4653).

[14] J. Wang, and M. Zhang, "Deepflash: An efficient network for learning-based medical image registration" In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, (pp. 4444-4452).

[15] M. Zhang and P. T. Fletcher, "Fast Diffeomorphic Image Registration via Fourier-Approximated Lie Algebras", In *International Journal of Computer Vision* 2019, 127, 61–73.

[16] Y. Zhang, Y. Pei, and H. Zha, "Learning Dual Transformer Network for Diffeomorphic Registration', In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, 2021. Lecture Notes in Computer Science, vol 12904. Springer, Cham.

[17] M.A. Pino, A. Imperato, I. Musca, R. Maugeri, G.R. Giammalva, G. Costantino, F. Graziano, F. Meli, N. Francaviglia, D.G. Iacopino, A. Villa, New Hope in Brain Glioma Surgery: The Role of Intraoperative Ultrasound. A Review. Brain Sci. 2018, 8, 202

[18] L. Dixon, A. Lim, M. Grech-Sollars, et al. Intraoperative ultrasound in brain tumor surgery: A review and implementation guide. Neurosurg Rev 45, 2503–2515 (2022).

[19] N. Colakovic, D. Zdravkovic, Z. Skuric, et al. Intraoperative ultrasound in breast cancer surgery—from localization of non-palpable tumors to objectively measurable excision. World J Surg Onc 16, 184 (2018).

[20] T.J. Sorenson, P.J. Nicholson, C.A. Hilditch, M.H. Murad, and W. Brinjikji, 2019. A lesson from cardiology: the argument for ultrasound-guided femoral artery access in interventional neuroradiology. World neurosurgery, 126, pp.124-128.

[21] M. Gaudino, S. Sandner, G. Di Giammarco, A. Di Franco, H. Arai, T. Asai, F. Bakaeen, T. Doenst, S.E. Fremes, D. Glineur, and T.M. Kieser, 2021. The use of intraoperative transit time flow measurement for coronary artery bypass surgery: sys-

tematic review of the evidence and expert opinion statements. Circulation, 144(14), pp.1160-1171.

[22] I. Alexandru et al. Ultrasound for kidney access in percutaneous nephrolithotomy: a contemporary review. Medical Ultrasonography, [S.l.], v. 20, n. 4, p. 508-514, dec. 2018. ISSN 2066-8643.

[23] C.O.C.O. Danilo and S. Leanza, 2020. Routine Intraoperative Ultrasound for the Detection of Liver Metastases during Resection of Primary Colorectal Cancer–A Systematic Review. Maedica, 15(2), p.250.

[24] P.I. Karagyozov, I. Tishkov, I. Boeva, K. Draganov. Endoscopic ultrasound-guided biliary drainage-current status and future perspectives. World J Gastrointest Endosc. 2021 Dec 16;13(12):607-618.

[25] S.M.D. Carstensen, M.J. Velander, L. Konge, M. Østergaard, M.P. Jensen, S.A. Just, and L. Terslev, Training and assessment of musculoskeletal ultrasound and injection skills—a systematic review, Rheumatology, Volume 61, Issue 10, October 2022, Pages 3889–3901,

[26] D. Diep, K.J.Q. Chen, and D. Kumbhare, Ultrasound-guided interventional procedures for myofascial trigger points: a systematic review Regional Anesthesia and Pain Medicine 2021;46:73-80.

[27] G.F. Gómez-Chiguano, M.J. Navarro-Santana, J.A. Cleland, J.L. Arias-Buría, C. Fernández-de-Las-Peñas, R. Ortega-Santiago, and G. Plaza-Manzano, 2021. Effectiveness of ultrasound-guided percutaneous electrolysis for musculoskeletal pain: A systematic review and meta-analysis. Pain Medicine, 22(5), pp.1055-1071.

[28] B. Jiao, H. Chen, M. Chen, P. Lu, J. Liu, and C. Chen, 2022. Opioid-sparing effects of ultrasound-guided erector spinae plane block for adult patients undergoing surgery: A systematic review and meta-analysis. Pain Practice, 22(3), pp.391-404.

[29] D. Viderman, M. Aubakirova, Y. Umbetzhanov, G. Kulkaeva, S.B. Shalekenov, and Y.G. Abdildin, 2022. Ultrasound-Guided Erector Spinae Plane Block in Thoracolumbar Spinal Surgery: A Systematic Review and Meta-Analysis. Frontiers in Medicine, 9.

[30] R. Kessner, D.A. Nakamoto, V. Kondray, S. Partovi, Y. Ahmed, and N. Azar, 2019. Contrast-Enhanced Ultrasound Guidance for Interventional Procedures. Journal of Ultrasound in Medicine, 38(10), pp.2541-2557.

[31] C.P. Nolsøe, A.B. Nolsøe, J. Klubien, H.C. Pommergaard, J. Rosenberg, M.F. Meloni, and T. Lorentzen, 2018. Use of ultrasound contrast agents in relation to percutaneous interventional procedures: a systematic review and pictorial essay. Journal of Ultrasound in Medicine, 37(6), pp.1305-1324.

[32] J.M. DeWitt, M. Arain, K.J. Chang, R. Sharaiha, S. Komanduri, V.R. Muthusamy, and J.H. Hwang, 2021. Interventional endoscopic ultrasound: current status and future directions. Clinical Gastroenterology and Hepatology, 19(1), pp.24-40.

[33] M. Antico, F. Sasazawa, L. Wu, A. Jaiprakash, J. Roberts, R. Crawford, A.K. Pandey, and D. Fontanarosa, 2019. Ultrasound guidance in minimally invasive robotic procedures. Medical image analysis, 54, pp.149-167.

[34] H. Yang, C. Shan, A.F. Kolen, et al. Medical instrument detection in ultrasound: a review. Artif Intell Rev (2022). pp.1-40.

[35] P. Beigi, S.E. Salcudean, G.C. Ng, et al. Enhancement of needle visualization and localization in ultrasound. Int J CARS 16, 169–178 (2021).

[36] Y.T. Shen, L. Chen, W.W. Yue, and H.X. Xu, 2021. Artificial intelligence in ultrasound. European Journal of Radiology, 139, p.109717.

[37] X. Jia, and K. Albuquerque, 2022, October. Artificial Intelligence and Deep Learning for Brachytherapy. In Seminars in Radiation Oncology (Vol. 32, No. 4, pp. 389-399). WB Saunders.

[38] C.X. Yan, B. Goulet, J. Pelletier, S.J.S. Chen, D.Tampieri, and D.L. Collins, "Towards accurate, robust and practical ultrasound-ct registration of vertebrae for image-guided spine surgery," *International journal of computer assisted radiology and surgery*, vol. 6, no. 4, pp. 523–537, 2011.

[39] T.K. Koo and W.E. Kwok, "Hierarchical ct to ultrasound registration of the lumbar spine: a comparison with other registration methods," *Annals of biomedical engineering*, vol. 44, no. 10, pp. 2887–2900, 2016.

[40] H. Rivaz, S.J.S. Chen and D.L. Collins, "Automatic Deformable MR-Ultrasound Registration for Image-Guided Neurosurgery," in IEEE Transactions on Medical Imaging, vol. 34, no. 2, pp. 366-380, Feb. 2015.

[41] N. Masoumi, Y. Xiao, H. Rivaz, "ARENA: Inter-modality affine registration using evolutionary strategy", In *IJCARS*, 14, 441–450 (2019). https://doi.org/10.1007/s11548-018-1897-1

[42] N. Masoumi, C. J. Belasso, M. O. Ahmad, H. Benali, Y. Xiao, H. Rivaz, "Multimodal 3D ultrasound and CT in image-guided spinal surgery: public database and new registration algorithms", In *IJCARS*, 16, 555–565 (2021). https://doi.org/10.1007/s11548-021-02323-2

[43] N. Singh, J. Hinkle, S. Joshi and P.T. Fletcher, "A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction", In *IEEE 10th International Symposium on Biomedical Imaging*, 2013, pp. 1219-1222.

[44] F. P. M. Oliveira and J. M. R. S. Tavares, "Medical image registration: a review", In *Computer Methods in Biomechanics and Biomedical Engineering*, 2014, 17:2, 73-93.

[45] W. Huang, H. Yang, X. Liu, C. Li, I. Zhang, R. Wang, H. Zheng, and S. Wang, 2021. A coarse-to-fine deformable transformation framework for unsuper-

vised multi-contrast MR image registration with dual consistency constraint. IEEE Transactions on Medical Imaging, 40(10), pp.2589-2599.

[46] M. Hoffmann, B. Billot, D.N. Greve, J.E. Iglesias, B. Fischl, and A.V. Dalca, 2021. SynthMorph: learning contrast-invariant registration without acquired images. IEEE transactions on medical imaging, 41(3), pp.543-558.

[47] A. Roche, G. Malandain, X. Pennec, N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration", In *Medical Image Computing and Computer-Assisted Intervention — MICCAI98*, 1998, Lecture Notes in Computer Science, vol 1496. Springer, Berlin, Heidelberg.

[48] A. Roche, G. Malandain, N. Ayache, and S. Prima, "Towards a better comprehension of similarity measures used in medical image registration", In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1999, September (pp. 555-566). Springer, Berlin, Heidelberg.

[49] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. Vlassenko, and M. E. Raichle, "OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease", 2019, MedRxiv,

[50] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults", In *Journal of cognitive neuroscience*, 2007, 19(9), pp.1498-1507.

[51] J.A. Long et al 2012. Development of a novel robot for transperineal needle based interventions: focal therapy, brachytherapy and prostate biopsies. J. Urol. 188, 1369–1374.

[52] A. Pourtaherian et al., ”Medical Instrument Detection in 3-Dimensional Ultrasound Data Volumes,” in *IEEE Transactions on Medical Imaging*, vol. 36, no. 8, pp. 1664-1675, Aug. 2017.

[53] S. G. Yuen et al., (2008). 3D Ultrasound-Guided Motion Compensation System for Beating Heart Mitral Valve Repair. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008. MICCAI 2008. Lecture Notes in Computer Science, vol 5241. Springer, Berlin, Heidelberg.

[54] F. Palhano Xavier de Fontes, G. Andrade Barroso, P. Coupé, *et al.*, Real time ultrasound image denoising. J Real-Time Image Proc 6, 15–22 (2011).

[55] H.D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, 2010. Automated breast cancer detection and classification using ultrasound images: A survey. Pattern recognition, 43(1), pp.299-317.

[56] W.Y. Chan, J. Qin, Y.P. Chui, and P.A. Heng, ”A Serious Game for Learning Ultrasound-Guided Needle Placement Skills,” in IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 6, pp. 1032-1042, Nov. 2012,

[57] A. Belaid, D. Boukerroui, Y. Maingourd and J.F. Lerallut, ”Phase-Based Level Set Segmentation of Ultrasound Images,” in IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 1, pp. 138-147, Jan. 2011.

[58] H. Zhou and H. Rivaz, ”Registration of Pre- and Postresection Ultrasound Volumes With Noncorresponding Regions in Neurosurgery,” in IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 5, pp. 1240-1249, Sept. 2016.

[59] A.K.Z. Tehrani, I.M. Rosado-Mendez, and H. Rivaz, ”Robust Scatterer Number Density Segmentation of Ultrasound Images,” in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 69, no. 4, pp. 1169-1180, April 2022

[60] N. Masoumi, H. Rivaz, I. Hacihaliloglu, M.O. Ahmad, I. Reinertsen, and Y. Xiao, “The big bang of deep learning in ultrasound-guided surgery: a review”, March, 2023.

[61] H. Yang, C. Shan, A.F. Kolen, et al. Catheter localization in 3D ultrasound using voxel-of-interest-based ConvNets for cardiac intervention. Int J CARS 14, 1069–1077 (2019).

[62] H. Yang, C. Shan, T. Tan, A.F. Kolen, P.H.N. de With, (2019). Transferring from ex-vivo to in-vivo: Instrument Localization in 3D Cardiac Ultrasound Using Pyramid-UNet with Hybrid Loss. In: , et al. Medical Image Computing and Computer Assisted Intervention - MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science(), vol 11768. Springer, Cham.

[63] H. Yang, C. Shan, A. F. Kolen and P. H. N. de With, ”Automated Catheter Localization in Volumetric Ultrasound Using 3D Patch-Wise U-Net with Focal Loss,” 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 1346-1350.

[64] H. Yang, C. Shan, A. F. Kolen and P.H.N. de With, ”Catheter Detection in 3D Ultrasound Using Triplanar-Based Convolutional Neural Networks,” 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 371-375.

[65] H. Yang, C. Shan, A. Bouwman, A.F. Kolen, and P.H. de With, 2021. Efficient and robust instrument segmentation in 3D ultrasound using patch-of-interest-FuseNet with hybrid loss. Medical Image Analysis, 67, p.101842.

[66] H. Yang, C. Shan, A. Bouwman, L.R.C. Dekker, A.F. Kolen and P. H. N. de With, ”Medical Instrument Segmentation in 3D US by Hybrid Constrained Semi-Supervised Learning,” in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 2, pp. 762-773, Feb. 2022.

[67] H. Yang, C. Shan, A.F. Kolen, and P.H. de With, 2022. Weakly-supervised learning for catheter segmentation in 3D frustum ultrasound. Computerized Medical Imaging and Graphics, 96, p.102037.

[68] H. Yang, C. Shan, A.F. Kolen, and P.H.N. de With, (2020). Deep Q-Network-Driven Catheter Segmentation in 3D US by Hybrid Constrained Semi-supervised Learning and Dual-UNet. In: , et al. Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science(), vol 12261. Springer, Cham.

[69] H. Yang, C. Shan, A. F. Kolen and P. H. N. de With, "Efficient Catheter Segmentation in 3D Cardiac Ultrasound using Slice-Based FCN With Deep Supervision and F-Score Loss," 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 260-264.

[70] H. Yang, C. Shan, A. F. Kolen and P. H. N. de With, "Improving Catheter Segmentation and Localization in 3D Cardiac Ultrasound Using Direction-Fused Fcn," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 1122-1126.

[71] L. Min, H. Yang, C. Shan, A.F. Kolen, and P.H.N. de With "Feasibility study of catheter segmentation in 3D Frustum ultrasounds by DCNN", Proc. SPIE 11315, Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, 1131521 (16 March 2020).

[72] A.F. Frangi, W.J. Niessen, K.L. Vincken, and M.A. Viergever, (1998) Multiscale vessel enhancement filtering. In: International conference on medical image computing and computer-assisted intervention, Springer, Berlin, Heidelberg, pp 130–137

[73] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, and O. Ronneberger, (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds) Medical Image

Computing and Computer-Assisted Intervention - MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science(), vol 9901. Springer, Cham.

[74] A. Karpathy, G. Toderici , S. Shetty, T. Leung , R. Suk- thankar, and F. Li, "Large-scale video classification with convolutional neural networks," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2014, pp. 1725-1732

[75] K. Simonyan, and A. Zisserman, 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[76] S. Chen, K. Ma, and Y. Zheng, Med3d: Transfer learning for 3d medical image analysis, arXiv:1904.00625 (2019)

[77] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, 2016. Learning deep features for discriminative localization. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2921–2929.

[78] Y. Zhang, Y. Lei, R.L. Qiu, T. Wang, H. Wang, A.B. Jani, W.J. Curran, P. Patel, T. Liu, and X. Yang, 2020. Multi-needle localization with attention U-Net in US-guided HDR prostate brachytherapy. Medical physics, 47(7), pp.2735-2745.

[79] Y. Zhang, Z. Tian, Y. Lei, T. Wang, P. Patel, A.B. Jani, W.J. Curran, T. Liu, and X. Yang, 2020. Automatic multi-needle localization in ultrasound images using large margin mask RCNN for ultrasound-guided prostate brachytherapy. Physics in Medicine and Biology, 65(20), p.205003.

[80] C. Andersén, T. Rydén, P. Thunberg, and J.H. Lagerlöf, 2020. Deep learning-based digitization of prostate brachytherapy needles in ultrasound images. Medical physics, 47(12), pp.6414-6420.

[81] F. Wang, L. Xing, H. Bagshaw, M. Buyyounouski, and B. Han, 2020. Deep learning applications in automatic needle segmentation in ultrasound-guided prostate brachytherapy. Medical Physics, 47(9), pp.3797-3805.

[82] D. Liu, DS. Tupor, J. Singh, et al. The challenges facing deep learning–based catheter localization for ultrasound guided high-dose-rate prostate brachytherapy. Med Phys. 2022; 49: 2442– 2451.

[83] K.B. Girum, A. Lalande, R. Hussain, et al. A deep learning method for real-time intraoperative US image segmentation in prostate brachytherapy. Int J CARS 15, 1467–1476 (2020).

[84] K.B. Girum, G. Créhange, R. Hussain, R. et al. Fast interactive medical image segmentation with weakly supervised deep learning method. Int J CARS 15, 1437–1444 (2020).

[85] N. Orlando, D.J. Gillies, I. Gyacskov, C. Romagnoli, D. D'Souza, and A. Fenster, 2020. Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images. Medical physics, 47(6), pp.2413-2426.

[86] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang, 2018 UNet++: a nested U-net architecture for medical image segmentation Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11045 3–11

[87] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang, 2020 UNet++: redesigning skip connections to exploit multiscale features in image segmentation IEEE Trans. Med. Imaging 39 1856–67

[88] N. Orlando, I. Gyacskov, D.J. Gillies, F. Guo, C. Romagnoli, D. D'Souza, D.W. Cool, D.A. Hoover, and A. Fenster, 2022. Effect of dataset size, image quality, and image type on deep learning-based automatic prostate segmentation in 3D ultrasound. Physics in Medicine and Biology, 67(7), p.074002.

[89] F. Milletari, N. Navab, and S.A. Ahmadi. V-Net: fully convolutional neural networks for volumetric medical image segmentation. Proceedings of International Conference on 3D Vision; 2016:565–571.

[90] Y. Lei, S. Tian, X. He, T. Wang, B. Wang, P. Patel, A.B. Jani, H. Mao, W.J. Curran, T. Liu, and X. Yang, 2019. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. Medical physics, 46(7), pp.3194-3206.

[91] E.M.A. Anas, et al. (2017). Clinical Target-Volume Delineation in Prostate Brachytherapy Using Residual Neural Networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D., Duchesne, S. (eds) Medical Image Computing and Computer Assisted Intervention - MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science(), vol 10435. Springer, Cham.

[92] D. Karimi, et al. (2018). Accurate and Robust Segmentation of the Clinical Target Volume for Prostate Brachytherapy. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science(), vol 11073. Springer, Cham.

[93] D. Karimi, Q. Zeng, P. Mathur, A. Avinash, S. Mahdavi, I. Spadinger, P. Abolmaesumi, and S.E. Salcudean, 2019. Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images. Medical image analysis, 57, pp.186-196.

[94] X. He, Y. Lei, Y. Liu, Z. Tian, T. Wang, W.J. Curran, T. Liu, and X. Yang, "Deep attentional GAN-based high-resolution ultrasound imaging," Proc. SPIE 11319, Medical Imaging 2020: Ultrasonic Imaging and Tomography, 113190B (16 March 2020);

[95] M. Golshan, D. Karimi, S. Mahdavi, J. Lobo, M. Peacock, S.E. Salcudean, and I. Spadinger, 2020. Automatic detection of brachytherapy seeds in 3D ultrasound images using a convolutional neural network. Physics in Medicine and Biology, 65(3), p.035016.

[96] Y. LeCun, et al 1989 Backpropagation Applied to Handwritten Zip Code Recognition Neural Comput. 1 541–51

[97] Y. Chen, L. Xing, L. Yu, W. Liu, R. Pooya Fahimian, T. Niedermayr, H.P. Bagshaw, M. Buyyounouski, and B. Han, 2021. MR to ultrasound image registration with segmentation-based learning for HDR prostate brachytherapy. Medical Physics, 48(6), pp.3074-3083.

[98] Q. Zeng, Y. Fu, J. Jeong, Y. Lei, T. Wang, H. Mao, A.B. Jani, P. Patel, W.J. Curran, T. Liu, and X. Yang, "Weakly non-rigid MR-TRUS prostate registration using fully convolutional and recurrent neural networks," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 113132Y (10 March 2020);

[99] J.R. Rodgers, D.J. Gillies, W.T. Hrinivich, I. Gyackov, and A. Fenster, "Automatic needle localization in intraoperative 3D transvaginal ultrasound images for high-dose-rate interstitial gynecologic brachytherapy," Proc. SPIE 11315, Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, 113150K (16 March 2020);

[100] H. Sun, K. Zhang, R. Fan, W. Xiong, and J. Yang, 2019. Stepwise Local Synthetic Pseudo-CT Imaging Based on Anatomical Semantic Guidance. IEEE Access, 7, pp.168428-168435.

[101] K. He, G. Gkioxari, P. Dollár, and R. Girshick, 2017 Mask R-CNN Proc. of the IEEE Int. Conf. on Computer Vision pp 2961–9

[102] E. Schubert, J. Sander, M. Ester, H.P. Kriegel, and X. Xu, 2017 DBSCAN revisited, revisited: why and how you should (still) use DBSCAN ACM Trans. Database Syst. 42 1–21

[103] N. Abraham, and N.M. Khan. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation; 2018. arXiv:1810.07842v1

[104] E. Elhamifar, R. Vidal, 2013. Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans. Pattern Anal. Mach. Intell. 35 (11), 2765–2781.

[105] W.T. Hrinivich, D.A. Hoover, K. Surry, C. Edirisinghe, J. Montreuil, D. D'Souza, A. Fenster, and E. Wong, "Simultaneous automatic segmentation of multiple needles using 3D ultrasound for high-dose-rate prostate brachytherapy," Med. Phys. 44(4), 1234–1245 (2017).

[106] D. Viderman, M. Dossov, S. Seitenov, and M.H. Lee, 2022. Artificial intelligence in ultrasound-guided regional anesthesia: A scoping review. Frontiers in Medicine, p.3037.

[107] C. Mwikirize, J.L. Nosher, and I. Hacihaliloglu, Convolution neural networks for real-time needle detection and localization in 2D ultrasound. Int J CARS 13, 647–657 (2018).

[108] C. Mwikirize, J.L. Nosher, I. Hacihaliloglu, (2019). Single Shot Needle Tip Localization in 2D Ultrasound. In: , et al. Medical Image Computing and Computer Assisted Intervention - MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science(), vol 11768. Springer, Cham.

[109] C. Mwikirize, A.B. Kimbowa, S. Imanirakiza, et al. Time-aware deep neural networks for needle tip localization in 2D ultrasound. Int J CARS 16, 819–827 (2021).

[110] J. Gao, P. Liu, G.D. Liu, et al. Robust needle localization and enhancement algorithm for ultrasound by deep learning and beam steering methods. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 36(2): 334–346 Mar. 2021.

[111] A. Pourtaherian, et al. (2017). Improving Needle Detection in 3D Ultrasound Using Orthogonal-Plane Convolutional Networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D., Duchesne, S. (eds) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science(), vol 10434. Springer, Cham.

[112] A. Pourtaherian, N. Mihajlovic, F.G. Zanjani, S. Zinger, G.C. Ng, H.H. Korsten, and P.H. De With, 2018, October. Localization of partially visible needles in 3D ul-

trasound using dilated CNNs. In 2018 IEEE International Ultrasonics Symposium (IUS) (pp. 1-4). IEEE.

[113] E. Smistad and F. Lindseth, "Real-Time Automatic Artery Segmentation, Reconstruction and Registration for Ultrasound-Guided Regional Anaesthesia of the Femoral Nerve," in IEEE Transactions on Medical Imaging, vol. 35, no. 3, pp. 752-761, March 2016.

[114] E. Smistad, K.F. Johansen, D.H. Iversen, and I. Reinertsen, 2018. Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. Journal of Medical Imaging, 5(4), p.044004.

[115] M. Baby, and A.S. Jereesh, 2017, April. Automatic nerve segmentation of ultrasound images. In 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA) (Vol. 1, pp. 107-112). IEEE

[116] H.M. Díaz-Vargas, C.A. Jimenez-Castaño, D.A. Cárdenas-Peña, O.D. Aguirre-Ospina, A.A. Orozco-Gutierrez, (2021). Peripheral Nerve Segmentation in Ultrasound Images Using Conditioned U-Net. In: Hernández Heredia, Y., Milián Núñez, V., Ruiz Shulcloper, J. (eds) Progress in Artificial Intelligence and Pattern Recognition. IWAIPR 2021. Lecture Notes in Computer Science(), vol 13055. Springer, Cham.

[117] E. Maneas et al., "Deep Learning for Instrumented Ultrasonic Tracking: From Synthetic Training Data to In Vivo Application," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 69, no. 2, pp. 543-552, Feb. 2022,

[118] R. Girshick, (2015) Fast R-CNN. In: Proceedings of IEEE international conference on computer vision, pp 1440–148

[119] C. Mwikirize, J.L. Nosher, and I. Hacihaliloglu: Learning needle tip localization from digital subtraction in 2D ultrasound. Int. J. CARS 14(6), 1017–1026 (2019)

[120] O. Ronneberger, P. Fischer, and T. Brox, (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham.

[121] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.(CVPR), Jun. 2016, pp. 770–778.

[122] H. Rivaz, et al. (2008). Ablation Monitoring with Elastography: 2D In-vivo and 3D Ex-vivo Studies. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008. MICCAI 2008. Lecture Notes in Computer Science, vol 5242. Springer, Berlin, Heidelberg.

[123] S. Zhang et al., "Detection and Monitoring of Thermal Lesions Induced by Microwave Ablation Using Ultrasound Imaging and Convolutional Neural Networks," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 4, pp. 965-973, April 2020,

[124] M. Wang et al., "Detection and Monitoring of Microwave Ablation by Ultrasound Imaging Based on Convolutional Neural Network," 2020 IEEE International Ultrasonics Symposium (IUS), 2020, pp. 1-3,

[125] R. Kondo, N. Koizumi, Y. Nishiyama, N. Matsumoto, and K. Numata, 2018, June. Out-of-plane motion detection system using convolutional neural network for US-guided radiofrequency ablation therapy. In 2018 15th International Conference on Ubiquitous Robots (UR) (pp. 729-731). IEEE

[126] M. Arif, A. Moelker, and T. van Walsum, 2019. Automatic needle detection and real-time bi-planar needle visualization during 3D ultrasound scanning of the liver. Medical image analysis, 53, pp.104-110.

[127] L. Mercier, R.F.D. Maestro, K. Petrecca, D. Araujo, C. Haegelen, and D.L. Collins, 'Online database of clinical MR and ultrasound images of brain tumors," In *Medical Physics*, vol. 39, p. 3253, 2012.

[128] Y. Xiao, H. Rivaz, M. Chabanas, M. Fortin, I. Machado, Y. Ou, M.P. Heinrich, J.A. Schnabel, X. Zhong, A. Maier, and W. Wein, 2019. Evaluation of MRI to ultrasound registration methods for brain shift correction: the CuRIOUS2018 challenge. IEEE transactions on medical imaging, 39(3), pp.777-786.

[129] Y. Xiao, M. Fortin, G. Unsgård, H. Rivaz, and I. Reinertsen, "REtroSpective Evaluation of Cerebral Tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries", In *Medical Physics*, 2017, 44: 3875-3882. https://doi.org/10.1002/mp.12268

[130] L. Canalini, J. Klein, S. Miller, et al. Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery. Int J CARS 14, 1697–1713 (2019).

[131] L. Canalini, J. Klein, D. Miller, and R. Kikinis, 2020. Enhanced registration of ultrasound volumes by segmentation of resection cavity in neurosurgical procedures. International journal of computer assisted radiology and surgery, 15(12), pp.1963-1974

[132] R.A. Zeineldin, M.E. Karar, J. Coburger, C.R. Wirtz, F. Mathis-Ullrich, and O. Burgert, 2020. Towards automated correction of brain shift using deep deformable magnetic resonance imaging-intraoperative ultrasound (MRI-iUS) registration. Current directions in biomedical engineering, 6(1).

[133] R.A. Zeineldin, M.E. Karar, Z. Elshaer, M. Schmidhammer, J. Coburger, C.R. Wirtz, O. Burgert, and F. Mathis-Ullrich, 2021. iRegNet: non-rigid registration of MRI to interventional US for brain-shift compensation using convolutional neural networks. IEEE Access, 9, pp.147579-147590.

[134] R.A. Zeineldin, M.E. Karar, F. Mathis-Ullrich, O. Burgert. A Hybrid Deep Registration of MR Scans to Interventional Ultrasound for Neurosurgical Guidance. In: Lian, C., Cao, X., Rekik, I., Xu, X., Yan, P. (eds) Machine Learning in Medical Imaging. MLMI 2021. Lecture Notes in Computer Science(), vol 12966. Springer, Cham.

[135] A. Pirhadi, S. Salari, M.O. Ahmad, H. Rivaz, and Y. Xiao, 2022. Robust landmark-based brain shift correction with a Siamese neural network in ultrasound-guided brain tumor resection. International Journal of Computer Assisted Radiology and Surgery, pp.1-8.

[136] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, P.H. Torr, (2016) Fully-convolutional Siamese networks for object tracking. In: European conference on computer vision. Springer, pp 850–865

[137] R.A. Zeineldin, A. Pollok, T. Mangliers, M.E. Karar, F. Mathis-Ullrich, and O. Burgert, 2022. Deep automatic segmentation of brain tumours in interventional ultrasound data. Current directions in biomedical engineering, 8(1), pp.133-137.

[138] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, and Y. Zhou, 2021. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.

[139] F.X. Carton, M. Chabanas, B.K. Munkvold, I. Reinertsen, and J.H. Noble, 2020, March. Automatic segmentation of brain tumor in intraoperative ultrasound images using 3D U-Net. In Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 11315, pp. 190-195). SPIE

[140] S. Cepeda, S. García-García, I. Arrese, G. Fernández-Pérez, M. Velasco-Casares, M. Fajardo-Puentes, T. Zamora, and R. Sarabia, 2021. Comparison of intraoperative ultrasound B-mode and strain elastography for the differentiation of glioblastomas from solitary brain metastases. An automated deep learning approach for image analysis. Frontiers in Oncology, 10, p.590756.

[141] M.P. Heinrich, and L. Hansen, (2020) Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5D displacement search. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 190–200

[142] J.H. Lee, I. Joo, T.W. Kang, Y.H. Paik, D.H. Sinn, S.Y. Ha, K. Kim, C. Choi, G. Lee, J. Yi, and W.C. Bang, 2020. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. European radiology, 30(2), pp.1264-1273.

[143] D.J. Gillies, J.R. Rodgers, I. Gyacskov, P. Roy, N. Kakani, D.W. Cool, and A. Fenster, 2020. Deep learning segmentation of general interventional tools in two-dimensional ultrasound images. Medical Physics, 47(10), pp.4956-4970.

[144] Y. Wang et al., "Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound," in IEEE Transactions on Medical Imaging, vol. 38, no. 12, pp. 2768-2778, Dec. 2019,

[145] Hu, Y. et al. (2018). Adversarial Deformation Regularization for Training Image Registration Neural Networks. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds) Medical Image Computing and Computer Assisted Intervention - MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science(), vol 11070. Springer, Cham.

[146] P.U. Pandey, N. Quader, P. Guy, R. Garbi, and A.J. Hodgson, 2020. Ultrasound bone segmentation: a scoping review of techniques and validation practices. Ultrasound in Medicine and Biology, 46(4), pp.921-935.

[147] I. Hacihaliloglu, 2017. Ultrasound imaging and segmentation of bone surfaces: A review. Technology, 5(02), pp.74-80.

[148] S. Schumann, (2016). State of the Art of Ultrasound-Based Registration in Computer Assisted Orthopedic Interventions. In: Zheng, G., Li, S. (eds) Computational

Radiology for Orthopaedic Interventions. Lecture Notes in Computational Vision and Biomechanics, vol 23. Springer, Cham.

[149] G. Ayana, K. Dese, and S.W. Choe, 2021. Transfer learning in breast cancer diagnoses via ultrasound imaging. Cancers, 13(4), p.738.

[150] A. Hering et al., "Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning," in IEEE Transactions on Medical Imaging, 2022.

[151] M. Antonelli, A. Reinke, S. Bakas, et al. The Medical Segmentation Decathlon. Nat Commun 13, 4128 (2022).

[152] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, and M. Shah, 2022. Transformers in vision: A survey. ACM computing surveys (CSUR), 54(10s), pp.1-41.

[153] A. Rasoulian, S. Salari, and Y. Xiao, (2022). Weakly Supervised Intracranial Hemorrhage Segmentation Using Hierarchical Combination of Attention Maps from a Swin Transformer. In: , et al. Machine Learning in Clinical Neuroimaging. MLCN 2022. Lecture Notes in Computer Science, vol 13596. Springer, Cham.

[154] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

[155] D. Minh, H.X. Wang, Y.F. Li, et al. Explainable artificial intelligence: a comprehensive review. Artif Intell Rev 55, 3503–3568 (2022).

[156] I.J. Gerard, M. Kersten-Oertel, S. Drouin, J.A. Hall, K. Petrecca, D. De Nigris, D.A. Di Giovanni, T. Arbel, and D.L. Collins, "Combining intraoperative ultrasound brain shift correction and augmented reality visualizations: a pilot study of eight cases," *Journal of Medical Imaging*, vol. 5, no. 2, p. 021210, 2018.

[157] G.M. Della Pepa, G. Sabatino, C.L. Sturiale, E. Marchese, A. Puca, A. Olivi, and A. Albanese, "Integration of real-time intraoperative contrast-enhanced ultrasound and color doppler ultrasound in the surgical treatment of spinal cord dural arteriovenous fistulas," *World Neurosurgery*, vol. 112, pp. 138–142, 2018.

[158] M. Ganau, N. Syrmos, A.R. Martin, F. Jiang, and M.G. Fehlings, "Intraoperative ultrasound in spine surgery: history, current applications, future developments," *Quantitative imaging in medicine and surgery*, vol. 8, no. 3, p. 261, 2018.

[159] J.F. Boursier, A. Fournet, J. Bassanino, M. Manassero, A.S. Bedu, and D. Leperlier, "Ultrasonography is more accurate than percutaneous palpation for identifying targeted thoracolumbar intervertebral disc spaces in dogs,"*Veterinary Radiology & Ultrasound*, 2018.

[160] W. Wein, S. Brunke, A. Khamene, M.R. Callstrom, and N. Navab, "Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention," *Medical image analysis*, vol. 12, no. 5, pp. 577–585, 2008.

[161] O. Mattausch and O. Goksel, "Image-based reconstruction of tissue scatterers using beam steering for ultrasound simulation," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 767–780, 2018.

[162] M. Tuzer, R. Türkay, M. Boyman, and B. Acar, "Multi-ray medical ultrasound simulation without explicit speckle modelling," *International journal of computer assisted radiology and surgery*, pp. 1–9, 2018.

[163] P. Rubi, E. F. Vera, J. Larrabide, M. Calvo, J. D'Amato, and I. Larrabide, "Comparison of real-time ultrasound simulation models using abdominal ct images," in *12th International Symposium on Medical Information Processing and Analysis*, vol. 10160, p. 1016009, International Society for Optics and Photonics, 2017.

[164] B. Burger, S. Bettinghausen, M. Radle, and J. Hesser, "Real-time gpu-based ultrasound simulation using deformable mesh models," *IEEE transactions on medical imaging*, vol. 32, no. 3, pp. 609–618, 2013.

[165] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository", In *Journal of Digital Imaging*, Volume 26, Number 6, December, 2013, pp 1045-1057.

[166] C. Roche, E. Bonaccio, and J. Filippini, "Radiology data from the cancer genome atlas sarcoma [tcga-sarc] collection," *The Cancer Imaging Archive*, 2016.

[167] A. Karamalis, W. Wein, and N. Navab, "Fast ultrasound image simulation using the westervelt equation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 243–250, Springer, 2010.

[168] S. Gill, P. Abolmaesumi, G. Fichtinger, J. Boisvert, D. Pichora, D. Borshneck, and P. Mousavi, "Biomechanically constrained groupwise ultrasound to ct registration of the lumbar spine," *Medical image analysis*, vol. 16, no. 3, pp. 662–674, 2012.

[169] J. Jaros, A. P. Rendell, and B. E. Treeby, "Full-wave nonlinear ultrasound simulation on distributed clusters with applications in high-intensity focused ultrasound," *The International Journal of High Performance Computing Applications*, vol. 30, no. 2, pp. 137–155, 2016.

[170] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 1115–1124, Springer, 1998.

[171] S. Boyd, S.P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[172] J. L. Kendall and J. P. Faragher, "Ultrasound-guided central venous access: a homemade phantom for simulation," *Canadian Journal of Emergency Medicine*, vol. 9, no. 5, pp. 371–373, 2007.

[173] H.J. Wilke, A. Kettler, K.H. Wenger, and L.E. Claes, "Anatomy of the sheep spine and its comparison to the human spine," *The Anatomical Record: An Official Publication of the American Association of Anatomists*, vol. 247, no. 4, pp. 542–555, 1997.

[174] J.A. Jensen, "Field: A program for simulating ultrasound systems," in *10th nordicbaltic conference on biomedical imaging, vol. 4, supplement 1, part 1: 351–353*, Citeseer, 1996.

[175] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger, "Plus: open-source toolkit for ultrasound-guided intervention systems," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 10, pp. 2527–2537, 2014.

[176] T. Ungi, A. Lasso, and G. Fichtinger, "Open-source platforms for navigated image-guided interventions," *Medical Image Analysis*, vol. 33, pp. 181–186, 2016.

[177] B. Brendel, S.W.A. Rick, M. Stockheim, and H. Ermert, "Registration of 3d ct and ultrasound datasets of the spine using bone structures," *Computer Aided Surgery*, vol. 7, no. 3, pp. 146–155, 2002.

[178] N. Masoumi, H. Rivaz, M.O. Ahmad, and Y. Xiao. DiffeoRaptor: diffeomorphic inter-modal image registration using RaPTOR. Int J CARS (2022). https://doi.org/10.1007/s11548-022-02749-2

[179] M. Hernandez, M.N. Bossa, and S. Olmos, "Registration of Anatomical Images Using Paths of Diffeomorphisms Parameterized with Stationary Vector Field Flows", *International Journal of Computer Vision* 85, 291–306 (2009).

[180] D. Tward, T. Brown, Y. Kageyama, J. Patel, Z. Hou, S. Mori, M. Albert, J. Troncoso, and M. Miller, "Diffeomorphic registration with intensity transformation

and missing data: Application to 3D digital pathology of Alzheimer's disease", In *Frontiers in neuroscience*, 2020, 14, p.52.

[181] M. Zhang, and P. Golland, "Statistical shape analysis: From landmarks to diffeomorphisms", In *Medical Image Analysis*, 2016, ISSN: 1361-8415, Vol: 33, Page: 155-158

[182] C. Blaiotta, P. Freund, M. J. Cardoso, and J. Ashburner, "Generative diffeomorphic modelling of large MRI data sets for probabilistic template construction", In *NeuroImage*, 2018, 166, pp.117-134.

[183] Z. Xu, C. P. Lee, M. P. Heinrich, M. Modat, D. Rueckert, S. Ourselin, R. G. Abramson, B. A. Landman "Evaluation of Six Registration Methods for the Human Abdomen on Clinically Acquired CT", in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1563-1572, Aug. 2016.

[184] F. X. Vialard, L. Risser, D. Rueckert, C. J. Cotter, "Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation", In *International Journal of Computer Vision*, 2012, 97(2), 229–241.

[185] J. Wu and X. Tang, "A Large Deformation Diffeomorphic Framework for Fast Brain Image Registration via Parallel Computing and Optimization", In *Neuroinformatics*, 18, 251–266 (2020).

[186] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces", In *Medical image analysis*, 2019, 57, pp.226-236.

[187] A. Hering, L. Hansen, T. C. Mok, A. Chung, H. Siebert, S. Häger, A. Lange, S. Kuckertz, S. Heldmann, W. Shao, and S. Vesal, "Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning", 2021, arXiv preprint arXiv:2112.04489.

[188] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units", In *Computer methods and programs in biomedicine*, 2010, 98(3), pp.278-284.

[189] S. C. Joshi and M. I. Miller, "Landmark matching via large deformation diffeomorphisms", In *IEEE Transactions on Image Processing*, vol. 9, no. 8, pp. 1357-1370, Aug. 2000.

[190] M. I. Miller, A. Trouvé, and L. Younes, "Geodesic Shooting for Computational Anatomy", In *Journal of Mathematical Imaging and Vision* 24, 209–228 (2006).

[191] B. B. Avants, N. Tustison, and G. Song, "Advanced normalization tools (ANTS)", 2009, Insight j, 2(365), pp.1-35.

[192] J. V. Manjón, S. Eskildsen, P. Coupé, J. Romero, L. Collins, and M. Robles, "NICE: Nonlocal intracranial cavity extraction", In *International journal of biomedical imaging*, 2014, 820205.

[193] J. V. Manjón, and P. Coupé, "volBrain: an online MRI brain volumetry system", In *Frontiers in neuroinformatics*, 2016, 10, p.30.

[194] G. Grabner, A. L. Janke, M. M. Budge, D. Smith, J. Pruessner, and D. L. Collins, "Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults", In *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, vol. 9, pp. 58–66, 2006.

[195] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable",2011, In *IEEE transactions on medical imaging*, 31(2), pp.153-163.

[196] B. Fischl, "Freesurfer," In *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[197] B. Patenaude, S. M. Smith, S. D. Kennedy, and M. Jenkinson,"A Bayesian model of shape and appearance for subcortical brain segmentation." 2011, In *Neuroimage*, 56, 907–922. doi: 10.1016/j.neuroimage.2011.02.046

[198] N. Masoumi, H. Rivaz, M.O. Ahmad, and Y. Xiao, "DLCR: Deep Learning-based deformable image registration using Correlation Ratio", In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 2023. Lecture Notes in Computer Science, Springer, Cham.

[199] T. Lindig, R. Kotikalapudi, D. Schweikardt, P. Martin, F. Bender, U. Klose, U. Ernemann, N.K. Focke, and B. Bender, 2018. Evaluation of multimodal segmentation based on 3D T1-, T2-and FLAIR-weighted images–the difficulty of choosing. Neuroimage, 170, pp.210-221.

[200] H. Zhang, C. Lai, R. Liu, T. Liu, W. Niu, K. Oishi, Y. Zhang, and D. Wu, 2019. Age-specific optimization of T1-weighted brain MRI throughout infancy. Neuroimage, 199, pp.387-395.

[201] M. Goubran, S. de Ribaupierre, R.R. Hammond, C. Currie, J.G. Burneo, A.G. Parrent, T.M. Peters, and A.R. Khan, 2015. Registration of in-vivo to ex-vivo MRI of surgically resected specimens: a pipeline for histology to in-vivo registration. Journal of neuroscience methods, 241, pp.53-65.

[202] Y. Zhang, E. Paulson, S. Lim, W.A. Hall, E. Ahunbay, N.J. Mickevicius, M.W. Straza, B. Erickson, and X.A. Li, 2020. A Patient-Specific Autosegmentation Strategy Using Multi-Input Deformable Image Registration for Magnetic Resonance Imaging–Guided Online Adaptive Radiation Therapy: A Feasibility Study. Advances in radiation oncology, 5(6), pp.1350-1358.

[203] S. Ghosal, and N. Ray, 2017. Deep deformable registration: enhancing accuracy by fully convolutional neural net. Pattern Recognition Letters, 94, pp.81-86.

[204] M. Avi-Aharon, A. Arbelle, and T. R. Raviv, "Deephist: Differentiable joint and color histogram layers for image-to-image translation", 2020. In *arXiv preprint arXiv:2005.03995*.

[205] A. Papoulis. Probability, Random Variables, and Stochastic Processes. McGrawHill, Inc., third edition, 1991.

[206] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, 2019. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.

[207] P. Dupuis, U. Grenander, and M.I. Miller, 1998. Variational problems on flows of diffeomorphisms for image matching. Quarterly of applied mathematics, pp.587-600.

[208] D.L. Collins *et al.*, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," J. Comput. Assist. Tomogr. 18(2), 192–205 (1994).

[209] A.J. Worth, N. Makris, D.N. Kennedy, and V.S. Caviness Jr, (2001). Accountability in methodology and analysis for clinical trials involving quantitative measurements of MR brain images, Technical Report TR20011117, Neuromorphometrics, Inc. (http://neuromorphometrics.com/TR20011117.pdf)

[210] V.S. Caviness Jr, N.T. Lange, N. Makris, M.R. Herbert, D.N. Kennedy: MRI based brain volumetrics: emergence of a developmental brain science.Brain Dev 1999, 21:289-295.

[211] D.P. Kingma and J. Ba. (2014). "ADAM: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980