Image segmentation and registration using machine learning for brain shift compensation in image-guided neurosurgery

François-Xavier Carton

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Informatique

Arrêté ministériel : 25 mai 2016

Présentée par

François-Xavier CARTON

Thèse dirigée par **Matthieu CHABANAS**, Univ. Grenoble Alpes
et codirigée par **Jack NOBLE**, Vanderbilt University, USA

préparée au sein du **Laboratoire Translational Innovation in Medicine and Complexity (TIMC)**
dans **l'École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

Segmentation et recalage d'images avec machine learning pour la compensation du brain shift en neurochirurgie guidée par image

Image segmentation and registration using machine learning for brain shift compensation in image-guided neurosurgery

Thèse soutenue publiquement le **13 décembre 2021**,
devant le jury composé de :

Monsieur Antoine SIMON
Maître de conférences, UNIVERSITÉ RENNES 1, Rapporteur
Monsieur Marc MODAT
Senior lecturer, KING'S COLLEGE LONDON, UK, Rapporteur
Monsieur Michel DESVIGNES
Professeur des Universités, GRENOBLE INP, Examinateur
Madame Jocelyne TROCCAZ
Directrice de recherche, CNRS, UNIVERSITÉ GRENOBLE ALPES, Présidente
Monsieur Thomas LANGO
Research director, SINTEF, Norway, Invité

Abstract

In brain tumor resection surgery, a preoperative magnetic resonance (MR) image is used for surgical planning and guidance. Brain tissue displacement during surgery, known as brain shift, severely limits image guidance with the preoperative MR (pMR), which no longer reflects the location of tissue and anatomical structures accurately. This can be mitigated by registering the pMR with intraoperative data, ie. updating the pMR so that it matches images acquired during surgery. Ultrasound (US) imaging has several benefits compared to other intraoperative imaging modalities such as MR. In particular, it has fast acquisition times and do not require moving the patients, which makes the surgery longer. Thus, there is growing interest in using US imaging in brain surgery, despite the low image quality.

While traditional registration methods have been proposed, deep learning (DL) are being actively researched as they have proven successful and achieve state-of-the-art results in many domains such as medical image segmentation, and more recently, image registration. DL models are complex models with parameters learned with data, with many applications. However, there are few studies proposing DL models for segmentation of intraoperative US (iUS) images of the brain, or for registration of the pMR image with iUS images.

We start with the segmentation of iUS images of the brain. This is can be used for several purposes, such as guiding a registration model, which is the main objective of this thesis. We first focus on the segmentation of the resection cavity, which is essential to model the tissue displacement due to tissue removal. Next, we address tumor tissue segmentation, which is more challenging but can be used not only for registration but also to determine if tissue resection is complete. The results show that DL models can successfully segment the resection cavity and tumor tissue in US images. Then, we train a multi-class segmentation model, that segments several structures at the same time. The results suggest that multi-class models can be more accurate than single-class models, and leverage inter-class relationships.

We then investigate DL models for registration of pMR with iUS images. Image registration is inherently difficult, and especially with different modalities such as MR and US. In addition, there is a high variability across cases, as different structures are visible in the limited US field of view, depending on tumor location. It is also difficult to estimate ground truth displacements for training DL models. While we were not able to train a registration model with the available data, we discuss our findings and possible areas of improvement.

DL models can successfully segment iUS images. We currently lack data to train registration models but expect that larger datasets will enable the training of such models.

Keywords: Brain-shift, image-guided neurosurgery, image segmentation, image registra-

tion, deep learning.

Résumé

Dans le cadre des chirurgies d'ablation de tumeurs cérébrales, un examen préopératoire par imagerie par résonance magnétique (IRM) est réalisé pour la planification chirurgicale et le guidage lors de la chirurgie. Le déplacement des tissus cérébraux pendant la chirurgie, appelé communément *brain shift*, limite considérablement le guidage par image utilisant l'IRM préopératoire (pMR), celui-ci ne représentant plus la position courante des tissus. Ces limitations peuvent être compensées en recalant le pMR avec des images peropératoires. L'échographie présente plusieurs avantages à d'autres modalités comme l'IRM, notamment sa rapidité d'acquisition, et la non-nécessité de déplacer les patients. L'échographie a donc un réel intérêt en neurochirurgie, malgré la moindre qualité de ces images.

Bien que des méthodes traditionnelles de recalage aient été proposées, l'apprentissage profond (AP) fait l'objet de nombreux travaux de recherche car il obtient les meilleurs résultats dans plusieurs domaines comme le traitement d'images médicales, notamment la segmentation et plus récemment, le recalage d'image. Les modèles d'AP sont des modèles complexes dont les paramètres sont appris à partir de données. Cependant, il y a peu d'études sur leur applicabilité à la segmentation d'images échographiques, ou au recalage avec un pMR.

L'étude porte tout d'abord sur la segmentation des images échographiques peropératoires (iUS) du cerveau. Ces segmentations peuvent notamment servir au recalage d'image, l'objectif principal de cette thèse. Nos travaux portent premièrement sur la segmentation des cavités de la résection, permettant de modéliser précisément le décalage des tissus lié à l'ablation de tissus. Nous proposons ensuite un modèle pour la segmentation des tumeurs, problème plus difficile mais utile non seulement pour le recalage d'image mais aussi pour déterminer si la résection est complète. Les résultats obtenus montrent que l'AP peut être utilisé pour la segmentation d'iUS. Nous avons ensuite entraîné un modèle multi-classes, segmentant directement plusieurs structures. Les résultats suggèrent que ces modèles multi-classes peuvent produire de meilleures segmentations, tenant compte des dépendances entre les classes.

Puis nous étudions les modèles de recalage des pMRs avec des iUSs. Ce problème est considérablement plus compliqué que la segmentation d'image, en particulier entre différentes modalités comme l'IRM et l'échographie. Il y a de plus une grande variabilité entre patients, les structures visibles dans le champ de vue limité des iUSs dépendant de la position de la tumeur. Il est aussi difficile de déterminer les vrais déplacements de tissus avec lesquels entraîner des modèles. Bien que n'étant pas parvenus à entraîner un modèle avec les données disponibles, nous analysons les résultats et identifions de potentielles pistes d'améliorations.

Les modèles d'AP peuvent segmenter les iUS. Les données sont actuellement insuffisantes

pour des modèles de recalage, que plus de données permettraient d'entraîner.

Mots-clés : Brain-shift, neurochirurgie guidée par image, segmentation d'image, recalage d'image, apprentissage profond.

Acknowledgements

I would like to first thank my advisors Matthieu Chabanas and Jack Noble for their continued support and guidance during these three years. Thank you for all the advice, support, and availability (even when I was abroad). Et merci Matthieu pour tous les moments partagés, dans le cadre mais aussi en dehors de la thèse. Et pour m'avoir aidé dans les segmentations, tu sais à quel point ça compte. Je sais que j'aurais quelqu'un vers qui me tourner la prochaine fois que j'apercevrai un *sulcus* ou une faux dans la nature !

Next, I would like to thank all the committee members: Antoine Simon, Marc Modat, Michel Desvignes, Jocelyne Troccaz, and Thomas Lango. In particular, thank you Antoine Simon and Marc Modat for reviewing the manuscript thoroughly and providing helpful comments. And thank you Michel Desvignes for agreeing to join the committee given the very short notice.

Je voudrais ensuite remercier Florian Le Lann, pour ses précieux conseils sur les images échographiques, ainsi que pour avoir pris le temps de vérifier et commenter les segmentations de la cavité de la resection. Merci également de m'avoir permis d'assister à une opération au bloc, ce qui m'a permis d'en apprendre beaucoup et de voir les contraintes pratiques pendant les chirurgies.

I would also like to thank Ingerid Reinertsen for the help, comments, and support. And thanks to Bodil Munkvold, Bahareh Behboodi, and Ingerid Reinertsen for sharing the tumor segmentations. I know segmenting ultrasound images is no fun, but especially delineating tumors is really difficult, so thank you very much for sharing them.

Many thanks to everyone I met at Vanderbilt University. Thank you Benoit Dawant for enabling me to come to Vanderbilt University as a visiting student. Special thanks to Bill for all the help and support! And thanks to all the students, for all we shared together. In no particular order, thanks to: Srijata, Xiaochen, Jianing, Ahmet, Yiyuan, Dongqing, Rueben, Erin, Ziteng, Mohammad, Ghassan, Yubo, Haley, Roza, Colin, Ben, Sarah, Nhung, Praitayini, Cailey. Thank you Jianing for the discussions on registration, and for sharing code and singularity images with me. Thank you Rueben for the serious and non-serious discussions, for being on the Linux/LaTeX side in this adverse environment, and for your TPS code. Thank you Srijata for supporting my rants, all the driving, and for the DnD "plans".

Merci à tous ceux que j'ai rencontrés au laboratoire TIMC. D'abord merci aux permanents de l'équipe GMCAO, Sandrine, Joce, Yohan, Céline, Pierre-Alain, Taha, Mahnu. Et bien sûr merci aux doctorants, post-docs, et intrus : Pablo, Matthias, Sylvain, Tamara, Clément, Maxime, Nolwenn, Nicolas, Jean-Loup, et Guillaume. Pour tous ces déjeuners, pauses café,

discussions, avis plus ou moins tranchés... merci !

Et enfin, un grand merci à ma famille pour tout le soutien.

Publications

Journal papers

F.-X. Carton, M. Chabanas, F. Le Lann, and J. H. Noble. **Automatic segmentation of brain tumor resections in intraoperative ultrasound images using U-Net.** In *Journal of Medical Imaging*, vol. 7, no. 3, pp. 1–15 (2020). DOI: 10.1117/1.JMI.7.3.031503

Conference papers

F.-X. Carton, J. H. Noble, and M. Chabanas. **Automatic segmentation of brain tumor resections in intraoperative ultrasound images.** In *Proceedings of SPIE*, vol. 10951 (2019). DOI: 10.1117/12.2513613

F.-X. Carton, J. H. Noble, B. K. R. Munkvold, I. Reinertsen, and M. Chabanas. **Automatic segmentation of intraoperative ultrasound images of the brain using U-Net.** In *Surgetica 2019 - Computer-Assisted Medical Interventions: scientific problems, tools and clinical applications*, pp. 66–68 (2019)

F.-X. Carton, M. Chabanas, B. K. R. Munkvold, I. Reinertsen, and J. H. Noble. **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 (2020). DOI: 10.1117/12.2549516

F.-X. Carton, J. H. Noble, F. Le Lann, B. K. R. Munkvold, I. Reinertsen, and M. Chabanas. **Multiclass segmentation of brain intraoperative ultrasound images with limited data.** In *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11598, pp. 149–154 (2021). DOI: 10.1117/12.2581861

M. Donnez, F.-X. Carton, F. Le Lann, E. De Schlichting, and M. Chabanas. **Realistic synthesis of brain tumor resection ultrasound images with a generative adversarial network.** In *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11598, pp. 637–642 (2021). DOI: 10.1117/12.2581911

Résumé détaillé

Introduction

Lors des chirurgies d'ablation de tumeurs cérébrales, un système de navigation permet le guidage en utilisant l'IRM préopératoire. Cela permet notamment d'afficher la position d'outils repérés à l'aide de marqueurs sur l'IRM. Cependant, l'IRM préopératoire ne représente pas la position courante des tissus cérébraux, à cause du déplacement de ces tissus pendant la chirurgie. Une solution pour compenser ce problème est de recaler l'IRM préopératoire avec des images peropératoires. Cela consiste à mettre à jour l'IRM préopératoire pour que celui-ci corresponde à des images acquises lors de la chirurgie.

L'échographie est une modalité d'imagerie qui possède plusieurs avantages sur d'autres modalités comme l'IRM. En particulier, l'acquisition d'images échographiques est rapide et ne nécessite pas de déplacer les patients, ce qui rallongerait l'opération chirurgicale. C'est pourquoi l'échographie est parfois utilisée en neurochirurgie et fait l'objet d'études, malgré la moindre qualité de ces images.

Les modèles d'apprentissage profond sont des modèles complexes entraînés avec des données, qui peuvent être utilisés notamment pour le traitement d'image. Bien que ces modèles soient l'objet de nombreux travaux de recherche, et en particulier pour le traitement d'images médicales, il n'existe peu d'études sur l'applicabilité de ces méthodes aux images échographiques, pour leur segmentation ou recalage avec un pMR.

Les travaux présentés dans cette thèse portent donc sur la segmentation d'images échographiques du cerveau, et leur recalage avec l'IRM préopératoire, avec des modèles d'apprentissage profond. Nous présentons d'abord des modèles de segmentations, pour la cavité de la résection et pour les tumeurs, puis des modèles multi-classes, et enfin des modèles de recalage.

Organisation du manuscrit

Le **chapitre 1** présente le contexte clinique, les motivations et les objectifs de ces travaux de thèse. Le chapitre commence par une présentation de l'anatomie du cerveau et des tumeurs

cérébrales. Il explique notamment comment celles-ci sont diagnostiquées et traitées. Puis, la chirurgie d'ablation des tumeurs cérébrales est détaillée en particulier, et les systèmes de guidage par image sont présentés. Enfin, les causes et conséquences des déplacements des tissus sont expliqués, ce qui permet d'exposer la problématique et les objectifs de cette thèse.

Le **chapitre 2** propose une revue de la littérature et pose les bases techniques des méthodes utilisées dans cette thèse. Le chapitre commence par une revue des méthodes de compensation du *brain shift*, et plus particulièrement les méthodes utilisant l'échographie peropératoire. Les différentes méthodes de segmentation d'IRM et images échographiques du cerveau sont ensuite synthétisées, par type de méthode. Cette revue générale est suivie par une revue plus détaillée des méthodes d'apprentissage profond, en commençant par les bases techniques puis en présentant les méthodes de segmentation et recalage. La fin du chapitre est consacrée à la présentation des deux bases de données publiques d'images échographiques du cerveau. Nous avons utilisé ces deux bases de données pour les travaux de cette thèse, pour l'entraînement et évaluation des modèles de segmentation et recalage.

Le **chapitre 3** présente les méthodes et résultats des modèles de segmentation de la cavité de la résection dans les images échographiques du cerveau. La méthodologie y est détaillée, avec en particulier l'architecture des réseaux de neurones artificiels, ainsi que les processus de pré-traitement, échantillonage et post-traitement. Nous comparons différentes variations de la méthode, avec plusieurs méthodes d'échantillonage et des architectures 2D et 3D. Cette étude est basée sur deux bases de données acquises sur des sites différents, ce qui permet d'évaluer la généralisabilité des modèles à des données acquises sur d'autres sites. Le chapitre est basé sur le papier que nous avons publié dans le *Journal of Medical Imaging* [1], suite à des résultats préliminaires présentés à la conférence *SPIE 2019 medical imaging* [2].

Le **chapitre 4** évalue et étend le modèle du chapitre 3 à la segmentation de tissus cancéreux. Cette tâche étant plus difficile, nous proposons d'ajouter l'IRM préopératoire, et éventuellement sa segmentation aux entrées du modèle, pour déterminer si le modèle peut bénéficier de cette information des images préopératoires pour l'estimation de la segmentation peropératoire. Des résultats préliminaires pour cette étude, qui ne testait pas l'ajout des informations préopératoires, avaient été présentés à la conférence *SPIE 2020 medical imaging* [4].

Le **chapitre 5** propose des modèles de segmentation multi-classes, c'est-à-dire des modèles capable de segmenter plusieurs structures directement. Cette étude est motivée par de précédentes études suggérant que les modèles de segmentation multi-classes peuvent tenir compte des relations spatiales entre les classes et donc obtenir de meilleurs résultats. Par conséquent, nous avons implémenté des modèles multi-classes et les avons comparés aux modèles mono-classe. Dans ce chapitre, nous présentons également une méthode de pondération des classes des échantillons, qui permet d'entraîner les modèles malgré l'important déséquilibre des classes, avec en particulier des classes complètement absentes de certains volumes. De plus, nous évaluons l'effet d'un terme de distance de Hausdorff dans la fonction d'ob-

jectif, dont l'implémentation a récemment été proposée. Des résultats préliminaires ont été présentés à la conférence *SPIE 2021 medical imaging* [5]. Ce chapitre présente une analyse plus détaillée des résultats, avec en particulier des résultats quantitatifs pour la segmentation des *sulci*. Cette analyse n'avait pas pu être faite pour la conférence, la vérité terrain des échantillons de test n'ayant pas encore été manuellement segmentée.

Le **chapitre 6** présente un modèle de recalage de l'IRM préopératoire avec les images échographiques peropératoires. Ce problème est extrêmement compliqué à cause du nombre limité de volumes disponibles, de l'importante variabilité inter-patient, et du contexte multimodal. Bien que nous ne soyons pas parvenus à entraîner un modèle de recalage satisfaisant, nous présentons une analyse des méthodes et de leurs limites et nous proposons des pistes d'améliorations.

Le **chapitre 7** conclut ces études avec un résumé des méthodes présentées et de leurs résultats. Puis, nous présentons les travaux en cours et futurs, qui constituent la suite des travaux présentés ici. Cette conclusion est reprise en français ci-après.

Conclusion

Nous avons présenté des modèles d'apprentissage profond pour la segmentation d'images échographiques du cerveau et le recalage de l'IRM préopératoire avec les images échographiques peropératoires. Nous avons d'abord proposé des modèles mono-classe pour la segmentation de la cavité de la résection et pour les tissus cancéreux. Puis nous avons adapté ces modèles à la segmentation multi-classe, avec plusieurs classes dont les *sulci* et la faux du cerveau (*falx cerebri*). Enfin, nous avons étudié les modèles de recalage de l'IRM préopératoire avec les images échographiques peropératoires.

Les résultats montrent que les modèles mono-classe peuvent segmenter la cavité de la résection et les tumeurs dans les images échographiques. Les modèles 3D sont plus précis que les modèles 2D, mais ont des temps d'exécution plus long dus à la nécessité d'évaluer plus d'échantillons pour reconstruire un volume complet. Les résultats des modèles multi-classes suggèrent que ces modèles peuvent obtenir de meilleurs résultats que les modèles mono-classe. Les modèles multi-classes que nous avons entraînés sont capables de segmenter la cavité de la résection, la tumeur et les *sulci*. En revanche, nous n'avions pas assez de volumes où la faux du cerveau ou les ventricules étaient visibles pour permettre d'obtenir des modèles de segmentation fiables pour ces deux classes.

Nous avons entraîné des modèles de recalage de l'IRM préopératoire avec les images échographiques peropératoires, mais nous ne sommes pas parvenus à obtenir de modèle satisfaisant. Nous pensons que cela est principalement dû au faible nombre de données sur lesquelles entraîner les modèles, ainsi qu'au manque d'une véritable vérité terrain, que nous avons es-

timée à l'aide d'un nombre très limité de points correspondants. De plus, c'est un problème de recalage multimodal, avec une grande variabilité des données d'entrée. La construction de bases de données plus importantes sera essentielle au développement de méthodes d'apprentissage profond de recalage d'IRM peropératoire avec des images échographiques.

Pour les travaux sur la segmentation et le recalage présentés dans cette thèse, nous avons manuellement segmenté les volumes. Les segmentations de la cavité de la résection ont été vérifiées par un neurochirurgien, et nous sommes en train de valider les segmentations des autres classes. Nous publierons ces segmentations une fois validées, afin que d'autres études puissent en bénéficier.

Une limitation importante des modèles présentés ici est le faible nombre de cas des bases de données. Nous avons principalement utilisé les images des 23 patients de la base de données RESECT, qui est la plus récente base de données publique d'images échographiques du cerveau. Nous avons aussi évalué les modèles de segmentation de la cavité de la résection avec la base de données BITE, une base de données antérieure de 14 patients. Ces nombres de patients sont considérablement inférieurs aux nombres usuels dans les études d'apprentissage profond pour le traitement d'images médicales, qui sont généralement de l'ordre de la centaine voire milliers de patients. Le faible nombre de volumes utilisés ici limite non seulement l'apprentissage mais aussi l'évaluation des modèles. Un nombre plus important de volumes permettrait un meilleur entraînement des modèles et une meilleure évaluation de le généralisabilité aux nouveaux volumes.

Perspectives

Validation des segmentations manuelles Les segmentations manuelles créées pour les besoins de cette thèse constituent des données utiles qui peuvent profiter à d'autres études. Nous sommes donc en train de valider ces segmentations dans le but de les publier.

Segmentation multi-classes Bien que la faux du cerveau soit utile pour le recalage d'image, il serait intéressant de comparer les modèles présentés dans cette thèse qui considèrent les *sulci* et la faux du cerveau comme deux classes distinctes, avec des modèles pour lesquels ces deux classes sont fusionnées.

Il serait également pertinent d'entraîner des modèles multi-classes 3D, les cartes graphiques récentes disposant de plus de mémoire vidéo et permettant donc d'utiliser des volumes plus grands. En effet, les modèles 3D seraient potentiellement capables de segmenter la faux du cerveau, ce qui est bien plus difficile sur une seule coupe 2D.

Modèle de recalage Plusieurs pistes peuvent être envisagées pour le modèle de recalage. Il est difficile d'établir une vérité terrain pour la déformation, et l'estimation avec les *thin plate splines* peut être améliorée. Par exemple, il serait possible d'utiliser les segmentations pour obtenir une déformation qui fasse correspondre non seulement les correspondances de points de RESECT mais aussi les segmentations. Une autre piste est de changer le mode de représentation de la transformation en sortie du modèle de recalage. Une transformation affine ou une grille dense de déformation moins détaillée que celle utilisée pourrait être utilisée soit en sortie du modèle, soit dans une première étape de recalage.

Base de données plus grandes Des bases de données plus grandes permettraient aux modèles de gagner en robustesse et d'être mieux évalués. Cela requiert l'acquisition de données, puisque les données publiques sont déjà utilisée dans les travaux de cette thèse. C'est donc une perspective à plus long terme, car acquérir de telles données est un long processus.

Organization of the dissertation

Chapter 1 introduces the clinical background, and presents the motivations and goals of this thesis. First, an overview of brain anatomy and brain tumors is given, including the diagnosis and treatment of brain tumors. Next, we focus on brain tumor resection surgery, and present the surgical procedure and image-based navigation. Finally, we discuss the causes and consequences of brain shift, which is the displacement of brain tissue during surgery. We state the objectives of this thesis and summarize the contributions.

Chapter 2 reviews related works and presents technical background for the methods used in this thesis. We start with reviewing brain shift compensation methods, with a focus on methods using intraoperative ultrasound. Then we give an overview of segmentation methods for brain magnetic resonance (MR) and ultrasound images. After this general review, we focus on deep learning methods, starting with general background and followed by its applications to image segmentation and registration. Finally, we introduce the two publicly available datasets of brain ultrasound images. These two datasets are used in this thesis for training and testing the segmentation and registration models.

Chapter 3 presents the methods we proposed and results we obtained for the segmentation of the resection cavity in brain ultrasound images. The methodology is detailed, including the model architecture as well as the pre-processing, sampling, and post-processing steps. We compare several variations of the sampling methods and several model architectures with 2D and 3D models. This study is based on two datasets acquired on different sites, which allows the evaluation of the model generalizability to data acquired in different sites. This chapter is based on the paper we published in the Journal of Medical Imaging [1], for which preliminary results were presented at the SPIE 2019 medical imaging conference [2].

Chapter 4 evaluates and extends the segmentation model to tumor tissue segmentation. As this task is more challenging, we propose to add preoperative images and optionally segmentations in the input, to assess whether this preoperative information can be leveraged by this model to find the intraoperative segmentations. Preliminary results that did not include this addition of the preoperative data were presented at the SPIE 2020 medical imaging conference [4].

Chapter 5 presents multi-class segmentation models, ie. models capable of segmenting several structures at once. This is motivated by other studies that suggested that multi-class segmentation networks can leverage spatial relationships between classes and obtain more accurate results. Thus, we implement multi-class models and compare the results with those of single-class models. In this chapter, we also present a weighting strategy that allows training the models with high class imbalance, and in particular with classes completely missing in some volumes. In addition, we evaluate the impact of a Hausdorff distance term in the loss function, whose implementation was recently proposed. Preliminary results were presented at the SPIE 2021 medical imaging conference [5]. In this chapter we present a more detailed analysis of the results. In particular, we show quantitative results for the *sulci* class, for which ground truth was not available for the preliminary report.

Chapter 6 proposes a registration model for preoperative MR and intraoperative US. This task is extremely challenging due to the limited data available, high inter-patient variability, and multimodal setting. While we were not able to train a successful registration model, we discuss the methods we have tried and the limitations of the proposed method and propose possible ways to mitigate these issues.

Chapter 7 concludes this study with a summary of the methods and results that were presented in this thesis. Last, we discuss works that are currently in progress and future works.

Table of contents

Chapter 1

Clinical context and motivations

The work presented in this thesis aims to improve the guidance system in the context of brain tumor resection surgery. This chapter provides the clinical context for this work.

We start by giving an overview of brain anatomy (section 1.1), which presents the anatomical structures that we refer to in the manuscript. Next, we describe brain tumors (section 1.2) and brain tumor resection surgery (section 1.3), which is the clinical context of this work.

The main challenge of tumor resection surgery is to locate the tumor tissue, and determine the extent of resection. Image navigation systems are used for guidance (section 1.4). Section 1.5 presents the imaging techniques that can be used, and in particular the ones that can be used during surgery.

Last, we describe brain shift, which is the displacement of tissue during surgery. Because of this displacement, the preoperative MRI (acquired before surgery) does not accurately match the tissue and anatomical structures in the operating room. This can be mitigated by updating the preoperative MRI using intraoperative imaging. This process of finding the transform between two images so that one can be mapped into the other is called registration. This thesis explores structure-based registration using machine learning models, for brain shift compensation. This choice is detailed in section 1.7.

1.1 Brain anatomy

The nervous system is a biological system of an animal that allows communication between different parts of its body and coordinates its actions. The nervous system is divided into two parts: the central nervous system, consisting of the brain and spinal cord, and the peripheral nervous system, which consists in nerves and ganglia outside the central nervous system. The peripheral nervous system connects the brain to end organs, allowing the transmission of signals from the brain to organs, and from senses to the brain.

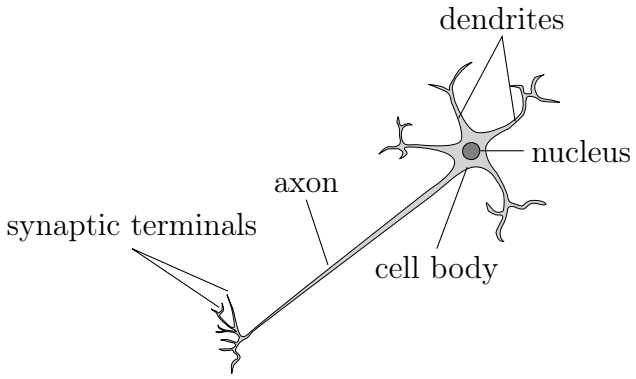


Figure 1.1 – Schematic of a neuron.

Signals are transmitted within the nervous system through cells called neurons. Neurons are composed of a cell body, an axon through which signals are transmitted, and dendrites through which signals are received (see figure 1.1). Neurons are connected to other neurons or cells at the end of their axons.

The cerebrum is the uppermost and largest part of the brain. It consists of two cerebral hemispheres (left and right) and other subcortical structures. The two cerebral hemispheres are connected by the *corpus callosum*, a large set of nerves that allow communication between the two hemispheres. Above the *corpus callosum* lies the *falx cerebri*, a vertical membrane located between the two cerebral hemispheres. The *falx cerebri* is named as such because of its crescent-like shape, which looks like the blade of a *falx* (sickle). At the back of the brain (anterior part), the two cerebral hemispheres lie on a vertical membrane called *tentorium cerebelli*. This membrane separates the cerebrum (upper part) from the cerebellum (lower part). Figure 1.2 shows a three dimensional representation of the aforementioned structures, as well as a slice from the corresponding MRI scan. The three dimensional view is useful to understand where the different structures are located relative to each other.

The two cerebral hemispheres are made of neurons which form brain tissue. The outer layer is called gray matter (GM) and mainly consists of neuron cell bodies, whereas the inner part, called white matter (WM), is mainly composed of axons. Brain tissue is folded, ie. its surface has many folds. The ridges are called *gyrus* (plural *gyri*), whereas the grooves are called *sulcus* (plural *sulci*) (see figure 1.3).

The brain is surrounded by cerebrospinal fluid (CSF). CSF protects the brain both mechanically against hits and immunologically against infections. CSF is produced by the brain's ventricular system, which is also responsible for CSF flow within the brain. The ventricular system is mainly composed of four cavities called ventricles: two lateral ventricles, and the third and fourth ventricles. The lateral ventricles are connected to the third one by the interventricular *foramina* (singular *foramen*). The third and fourth ventricles are connected by the cerebral aqueduct. Figure 1.4 shows a three dimensional render of the

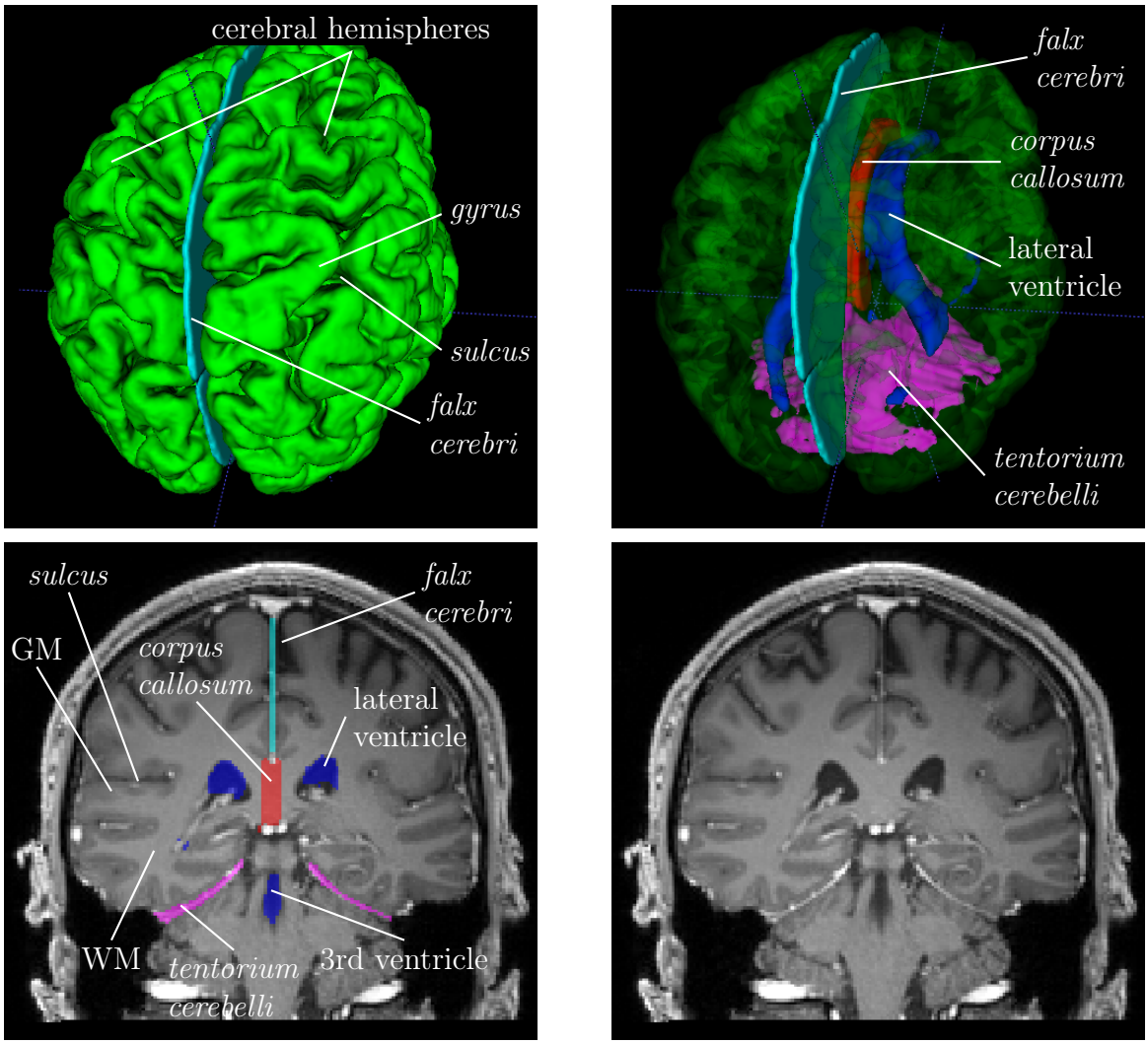


Figure 1.2 – Localization of the human brain's main structures (top: 3d reconstruction, bottom: coronal slice of a T1-weighted MRI scan).

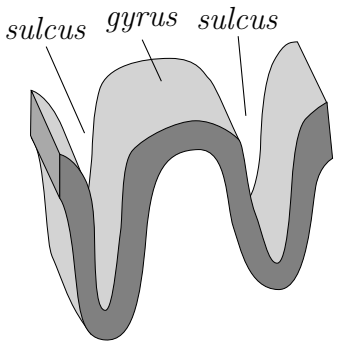
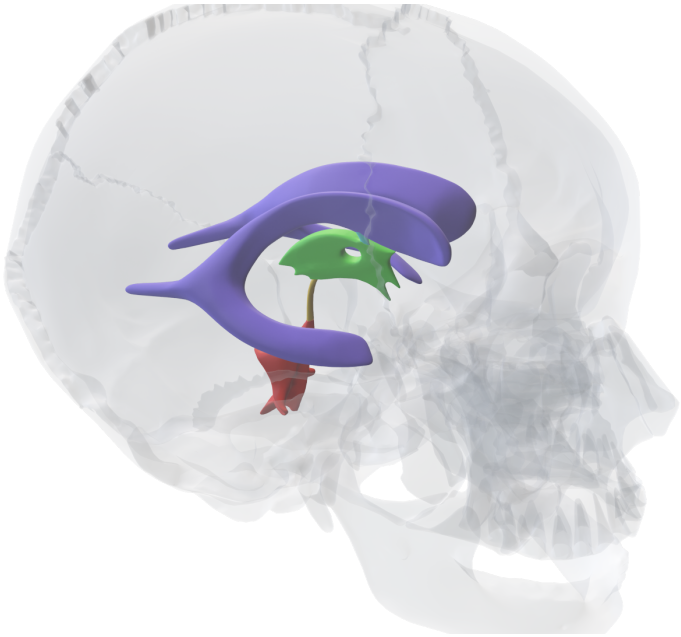


Figure 1.3 – *Sulci* and *gyri*.



Created with blender (https://www.blender.org), using meshes from BodyParts3D © The Database Center for Life Science, licensed under CC Attribution-Share Alike 2.1 Japan.

Figure 1.4 – Brain ventricular system: lateral ventricles (blue), third ventricle (green), fourth ventricle (red), interventricular *foramina* (cyan) and cerebral aqueduct (yellow).

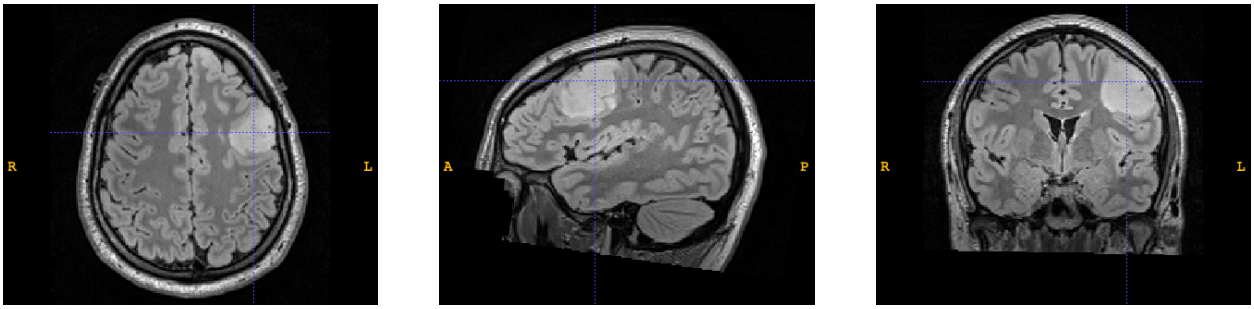
brain ventricular system.

1.2 Brain tumors

A tumor is an abnormal growth of tissue due to dysfunctional cells. Tumors can be benign or malignant. Benign tumors do not evolve into cancer, whereas malignant tumor (most commonly called cancer) destroy surrounding tissue and generally cause death.

Brain tumors can be classified by the type of affected cells. For example, a tumor that starts in the glial cells of the brain is called a glioma. Gliomas represent about 80% of all malignant brain tumors [7]. Other tumor types include meningiomas, pituitary adenomas, and nerve sheath tumors. Brain tumors are also given a grade which correspond to their growth speed [8], from I to IV, where IV is the highest speed. Grade I and II tumors are called low-grade tumors, while grade III and IV tumors are high-grade tumors.

There are many possible symptoms of brain tumor, depending on several factors such as the location and size of the tumor. The increased intracranial pressure due to tissue growth may cause headaches. Depending on the tumor location, additional troubles such as loss of vision or hearing, difficulty with movement or balance, behavioral issues, and troubles with language or memory.



Axial view | Sagittal view | Coronal view

Figure 1.5 – A FLAIR MRI, in which tumor tissue is in the brighter area indicated by the blue cross.

Brain tumors are usually diagnosed following symptoms that may be caused by brain tumors, if there is no sign that may indicate other causes. Suspicions of a brain tumor are then further investigated using medical imaging, with which tumor tissue can be detected and located. The standard imaging method used for this purpose is magnetic resonance imaging (MRI) [9, 10]. Figure 1.5 shows an MRI scan in which a tumor can be seen.

There are several possible treatments depending on the tumor type, grade, and location. The most common type of treatment are surgery, radiotherapy, and chemotherapy [10]. Surgical resection of tumor tissue is the preferred method when it is applicable. This procedure is a surgery during which a surgeon partially or completely removes tumor tissue, after a small opening in the skull is made (craniotomy). Radiotherapy is another treatment procedure which aims to destroy tumor cells (and only tumor cells) using radiations. This process is generally repeated several times, and may be used after surgery to eliminate residual tumor tissue. Chemotherapy is a procedure which uses drugs to kill tumor cells. It may be used in addition to other methods, or when surgery and/or radiotherapy is not possible.

The next section gives more details about resection surgery, as it is the clinical context of this thesis.

1.3 Brain tumor resection surgery

Brain tumor resection surgery aims to remove the maximum of tumor tissue that is safely removable. Tumor tissue may only be partially removed if it is located in critical areas that will cause severe troubles if removed (eloquent areas). Removal of tumor tissue has been shown to improve patient outcomes [11, 12].

Before surgery, an MRI is acquired to establish the operative plan. In particular, it is used to determine where to make the craniotomy. For tumors that are not located directly

on surface, a path must be chosen to access the tumor. Then, tumor tissue is removed with surgical tools. This is a challenging task, because tumor tissue may not be clearly distinguishable from healthy tissue, and there is usually no clear boundary. Care must also be taken not to damage critical structures such as blood vessels.

After surgery, another MRI is acquired to assess the extent of resection, and determine if there is residual tumor tissue and if further treatment is needed.

The main challenge of this process is to locate tumor tissue, and particularly to determine its boundaries, ie. where to stop resecting tissue. Image navigation provides visual guidance to surgeons to help with this difficult task. These systems are described in the next section.

1.4 Image navigation and intraoperative imaging

To help surgeons with localizing the tumor, a navigation setup is often used. In such setup, the preoperative MRI is displayed on a screen, to guide the surgeons. It is rigidly registered to the patient position in the operating room, thus the image displayed matches the patient.

To register the preoperative MRI, a patient-to-image mapping is computed using corresponding points in the MRI and on the patient. There are different methods to establish such correspondences, depending on the navigation system. A common method is to use fiducials, which are markers that are placed on the patient. These markers are visible both physically on the patient and in the MRI (see figure 1.6). The locations of the markers are then measured by pointing them sequentially. The transform between the two sets of points is then computed, allowing to rigidly register the MRI to the patient. There are alternative methods to fiducial registration, that do not require markers, such as surface-based and image-based methods. Surface-based methods use scanners to acquire 3D surfaces of the face and forehead, and find a mapping that aligns the surfaces. Image-based methods use intraoperative images to estimate the mapping, by aligning the preoperative and intraoperative images.

Once the MRI has been registered, the position of tracked tools and pointer can be shown on the MRI display. The position of the tools in the operating room is known thanks to the tracking system, and that position can be translated in MRI space using the previously computed transform. There are several types of tracking systems. Figure 1.7 shows a picture of a surgical setup with navigation, using a tracking system with rigid bodies (small optical markers) that are tracked with a stereo camera.

However, navigation is impaired by tissue displacement during surgery (called brain shift). This phenomenon is further described in section 1.6. Because of that displacement, the

photo from Unsgaard et al., 2002 [13]

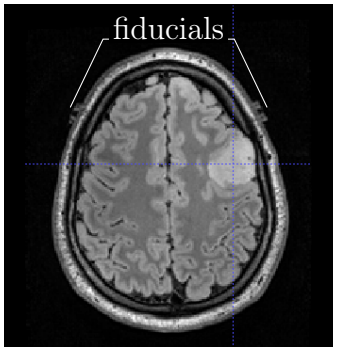


Figure 1.6 – Fiducial markers on a patient (left) and on an MRI (right).

previously established patient-to-image mapping is invalidated. Thus, the displayed image does not match the position of tissue in the operating room. As such, it is often complemented with intraoperative imaging: images acquired during surgery. The next section shows which imaging techniques can be used, and which ones are best suited for intraoperative imaging.

1.5 Imaging modalities

This section details medical imaging systems that can be used in brain tumor resection surgery. The last subsection focuses on the ones that can be used intraoperatively.

1.5.1 Magnetic resonance (MR)

MR is an imaging technique based on the magnetization of hydrogen atoms (which consist in a single proton). A strong uniform magnetic field is used to align the protons parallel to the direction of the magnetic field. The protons are said to be magnetized, which means that their spins are aligned with the same direction. A radiofrequency (RF) pulse is then applied to change the alignment, and bring the protons into phase with the RF pulse, thus with each other (phase coherence). After a certain amount of time, protons return in their equilibrium state (parallel to the external magnetic field). This return to equilibrium state is called relaxation, and can be decomposed into two processes: longitudinal relaxation and transverse relaxation. Longitudinal relaxation (or spin-lattice relaxation) refers to the recovery of the longitudinal component of magnetization, that is the component parallel to the external magnetic field. Transverse relaxation (or spin-spin relaxation) refers to the recovery of the component perpendicular to the external magnetic field, and corresponds to the loss of phase coherence. The two processes do not take the same amount of time, and are associated different time constants: T_1 for longitudinal relaxation and T_2 for transverse

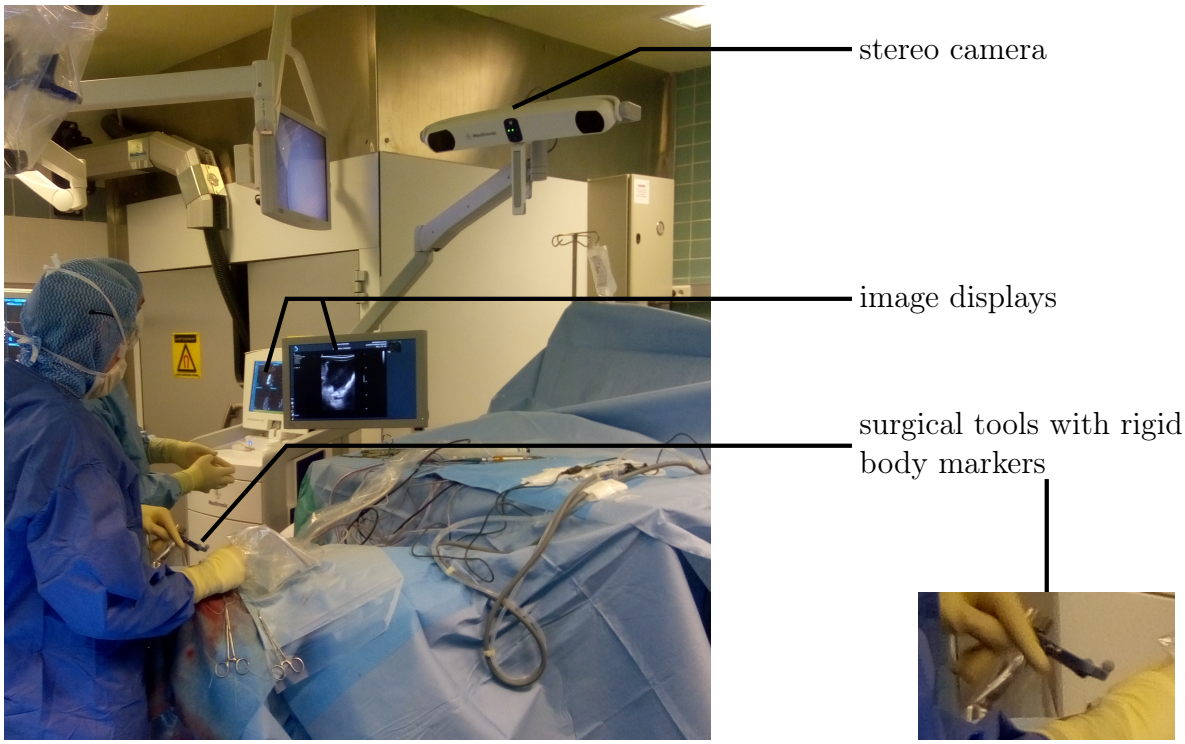


Figure 1.7 – Surgical setup with neuronavigation.

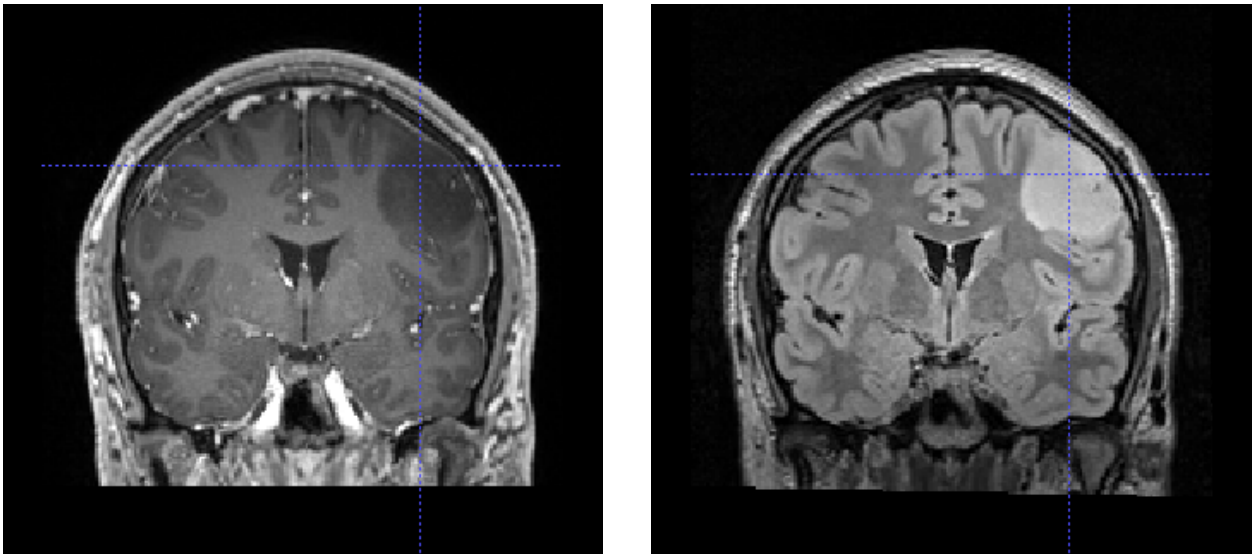
relaxation. For soft tissue, T_1 is around 1s while T_2 is on the order of 10ms.

With relaxation, protons emit RF energy, which is measured with receiver coils. The amplitude and duration of the emitted signal depends on several parameters, including T_1 and T_2. The measured signal is then converted to image intensities with Fourier transform.

The measure is made after a time T_E (echo time) after the RF pulse is applied. The process is repeated several times, with a time T_R (repetition time) between RF pulses. T_E and T_R are parameters of the MRI acquisition, and a particular set of values is called an MRI sequence. There are many MRI sequences, the most commonly used being T1-weighted, T2-weighted, and Fluid-attenuated inversion recovery (FLAIR). These sequences show different tissue characteristics, thus are complementary.

T1-weighted (T1w) MRI measures longitudinal relaxation, by using short T_R and T_E. In T1w MR images, content with more water appears darker. In particular, regions filled with CSF (such as ventricles and *sulci*) appears very dark. Gray matter appears gray, while white matter appears light gray. Tumor tissue is usually darker than healthy tissue (see figure 1.8a).

T2-weighted (T2w) MRI measures transverse relaxation, by using long T_R and T_E. In contrast to T1w, CSF appears very bright, while gray matter appears light gray and white



(a) T1w **(b)** FLAIR

Figure 1.8 – T1w and FLAIR MRI sequences (tumor tissue is pointed by the blue cross).

matter, gray.

FLAIR aims to suppress fluid signal by using even longer T_R and T_E. It looks similar to T2w, except that fluids such as CSF are very dark instead of very bright. Figure 1.8b shows an example of a FLAIR MR image.

MR angiography (MRA) is a specialized type of MR imaging which aims at imaging blood vessels. Different techniques and MRI sequences can be used for that purpose. A common method is to use a contrast agent that shortens the T_1 of blood, in combination with a short T_R. This will result in a very bright signal for blood, and thus blood vessels are visible in the resulting images.

1.5.2 Ultrasound (US)

US imaging is based on US waves, which are sound waves with frequencies higher than human hearing range (greater than 20kHz). An US probe is used to send US RF pulses into tissue, which are then echoed by the tissue with intensity and angle depending on tissue type. The echo is measured by the US probe, which converts the RF signal into current. The measured signal is then post-processed by the US hardware, including filtering and envelope detection.

The resulting signal can be used to create different images, the most common being the B-mode (for brightness) image, which represents the acoustic impedance of tissue. Doppler

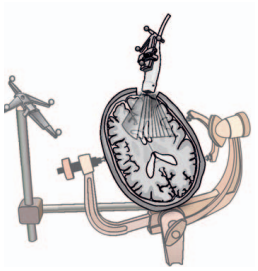


photo from Unsgård et al. [14]

(a) swiping motion

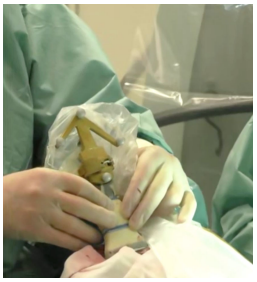


photo from Munkvold et al. [15]

(b) tracked ultrasound probe

Figure 1.9 – 3D US image reconstruction by acquering 2D slices in a swiping movement (left) with a tracked probe (right).

mode is another US mode that uses the Doppler effect to display the movement of tissues and fluids such as blood.

There are several type of US probes. Two-dimensional (2D) probes measure signal in a 2D region (slice), generating a 2D image, while three-dimensional (3D) probes measure signal in a 3D region. 2D probes can be used to acquire several slices with different angle, in a swiping motion, to acquire data in a 3D region that can be used to reconstruct a 3D image with interpolation (see figure 1.9a). This requires to track the probe (figure 1.9b), so that the position and orientation of each slice is known. Alternatively, 3D probes can acquire 3D volumes directly without reconstruction and interpolation.

In US images, healthy tissue appears dark gray, while tumor tissue usually appears a little brighter or darker depending on tumor type and grade. Interfaces between tissues and fluids such as CSF appear very bright, while fluids appear very dark. As such, ventricles are represented by dark cavities with a bright border, and *sulci* with bright lines, as they are thin for the most part. Figure 1.10 shows two examples of brain US images, with corresponding MRI for better visualization of the imaged structures. Note that the MRI has been registered with the US volume, so that the structures match.

1.5.3 Optical systems

Stereo vision cameras can be used to acquire images of the surface of the brain. With this data, the cortical surface can be reconstructed in three dimensions, for use in brain shift compensation methods [16]. Figure 1.11 shows an example of image of the cortical surface acquired with a stereo vision camera.

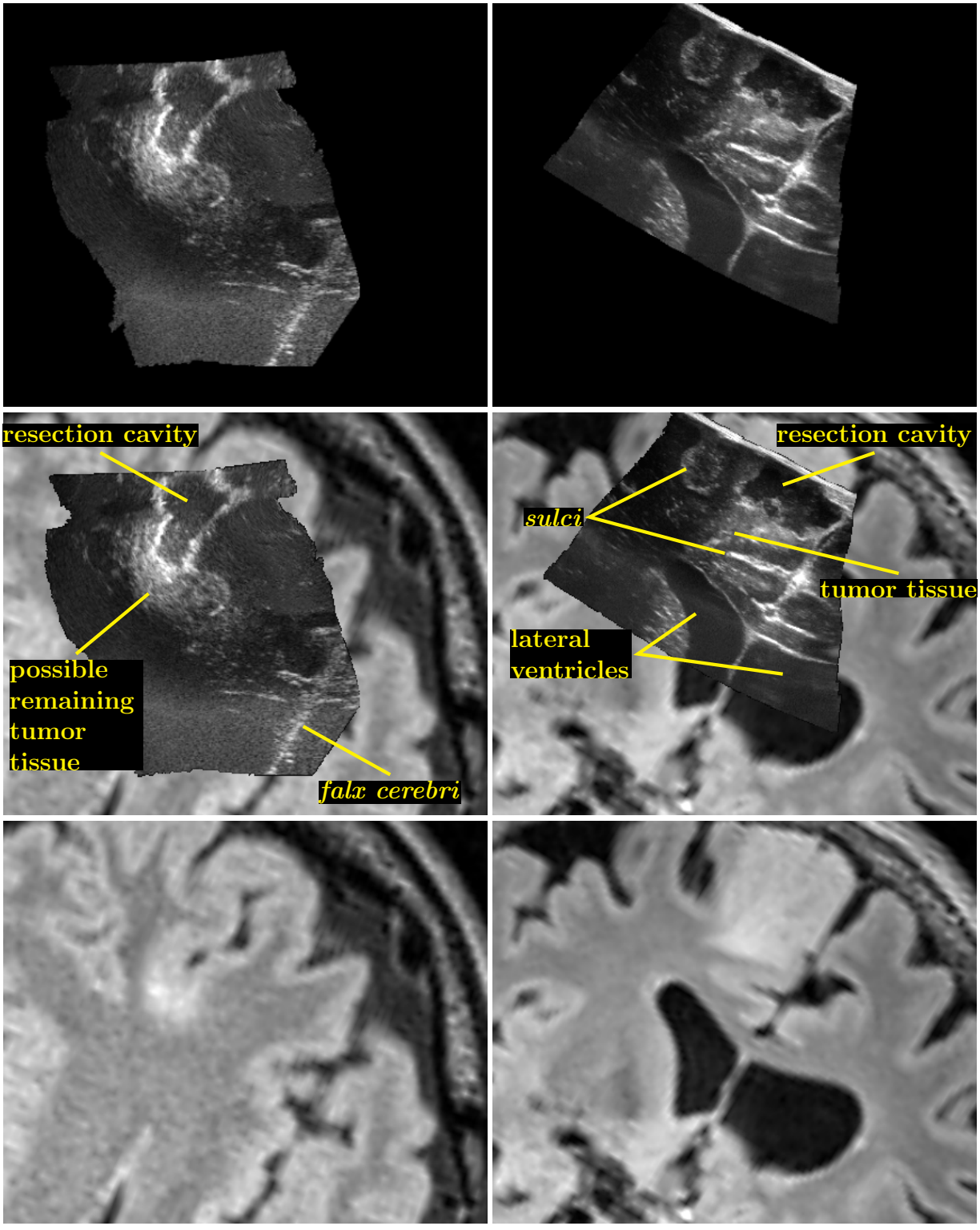
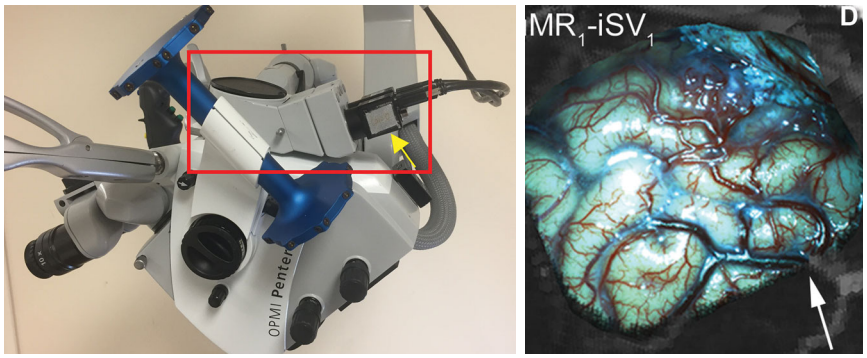


Figure 1.10 – Intraoperative ultrasound images of the brain (top row), overlaid over the registered MRI (middle row), registered MRI only (bottom row).



Images from Fan et al. [16].

Figure 1.11 – Stereo vision camera (left) and image of the cortical surface acquired with a stereo vision camera (right).

Laser range scanners are another imaging technique for acquisition of surface data. As for stereo vision cameras, it has been used for brain shift compensation [17].

1.5.4 Computed tomography (CT)

CT is rarely used for brain tumor surgery, due to lower contrast for soft tissue than MR, and the use of ionizing radiation (in contrast to MR). It is even less used intraoperatively due to the added contraints in the operating room.

1.5.5 Intraoperative imaging

MR is the standard preoperative imaging method used for brain tumor surgery [9, 10], due to good contrast of soft tissue in the resulting images. While it is commonly used preoperatively, it is also sometimes used intraoperatively. This is less common, due to the need of specialized equipment and changes in surgical procedures. Intraoperative MR (iMR) equipment is costly, and requires a dedicated operating room. In addition, iMR acquisition has been reported to increase surgery time by about one hour per scan [18].

Intraoperative ultrasound (iUS) is sometimes used as an alternative to iMR, due to its lower cost and easier integration in the operating room [13, 14]. There are however several differences that make understanding ultrasound (US) images challenging. Due to the physics of US, the intensities depend not only on tissue type but also on probe orientation and depth, and there may be artifacts and noise. In addition, the field of view is limited to the maximum depth of the probe, and to the angular movement during acquisition. Thus, it contains only the part of the brain around the tumor that is of interest, unlike MRI which covers the whole brain.

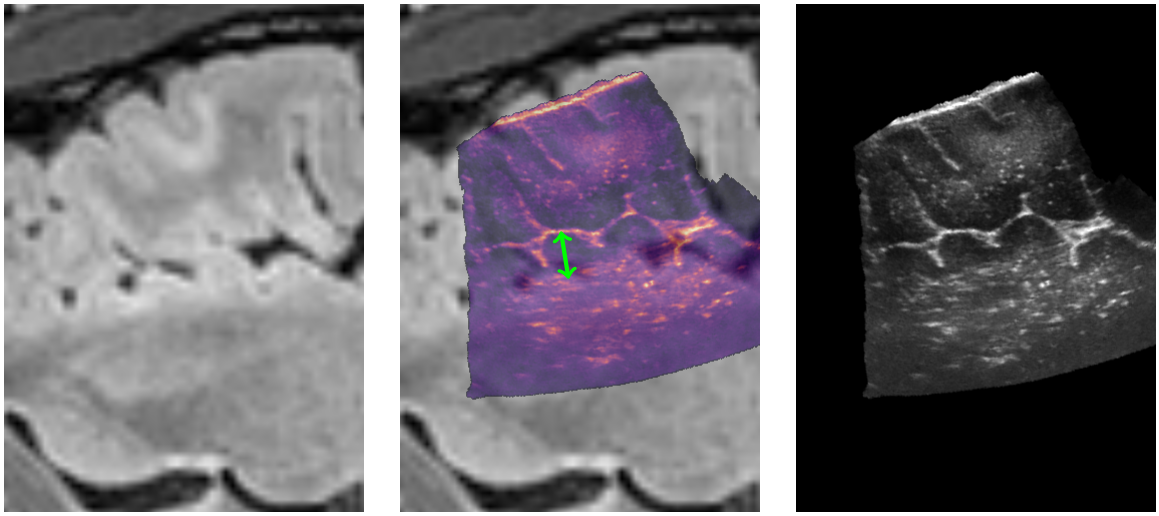


Figure 1.12 – Brain shift illustrated with a pMR (left) and iUS (right).

Optical systems such as stereo vision cameras and laser range scanners can also be used for intraoperative imaging. The main limitation of these systems is that they can only acquire surface data. They are nonetheless useful in applications such as image registration.

1.6 Brain shift

Brain shift denotes the displacement of brain tissue during surgery, which invalidates the patient-to-image mapping used for neuronavigation. Is has several causes, which can be physical, surgical, and biological [18]. Displacement is usually larger on the surface, on which displacements up to 20mm were reported. Several studies reported mean displacements of around 6 to 7mm. Figure 1.12 illustrates brain shift highlighting differences between the pMR and an iUS volume. The *sulcus* is the middle of the image is shifted, and the footprint in the top of the iUS is slightly below the tissue boundary in the pMR.

Physical factors include the effect of gravity, patient positioning, and movement of the head relative to the clamp used to prevent movement. Gravity is one of physical factors that contribute the most to brain shift. Gravity exerts a downward force, causing tissue sag. Patient positioning, which is usually different during surgery from the position used during preoperative MRI acquisition, also affect the position of brain anatomical structures. During surgery, the head of the patient is positioned such that the target location for craniotomy is accessible to surgeons, and such that CSF loss is minimized. During MRI acquisition, however, patients are commonly in supine position, with the head facing up.

Surgical factors are related to the use of surgical equipment and effects of surgery. Sur-

gical equipment such as skin retractors, surgical drapes, and tools for craniotomy can cause displacement of the head relative to the reference frame. The loss of CSF after the craniotomy and the *dura mater* has been opened is also a displacement factor. Finally, large displacements can be caused by tissue resection. In particular, the remaining tissue above resected tissue will move due to gravity as it is not supported anymore.

Biological factors include the use of drugs to lower intracranial pressure, and tumor type. For example, it has been observed that gliomas often cause more shift than meningiomas.

1.7 Aims

Brain shift impairs image navigation in brain tumor resection surgery, as the navigated image is the preoperative MRI. Intraoperative imaging is needed for accurate navigation; however, intraoperative MRI is impractical and other modalities have a more limited field of view and lesser quality than MRI (section 1.5.5). As such, several methods have been proposed to register the preoperative MRI with intraoperative data. This process consists in estimating the brain shift displacements with intraoperative imaging, and updating the preoperative MRI accordingly. The updated MRI can then be used for navigation. Section 2.1 reviews such methods in details.

In this thesis, we chose to use US intraoperative images for registration, because it is convenient to use in the operating room and has subsurface data. More specifically, we aimed to investigate which anatomical structures can be extracted in brain US images to register the preoperative MRI by matching these structures. There are very few such structure-based registration methods for brain US images, which motivated our study. We aimed to propose automated methods to 1) delineate the structures (image segmentation) and 2) register the preoperative MRI with the intraoperative US image and the extracted structures.

Many image segmentation and registration methods have been proposed. In this thesis, we focused on machine learning models, which have seen increased use for image processing, and have proven successful in many medical image processing tasks. They have been less studied for complex tasks such as US images processing, and image registration, although it is an active research domain. US images processing is challenging due to the limited field of view and lower image quality. Image registration aims to determine the transformation between two images. While there are many transforms that would correspond, only the actual transformation between the two objects that were imaged is the solution that should be found by image registration algorithms, making the problem ill-posed.

Since our main goal is structure-based registration, we first studied segmentation methods for US images of the brain using machine learning. In addition to registration, the segmentation of brain structures is valuable for several other purposes. The delineation of

tumor tissue is useful not only to guide surgeons when navigating the images, but also to help localizing remaining tumor tissue that should be removed for the resection to be complete. The resection cavity is also useful for brain shift compensation. Tissue resection is one of the causes of brain shift, which further extents tissue displacement during surgery (see section 1.6). As such, the extent of resection could be used in brain shift compensation methods to account for the shift induced by tissue resection. Finally, other structures such as *sulci*, *falx cerebri* and ventricles can help surgeons with navigating US images, as well as provide landmarks for image registration algorithms.

We also studied registration of the preoperative MRI with intraoperative US images using deep learning. This process aims to update the preoperative MRI so that it matches the intraoperative US image, and thus is corrected for brain shift. The updated MRI can be then used for navigation, which is more convenient than the US images due to higher quality and better contrast.

1.8 Contributions

We proposed single-class segmentation models for the resection cavity and tumor tissue, and studied the impact of various method parameters, such as the sampling strategy, using 2D or 3D models, and the inclusion of preoperative data. The, we studied multi-class segmentation of several structures, including the *sulci* and *falx cerebri*. We proposed a weighting strategy which allowed training of the models despite the high class imbalance, with classes entirely missing in some of the volumes. In addition, we implemented a recently proposed Hausdorff distance loss function term, and compare models trained with and without this term. Finally, we investigated registration models for registering the pMR with iUS volumes.

In addition, we have manually labelled the resection cavity, *sulci*, *falx cerebri* and ventricles in all volumes of the RESECT database, and the resection cavity in the post-resection iUS volumes of the BITE dataset. The resection cavity segmentations were then reviewed by a neurosurgeon, and corrected accordingly. We are in the process of validating the other segmentation masks, and plan to publish these segmentations, so that further studies may benefit from them.

Chapter 2

Literature review

In this chapter, we first review existing brain shift compensation methods. Next, we give an overview of brain MR and US segmentation. We then focus on deep learning methods, starting with technical background, followed by its application to image segmentation and registration. We conclude by presenting the two publicly available datasets of brain US images.

2.1 Brain shift compensation

The preoperative MRI (pMR) does not reflect the tissue displacement occurring in the operating room. In the other hand, intraoperative imaging modalities have limitations for use in brain tumor resection surgery (see section 1.5.5). No imaging system provide both ease of use in the operating room and sufficient image quality. In particular, iMR can have sufficient image quality but makes the surgical procedure longer and requires dedicated expensive equipment. As such, the pMR remains the reference image despite brain shift. Thus, several studies aimed to update the pMR to compensate for brain shift, using intraoperative data and/or modeling of the brain [18–20].

This process consists in finding the transform that maps one image into the other, and then applying this transform to generate the updated image. There are several methods for finding the transform, which can be grouped into two categories: image-based methods, which use image registration algorithms with the pMR and intraoperative images, and model-based methods, which model brain shift with biomechanical models.

2.1.1 Image-based methods

These methods use the pixel intensities of the images to estimate the transform between the two images. This class of methods have been widely studied, and a detailed and compre-

hensive review is out of the scope of this literature review. Instead, this section focuses on methods that were applied to brain shift compensation, and suggest a more general review of medical image registration by Sotiras et al [21]. Image-based methods can be classified into methods that optimize a similarity metric [22–29] and methods that use point or feature correspondences [30–39].

2.1.1.1 Similarity metric optimization

Similarity metrics are functions that measure the similarity between two images. They can be used for image registration by optimizing the transformation parameters such that the similarity is maximized. In the multimodal case, such as pMR-iUS registration, intensities cannot be directly used because of the different modalities. One possible approach is to generate pseudo images to transform the problem into a monomodal registration problem [40]. Alternatively, information-based metrics such as mutual information or cross-correlation can be used for multimodal registration.

Rivaz et al. [22] propose to use the correlation ratio (CR) for multimodal pMR-iUS registration, and normalized cross-correlation (NCC) for monomodal US registration. The correlation ratio was also used in other multimodal registration studies [28, 29]. They evaluate three registration strategies, starting with registering the pMR with the pre-resection iUS, then with the post-resection iUS directly, and finally using the composition of the pMR to pre-resection iUS and pre- to post-resection iUS registration. Rivaz et al. [22] report a reduced mean target registration error (mTRE) from 4.1mm to 2.4mm for MR-US registration and from 3.7mm to 1.5mm for US-US registration, averaged on 13 patients. Xiao et al. [28] reduce the mTRE from 7.2mm to 1.7mm on average for 4 patients.

Normalized gradient field (NGF) is another similarity metric that can be used for MR-US registration [41, 42].

Recent brain shift compensation studies [23–26] used the linear correlation of linear combination (LC2) metric [43]. In [23], NCC is additionally used for iUS-iUS registration for iterative updates during surgery.

Heinrich [27] proposed to optimize a similarity metric based on image self-similarities using the dense displacement sampling (deeds) optimization algorithm.

Results for these methods are reported on table 2.1. While they are generally not evaluated on the same datasets, making comparison difficult, LC2 and deeds appear to be the state-of-the-art methods based on similarity metric optimization for multimodal MR-US registration.

	modalities	method	patients	mTRE before registration	mTRE after registration
Rivaz et al. [22]	MR-US	similarity metric (CR)	13	4.1mm	2.4mm
Xiao et al. [28]	MR-US	similarity metric (CR)	4	7.2mm	1.7mm
Mok et al. [41]	MR-US	similarity metric (NGF)	10	6.4mm	5.7mm
Hager et al. [42]	MR-US	similarity metric (NGF)	10	6.4mm	7.0mm
Riva et al. [23]	MR-US	similarity metric (LC2)	8	5.9mm	2.7mm
Iversen et al. [24]	MR-US	similarity metric (LC2)	13	7.7mm	4.5mm
Wein [25]	MR-US	similarity metric (LC2)	27	5.4mm	1.8mm
Heinrich [27]	MR-US	similarity metric (deeds)	22	5.4mm	1.7mm
Rivaz et al. [22]	US-US	similarity metric (NCC)	13	3.7mm	1.5mm
Masoumi et al. [29]	US-US	similarity metric (CR)	22	5.4mm	2.8mm
Riva et al. [23]	US-US	similarity metric (NCC)	8	6.2mm	4.2mm
Reinertsen et al. [31]	MR-US	feature-matching (blood vessels)	7	7.4mm	2.8mm
Nitsch et al. [39]	MR-US	feature-matching (*falx*, *tentorium*)	11	16.9mm	2.2mm
Hansen et al. [44]	MR-US	deep learning (PDD-net)	10	6.4mm	3.1mm
Canalini et al. [38]	US-US	feature-matching (*sulci*, *falx*)	17	3.5mm	1.4mm
Machado et al. [36]	US-US	feature-matching (SIFT)	9	3.3mm	1.5mm
Miga et al. [17]	MR-LRS	biomechanical model	16	10.1mm	2.7mm
Morin et al. [45]	MR-US	biomechanical model	4	5.3mm	1.7mm

Table 2.1 – Quantitative evaluation of registration methods for brain shift compensation.

2.1.1.2 Feature-based methods

These methods use point or feature correspondences to compute the best transformation parameters. They first establish a set of homologous points or corresponding features in the two images, and then use these correspondences to determine the transform parameters.

Several methods using blood vessels for correspondences have been proposed [31, 32]. Reinertsen et al. [31] proposed a method in which blood vessels centerlines are extracted from angiographic pMR and Doppler-mode iUS, which are then registered using the iterative closet point (ICP) algorithm. Ilunga-Mbuyamba et al. [32] extracted blood vessels from contrast-enhanced US images using Hessian filters, and used the vasculature to register the pMR to pre- and post-resection US.

Other methods aims to match hyperechogenic structures such as the *sulci* and *falx cerebri* [30, 33, 34, 38, 39]. Farnia et al. [33] first proposed a method matching hyperechogenic structures and optimizing the residual complexity similarity metric [46] in the wavelet domain. In a later study [34], they report improved results with a co-sparsity function.

Drobny et al. [35] proposed a block matching algorithm using NCC, to establish point correspondences. The transformation parameters are then determined using a least trimmed squares regression. The image is then resampled using the transform, which can be used for another iteration to refine the transform.

Machado et al. [36] used Sift feature descriptors to find points of interest in iUS volumes, which were then paired based on descriptor values and finally used to compute the transform using thin-plate splines (TPS). This method was then compared to a model-based method (see next subsection) in a recent study [47]. Machado et al. [37] then proposed a novel method using a correlation-based similarity metric with multi-orientation and multi-scale features.

Deep learning models have also been proposed for pMR-iUS registration [44,48–50]. Deep learning for image registration is reviewed in more depth in section 2.3.3.

2.1.2 Model-based methods

Model-based methods model brain shift with a biomechanical model of the brain [16, 17, 51–64]. The physical behavior of tissues is modelled with physically-based equations. These equations are then discretized and numerically solved using the finite element method (FEM).

The equations describing the behaviour of brain tissue are based on a physical model of tissue. Many proposed models use linearly elastic equations [52,55–57,60–62,64], while more complex ones model tissue as an elastic body with interstitial fluid [17, 51, 53, 54, 58, 59, 63]. More advanced models include viscoelastic models, in which viscosity is taken into account.

Anatomical structures such as the *dura mater*, *falx cerebri*, and *tentorium cerebelli* can be modeled to improve the model accuracy. The *falx cerebri* can be modeled as a membrane with a higher stiffness and/or with boundary conditions [17, 51, 59, 63, 65]. Several studies suggested that modeling the *falx cerebri* can reduce errors in that area [52,59]. The *tentorium cerebelli* has also been modeled in some methods [17, 59, 62, 63]. In Chen et al. [59], nodes located at boundaries are assigned slip boundary conditions, such that they cannot move in the normal direction (through the *tentorium*). In DeLorenzo et al. [62], nodes corresponding to the *tentorium cerebelli* are fixed.

Intraoperative imaging may be used to constraint the model such that the estimated deformation matches the acquired intraoperative data. This can be achieved by adding boundary conditions to the FEM model corresponding to the observed displacement between the pMR and intraoperative measurements. Several methods [52, 53, 57, 61] proposed to use iMR to estimate the displacements of anatomical structures. Manual segmentation of the iMR images is not feasible in the operating room [61], however iMR segmentations can be estimated from pMR segmentations after registering the iMR with the pMR [52, 53]. Other studies used intraoperative acquisitions of the cortical surface with laser range scanners [17, 59,63] or stereo vision cameras [16,54,62] to estimate the displacement of the cortical surface. Bucki et al. [60] and Morin et al. [45] extracted the blood vessels from preoperative MRA and Doppler-mode iUS, between which the displacements are computed. Mohammadi et al. [64]

used a combination of stereo vision and Doppler-mode iUS to estimate the displacements.

During surgery, surgeons often need to spread out tissue to access subsurface tumor tissue. This can be achieved with retractors, which are surgical tools used to move and retain tissue. Tissue retraction introduces a discontinuous deformation of tissue, which is complex to model. Miga et al. [51] used specific boundary conditions to model tissue retraction, while Ferrant et al. [52] removed corresponding mesh nodes in the model. In Sun et al. [54], the tissue displacement due to retraction is estimated with intraoperative stereo vision, which is used to set boundary conditions. Vigneron et al. [61] proposed to use the extended FEM which can handle discontinuities, to model retraction as a discontinuous deformation.

Tissue resection induces further tissue displacement during surgery. Few biomechanical models incorporate such displacements [16,51,52,60], and have limitations that are discussed in the following section.

2.1.3 Modeling resected tissue

Most existing studies have focused on registration before tissue resection. Once tissue resection has started, brain tissue is further deformed due to tissue removal. Additionally, while the resection cavity is visible on intraoperative images, it is not present on the pMR. Thus, it is essential to account for the resection cavity to accurately register the pMR.

Existing methods that estimate the resection cavities have limitations. Some methods use manual or semi-automatic segmentation methods that require user input [51,52], which is impractical in the operating room. Other methods use simple estimations [60] or estimations with surface data only [16], which can be problematic as the shape of the resection cavity can be complex. More recently, Kochan et al. [66] proposed to adapt the NCC metric in their pMR-iMR registration method for better handling the resection cavity, which is not present in the pMR. This method was only tested for monomodal registration, and iMR has limitations (see section 1.5.5).

In contrast, iUS provides subsurface information and iUS volumes can be acquired such that the whole resection cavity is covered. Furthermore, iUS is convenient to use in the operating room. Therefore, using iUS to accurately delineate the resection cavity looks promising for improving the registration methods. Previous studies have shown that deep learning can successfully segment structures in iUS images automatically [1, 4, 38, 39, 67], including the resection cavity [1, 67] (see chapter 3). Only one study [67] applied this to registration, more specifically iUS-iUS registration. To the best of our knowledge, no pMR-iUS registration method using iUS segmentations of the resection cavity has been proposed.

2.2 Brain segmentation

In this section, methods for brain segmentation are reviewed. Image segmentation consists in classifying each voxel into several groups, which are called classes. Segmentation images are the images that are the result of the segmentation process, in which each pixel is given a value corresponding to a class. In brain images, classes of interest typically include tissue type (WM, GM, tumor tissue), CSF, and anatomical structures. The segmentations of brain images are not only useful for analysis and surgical planning, but also for many brain shift compensation methods, which can use the segmentations of brain tissue and anatomical structures in their process.

Segmentation of brain structures has been mostly studied for MR images. As iUS quality has improved in the last decade, it has seen increased usage for various applications in the operating room, including brain tumor resection surgery. This prompted research on brain iUS segmentation, although there are fewer studies of brain US segmentation than other US applications (such as US imaging of the heart, breast, and prostate) and this research domain is still active.

2.2.1 In MR images

Segmentation of brain MR images has been widely studied [68–70]. The MICCAI BRATS challenge on brain MR segmentation has been organized yearly since 2012 [71]. Many studies have focused on classifying brain voxels into the three main components: WM, GM, and CSF. Segmentation of tumor tissue is also highly relevant for surgical planning and brain shift compensation methods, thus, many methods have been proposed to that end [70].

This section gives an overview of brain MRI segmentation techniques, starting with common preprocessing methods, and then summarizing the different types of methods.

2.2.1.1 MRI preprocessing

MR images are often preprocessed prior to segmentation, to increase robustness and accuracy of the segmentation methods.

MR non-uniformity correction is a preprocessing step which compensates for signal variation due to the MR acquisition process, such as inhomogeneity of the magnetic field, and sensitivity of the receivers. Because of this signal variation, the intensity of tissue or CSF may vary spatially even though there is no change in the physical properties of the material.

This complicates the segmentation process, as many segmentation methods assume that intensity is relatively homogeneous in each class, and that there is limited intensity overlap between classes. Thus, MR images are commonly preprocessed to correct this non-uniformity, for which various methods have been proposed [72].

Brain mask extraction consists in separating brain voxels from non-brain voxels. This allows non-brain voxels to be discarded, such that only brain voxels are segmented. The benefit of such masking is to prevent non-brain structures, whose intensities may overlap with those of brain tissue, from being classified as brain voxels.

Other common preprocessing steps may include registration to standard coordinate systems and/or to a brain atlas, denoising, and intensity normalization.

2.2.1.2 Thresholding methods

Thresholding methods consists in finding thresholds that separates intensities into groups that correspond to segmentation classes. The thresholds may be global (one value for the entire image or volume) or local (corresponding to a neighborhood of voxels).

These methods are rarely used for brain MRI segmentation, as determining an satisfying threshold is difficult due to inhomogeneity, noise, and complex intensity distributions.

2.2.1.3 Clustering methods

Clustering algorithms classify data into several groups based on patterns.

K-means is a clustering algorithm in which each cluster is given an initial center point, which is then updated iteratively until convergence. The center points are updated such that a similarity metric is maximized, or a dissimilarity metric is minimized. The sum of squared distances is a commonly used dissimilarity metric.

K-means can be used for WM, GM, and CSF segmentation in brain MRI [73]. Abdel-Maksoud et al. [74] notes that while K-means can locate tumor tissue, the resulting segmentations may be incomplete. They propose a hybrid method combining K-means and fuzzy C-means (see next paragraph) to overcome this limitation.

Fuzzy C-means (FCM) [75] is a clustering similar to K-means, which differs from K-means in allowing data points to belong to several clusters. In constrast, points are assigned to a single cluster in K-means.

Several works have used FCM for brain MRI segmentation. In particular, Ahmed et al. adapted the similarity metric to estimate and account for MRI non-uniformity [76]. FCM is also used to compute initial segmentations in other methods [77, 78].

Expectation maximization (EM) is an algorithm which finds the maximum likelihood parameters of a statistical model given the observed data points. The maximum likelihood parameters are the parameters maximizing a likelihood function, and thus can be thought to be the most probable parameters. EM works iteratively by alternating the expectation and maximization steps. The expectation step updates latent variables given the model parameters. The maximization step updates the model parameters given the latent variables.

It can be used for clustering, using a gaussian mixture model in which each class correspond to a gaussian function in the model. In this case, latent variables are the class the observation belong to. In the case of image segmentation, the observations are the voxel intensities.

2.2.1.4 Atlas-based methods

Atlas-based methods consist in registering the MR image to a reference MR image called atlas, turning the problem into an image registration one. Once the mapping between the current image and the atlas is known, a segmentation of the atlas can be mapped to the current image. The segmentation of the atlas is only done once, so it can be done manually (while time consuming, manual segmentation are the most accurate).

Several image registration methods have been proposed. They usually consist in maximizing an image similarity metric, with regularity constraints. An overview of image registration methods can be found in this review [21]. Popular software packages include FreeSurfer [79] and Caret [80], which have been compared [81].

The main benefit of atlas-based methods are that they do not dependent on which classes need to be segmented. These classes only need to be segmented once in the atlas, possibly manually if no other method exists and/or for high accuracy. However, atlas-based methods may suffer from high inter-patient variability, and the reference image need to be chosen as a representative sample.

2.2.1.5 Surface methods

Surface methods find surface boundaries that delimit the classes in the volumes. These boundaries are normally located where intensity values are changing.

Active surfaces Active surfaces are surfaces defined explicitly with a set of controlling points, which are iteratively updated such that they move towards the boundaries in the images. Active surfaces are the 3D equivalent to active contours in 2D.

An initial surface needs to be initialized, either manually or with another method. This initial surface is then iteratively updated given image intensity and surface shape constraints. The image intensity constraints ensure that the curve moves to boundaries in the images, where intensity values are changing. The surface shape constraints force a certain regularity and smoothness for the surface.

Level-set are similar to active surfaces, except that the surfaces are defined with a function. The surface correspond to the points for which the function takes a given constant value, called level. The surface is moved by changing the level.

Level-set methods have been proposed for WM and GM segmentation [77, 82] as well as tumor tissue segmentation [78].

2.2.1.6 Machine learning methods

Machine learning refers to a class of methods that use data to determine the parameters of a model. A model is a function with a number of parameters, which use given input values and model parameters to compute output values. Machine learning consists in determining the best set of model parameters from sample data, such that the model outputs expected values for a particular task. In the case of image segmentation, the inputs are the image (or volume in 3D) as a 2D or 3D array of intensities, and the output is the corresponding segmentation image, which gives the class each pixel belongs to.

Deep learning is a subset of machine learning in which models contain a very large number of parameters. These models are usually composed of several interconnected layers, which form a network called artificial neural network. Deep learning for image segmentation is reviewed in details in section 2.3.2.

Determining the model parameters is called the training phase. This is done with a set of sample data, called training set; thus, parameters are said to be learned from the training data. Once the model parameters have be determined, the accuracy of the model

is evaluated with a distinct set of data called test set (not seen during the training phase). During the test phase, the model outputs for the test set is compared to the actual expected values (called ground truth) using distance or similarity metrics. These metrics indicate the accuracy of the model, and its capacity to generalize to unseen data.

k-nearest neighbors (k-NN) is a learning algorithm for classification. The training phase only stores the training values and labels. The output of the model is the most frequent label among the k most similar inputs in the training samples, according to a given similarity metric, and where k is a constant. For image segmentation, the input values are usually the intensities, although it can also be specific features, obtained by a transformation of image intensities. k-NN has been used in MRI segmentation, for example to classify CSF, WM and GM [83] and to segment lesions [84].

Support vector machines (SVM) are classification models in which hyperplanes separate the classes in the input space. Given an input space of dimension n, the training of an SVM model consists in finding the optimal hyperplanes (space of dimension $n-1$) separating the classes. As such, the input space needs to be linearly separable. Thus, the input is generally processed to extract features, such that the feature space is linearly separable.

In brain MRI segmentation, various features can be used, such as intensity and texture-based statistics [85], wavelet transform coefficients [86], or a combination of both [87].

Random forests are classification models that are composed of multiple decision trees. Several studies proposed to use random forests for tumor segmentation in brain MRI [88,89].

Neural networks are deep learning models that are composed of several layers. Deep learning segmentation networks are reviewed in detail in section 2.3.2. These models achieve state-of-the-art results for image segmentation, and are actively researched.

2.2.2 In ultrasound images

Segmentation of ultrasound images is much more difficult than MRI segmentation, due to several limitations of ultrasound. As explained in section 1.5.5, ultrasound images contain noise and artifacts, are dependant on probe orientation and depth, and their field of vue is limited. As such, many methods that can segment MR images cannot be used as is.

Speckle is one of the main reasons that make US segmentation challenging for traditional segmentation methods. Speckle refers to the granular noise in US images that is due to

backscattered echoes of the US signal. Speckle reduction methods have been proposed to improve the reliability of traditional segmentation methods for US images [90, 91]. Common techniques for speckle reduction include adaptive filters [92], anisotropic diffusion filters [93, 94], wavelet transforms [95, 96], and nonlocal means filters [97, 98]. More recently, several studies proposed deep learning methods for speckle reduction [99–101].

Noble et al. reviewed US segmentation methods [102]. Since then, research has been mostly focused on machine learning [103]. Machine learning is particularly interesting for US segmentation, which is generally difficult to apply traditional segmentation methods to. Research on brain US segmentation is however rather limited, compared to other US imaging applications. This is partly due to US imaging being not that common in neurosurgery, although it has seen increased usage and research interest in the last decade. Segmentation methods have been proposed for the midbrain region [104], resection cavities [1,67, chapter 3], tumor tissue [4, 105, 106, chapter 4], *falx cerebri* and *sulci* [5, 38, 39, chapter 5]. Some of the aforementioned studies [1, 4, 5] were conducted as part of this thesis and are detailed in the next chapters.

2.3 Deep learning for image segmentation and registration

This section starts with technical background on artificial neural networks. Deep learning methods for image segmentation and registration are then reviewed.

2.3.1 Artificial neural networks

2.3.1.1 General structure

Artificial neural networks are a class of machine learning models which are composed of several interconnected units called layers. The layers are functions taking inputs and computing outputs. The layers are connected such that a layer can take the output of another layer as input, and its output may be used as input of another layer. This forms a network, ie. a graph of functions, which itself has inputs and outputs. The inputs and outputs of the network are inputs and outputs of layers that are not connected to other layers. Note that a neural network may be defined hierarchically, because a neural network can be used as a layer in a bigger neural network.

A layer computes an output given its inputs and its parameters. Parameters are variables used in the computation of a layer's outputs, that are learned from data during the training

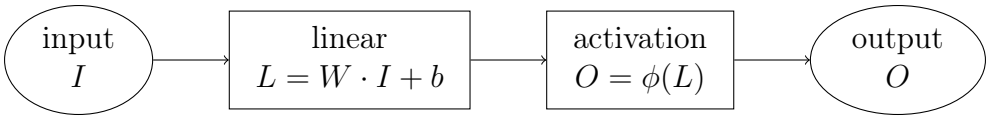


Figure 2.1 – A perceptron: $O = \phi\left(W \cdot I + b\right)$ where I is the input, O is the output, W is the weight matrix, b is the bias vector, and ϕ is the activation function.

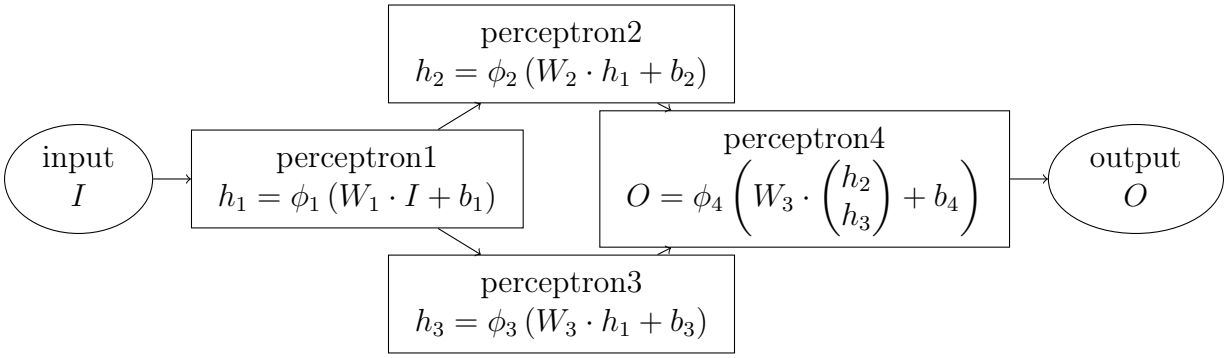


Figure 2.2 – An example of a multi layer perceptron: a neural network composed of several perceptrons.

process.

The most basic unit of a neural network is the linear operation $O = W \cdot I + b$, where I is the input vector of size n, O is the output vector of size p, and W and b are the parameters (a $p \times n$ weight matrix and a bias vector of size p). In order to build more complex operations, non linear functions are used in between linear operations. These non linear functions are called activations. The combination of a linear operation followed by an activation is called a perceptron (see figure 2.1). The most simple neural networks are composed only of perceptrons, organized in layers. Figure 2.2 shows a schematic of such neural network.

2.3.1.2 Learning parameters

The parameters of a neural network are the parameters of all the layers of the network. Their values are learned during the training process, which finds optimal parameter values given training data and a loss function. The loss function sets the objective of the model, and is a measure of how wrong an output is, given some input data. Thus, the training process aims to find the model parameters that minimize the loss function.

This is done with an iterative process called gradient descent. The gradient of the loss function f with respect to the parameters is the vector of the partial derivatives of f. Thus, by moving the parameters towards the opposite direction, the value of the loss function

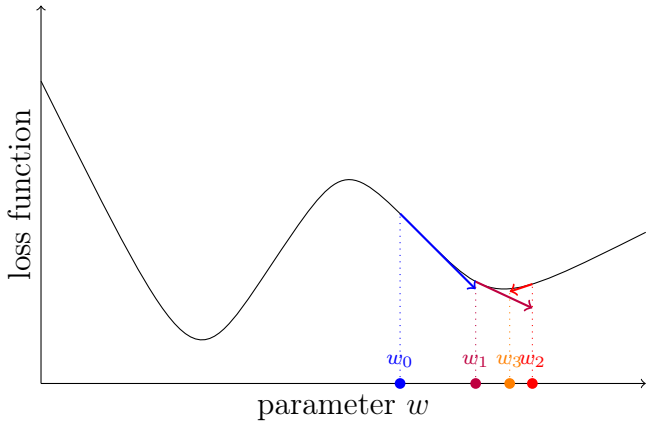


Figure 2.3 – Illustration of the gradient descent of a model with one parameter w.

decreases. Gradient descent consists in iteratively updating the parameters in the opposite direction of the gradient, until a local minimum is reached.

Figure 2.3 illustrates gradient descent, with a simplified example in which there is only one parameter w. In this case, the gradient is simply the derivative of the loss function. The parameter w is initialized with a value w_0, and is subsequently updated to $w_1, \ldots, w_3$ by increasing (respectively decreasing) w if the derivative is negative (respectively positive). It should be noted that the initial value w_0 is important and will determine which local minimum the algorithm will converge to. In the illustrated example, the algorithm converges to a local minimum which is not the absolute minimum.

At each iteration of the gradient descent, the gradient determines the direction towards which the parameters will be moved. The learning rate is a scalar weight applied to the gradient, which determines by how much the parameters are updated in that direction. A large learning rate induces a large displacement, and thus allows fast convergence of the training process. However, local minima may be skipped if the learning rate is too large, and may prevent convergence to a minimum. In contrast, small learning rates lead to more robust convergence, at a slower rate.

The training process is done with a training set containing several samples. The loss function should be minimized for all training samples, thus the total loss function is the sum of the loss terms for all samples. To reduce the memory complexity of computing the gradient of the total loss function, neural networks are usually trained by considering samples in smaller groups, called mini-batches. The training samples are divided in mini-batches which are randomly shuffled. The parameters are then updated for each mini-batch, using the gradient of the loss function for that mini-batch. An epoch refers to an iteration over all mini-batches, which is repeated until convergence. This process approximates the gradient descent process, and is called stochastic gradient descent (SGD). Various extensions of SGD have been proposed; in particular, the learning rate is adapted for each parameter

is algorithms such as AdaGrad, RMSProp, and Adam [107].

Due to the structure of neural networks, the function of which the gradient is computed is very complex, and includes many compositions of functions. Each block introduces a new composition of the block's function with it's input. The gradients are computed efficiently by using the derivation chain rule from the last output to the first blocks connected to inputs, storing intermediate values to avoid computing the same values several times. Because the gradients are computed in reverse order compared to when the function is evaluated, this process is call backpropagation.

2.3.1.3 Basic building blocks

This section introduces some of the most common layers used in neural networks.

Fully connected layers are layers which connects every outputs of the preceding layers to every inputs of the following layer, using linear operations and an activation function. Thus, it is a perceptron connecting all previous outputs.

The parameters of the layers are the weight matrix and the bias vector.

Convolutional layers are layers whose output is the convolution of the input with a kernel. The 2D convolution of an input image s with an $n \times p$ kernel k is the image d such as:

$$\forall i, j; d_{i,j} = \sum_{x=1}^{n} \sum_{y=1}^{p} k_{x,y} s_{i+x-\lceil \frac{n}{2} \rceil - 1, j+y-\lceil \frac{p}{2} \rceil - 1} \tag{2.1}$$

Each pixel in the output image is computed by looking at a neighborhood around the corresponding pixel in the source image, the size of the neighborhood being that of the kernel. The values of the neighborhood in the source image are multiplied element-wise with the kernel values. The output pixel is the sum of all these resulting values. Figure 2.4 is a graphical representation of this operation.

A 3D convolution is the same operation defined on 3D volumes, with a 3D kernel.

In a convolutional layer, the kernel values are parameters. Thus, these are learned during the training process. Since the input and output may be composed of several images (or volumes in 3D), the parameters are actually several kernels: a convolutional layer taking n_i images and outputting n_o images has $n_i \times n_o$ kernels corresponding to each pair of input and output images. The number of input or output images is also referred to as number of

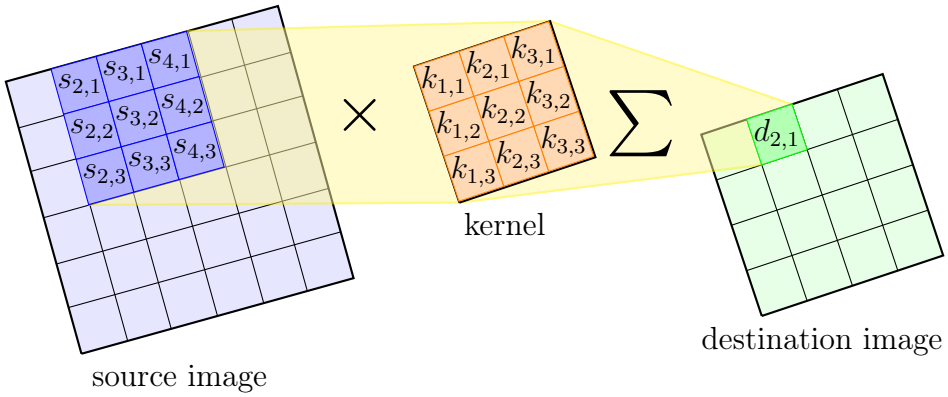


Figure 2.4 – A 2D convolution operation with a 3×3 kernel.

input or output features. A convolutional layer may also have a bias vector as for a fully connected layer.

Input images are usually padded such that the output images have the same size as the input images. This consists in defining a value for pixels out of the input images, used for the neighborhood of pixels close to the border. They may be set to zero (referred to as zero padding), or to the value of the closest pixel.

Pooling layers are similar to convolutional layers, except that the operation is not a convolution, but a given operation on the pixel neighborhood. For example, max pooling layers compute the maximum value in the neighborhood of the pixel. Max, min, and average pooling are the most common pooling layers. These layers do not have any parameters. The most common application of pooling layers is to decrease the spatial size while increasing the number of features.

Upsampling layers upsample an image using nearest neighbors or linear upsampling. These layers do not have any parameters. They may be used as the reverse of a pooling layer.

2.3.1.4 Activation functions

The most commonly used activation functions are threshold functions, which introduce non linearity. The following paragraphs present common threshold activation functions.

Rectified linear unit (ReLU) is the positive part function, ie. $\text{ReLU}(x) = max(x, 0)$. Several variations of this function have been introduced, for example to make it smooth.

This activation function is a general activation that is can be used after any layer.

Sigmoid (or logistic function) is a smooth approximation of the step function. It is defined as $\text{sigmoid}(x) = \frac{1}{1+\exp(-x)}$. Its output in in the range $[0, 1]$, and positive inputs are mapped to $[0.5, 1]$ whereas negative inputs are mapped to $[0, 0.5]$.

Sigmoid is often used as the last activation layer to constraint the output of the model to the range $[0, 1]$.

Softmax is a multidimensional generalization of sigmoid. For n inputs $x_1, \ldots, x_n$, it is defined as $\text{softmax}(x_1, \ldots, x_n) = \left(\frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}, i \in \{1, \ldots, n\}\right)$. The output is a vector of n number in range $[0, 1]$, whose sum is 1.

2.3.2 Image segmentation

This section gives an overview of the main neural network architectures and loss functions for image segmentation. For a more comprehensive review of medical image segmentation with deep learning, the reader is referred to this review [108].

2.3.2.1 Neural network architectures

Convolutional neural networks (CNN) are neural networks made of successive convolutional layers. Max pooling layers may be inserted between convolutional layers, so that features at learned at several scales. Depending of the end goal, the last layers may be fully connected layers to provide dense outputs, such as segmentation maps. Figure 2.5 shows a schematic of a typical CNN architecture. While CNNs were first designed for image classification, they have also had some success in image segmentation tasks, and were among the first neural network architectures used for image segmentation.

Fully convolutional networks (FCN) are CNNs in which the last layers are convolutional instead of fully connected. They were introduced by Long, Shelhamer, et al. [109] as the first pixel to pixel segmentation network trained end-to-end (ie. optimizing the entire model in one pass).

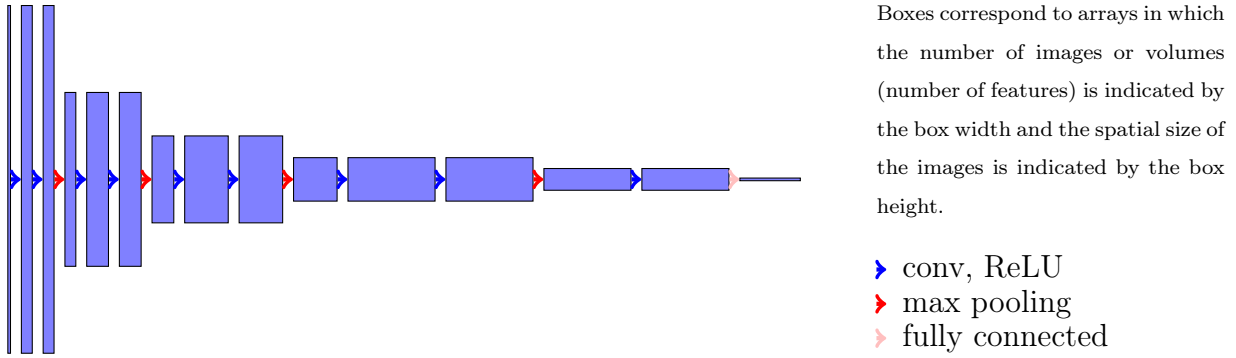


Figure 2.5 – Example of a CNN architecture.

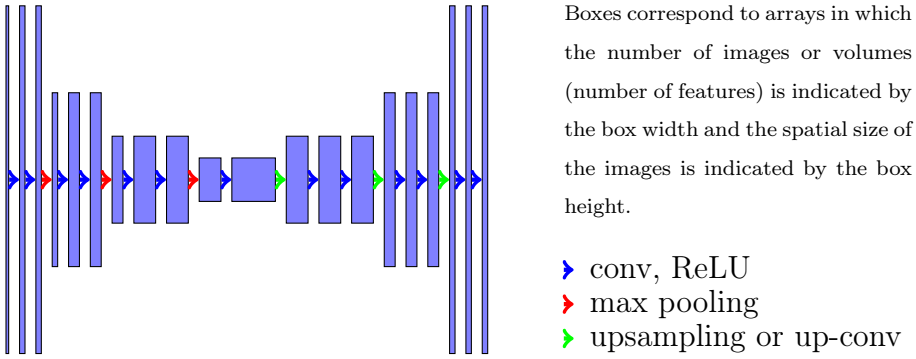


Figure 2.6 – Encoder-decoder network architecture.

Encoder-decoder networks are an evolution of FCNs to support larger input sizes. The first part of the network is the encoder, consisting of convolutional and max pooling layers as in FCNs, whose goal is to extract features from the input image. The second part is the decoder, which mirrors the encoder and outputs the segmentations from the extracted features. The decoder is structured like the encoder, except that max pooling operations become upsampling operations. Figure 2.6 shows a schematic of an encoder-decoder network.

U-Net is an encoder-decoder network in which skip connections are added. Skip connections connect the last layer of each scale in the encoder to the first layer of corresponding scale in the decoder. U-Net was proposed by Ronneberger et al. [110], and has been widely adopted and adapted to many medical image segmentation problems, due to its capacity to learn from small datasets. It is the basis of the models proposed in this thesis, which are illustrated in the next chapter, in figure 3.5.

2.3.2.2 Loss functions

This sections presents the most common loss functions for image segmentation. Ma et al. [111] conducted an extensive review of segmentation loss functions, and suggest the use of combinations of several loss functions.

In the following equations, y_t denotes the ground truth (as binary, with 0 for background and 1 for foreground), and y_p is the segmentation output of the model (in range $[0, 1]$).

Binary cross-entropy is a dissimilarity metric based on relative entropy (or Kullback-Leibler divergence). Its definition is provided in equation 2.2.

$$l(y_t, y_p) = -\sum_i y_t[i] \log y_p[i] + (1 - y_t[i]) \log(1 - y_p[i]) \tag{2.2}$$

Weighted binary cross-entropy is a weighted version of standard binary cross-entropy. Each class is assigned a weight: w_b for background and w_f for foreground (equation 2.3). This can improve the convergence of the models in case of imbalanced datasets.

$$l(y_t, y_p) = -\sum_i w_f y_t[i] \log y_p[i] + w_b(1 - y_t[i]) \log(1 - y_p[i]) \tag{2.3}$$

Dice loss is based on the Dice coefficient (also called F1-score) [112]. It was first proposed as a loss function in V-Net [113] to mitigate class imbalance. It is a smooth version of the Dice score that is suitable for backpropagation, as defined in equation 2.4.

$$l(y_t, y_p) = 1 - \frac{2 \sum_i y_t[i] y_p[i] + \epsilon}{\sum_i y_t[i] + \sum_i y_p[i] + \epsilon} \tag{2.4}$$

Hausdorff distance can be estimated in a loss function through various methods that were recently presented [114]. One method for approximating the Hausdorff distance consists in evaluating convolutions with several kernel sizes to estimate the largest distance.

2.3.3 Image registration

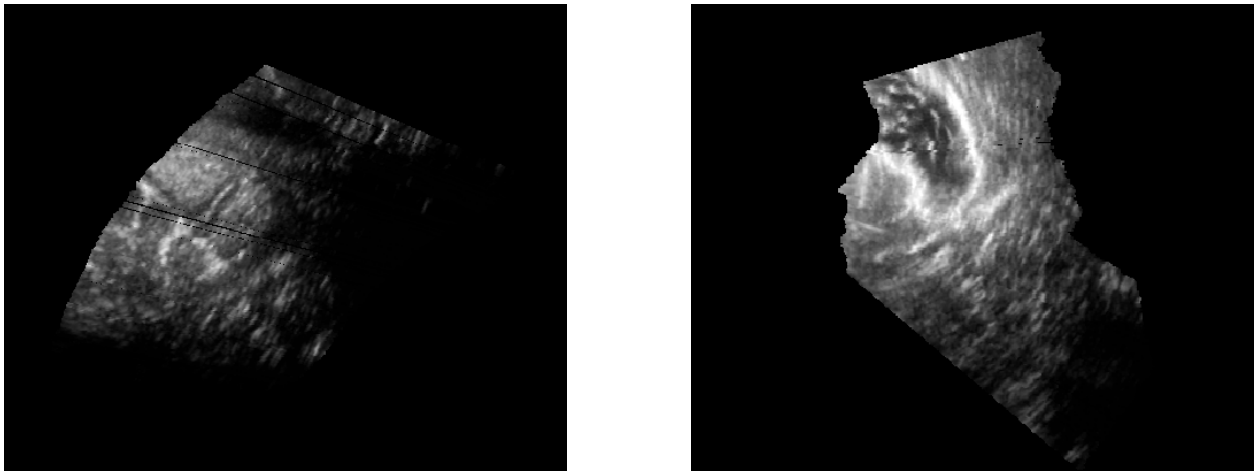
After having been extensively used in image classification and segmentation, deep learning techniques have been recently explored for image registration. Most proposed models are using well-known architectures such as CNNs [115–121] or encoder-decoder networks [122–127]. Less common alternatives include generative adversarial networks (GAN) [128–130]. Most of these methods are trained to output a dense displacement field (DDF) corresponding to the transformation between the two images. This DDF can then be used to transform the original image (moving) and obtain the registered image matching the fixed image.

The loss function used to train the model consists of several terms. In monomodal registration, an intensity term such as the L2-norm [125, 127] or NCC [119, 120, 125, 126] can be used. In multimodal registration, it is more challenging to use intensity-based metrics as the modalities are different. If ground truth DDFs are available, the DDFs can be compared directly with a metric such as the L1-norm [117, 123], L2-norm [115, 122, 127], or NCC [120]. Otherwise, the loss function may rely only on the similarity term (for monomodal registration), or use generated transformations. To keep the output DDF smooth, a regularization term such as the norm of the gradient of the DDF [119, 124, 125] can be added.

A novel approach using ground truth segmentations has recently been proposed by Hu et al. [124]. In their work, another term in the loss function using ground truth segmentation labels. In the training phase, ground truth segmentations corresponding to the moving image are registered along with the images. The registered segmentations are then compared to the segmentations of the fixed image by computing the Dice score. The segmentations are not part of the network input, thus they are not needed during the testing phase, which is a significant advantage of this method.

While the outputs of most models are DDFs, Heinrich designed a model with less transform parameters and builtin regularization constraints, pdd-net [131]. He reports better results than existing DDFs models.

Few deep learning models have been proposed for pMR-iUS registration for brain-shift compensation [44, 48–50]. Two MICCAI challenges were organized on this task using the RESECT database, in 2018 [132, 133] and 2020 [134]. Zhong et al. [48] used imitation-based learning to train a model that outputs piece-wise linear transformations. Sun et al. [49] proposed a model composed of two CNN encoders for extracting features from the pMR and iUS, respectively, and of a U-Net-like network generating a DDF from the extracted features. The loss function is a similarity metric between the transformed moving image and fixed image, that depends on both image intensity and gradient. Gunnarsson et al. proposed an method based on the multi-scale registration network PWC-Net [135], using a loss function with NCC intensity similarity, Dice label similarity, and an L2 regularizer. Pdd-net [131] was also evaluated on pMR-iUS registration [44], achieving the best results of the Learn2reg 2020 challenge [134].



(a) 8*a* **(b)** 8*u*

Figure 2.7 – Pre- and post-resection iUS volumes from BITE [136] dataset.

2.4 Datasets with brain intraoperative ultrasound images

There are only two publicly available datasets of brain iUS images: BITE and RESECT. Thus, we use these two datasets in this work.

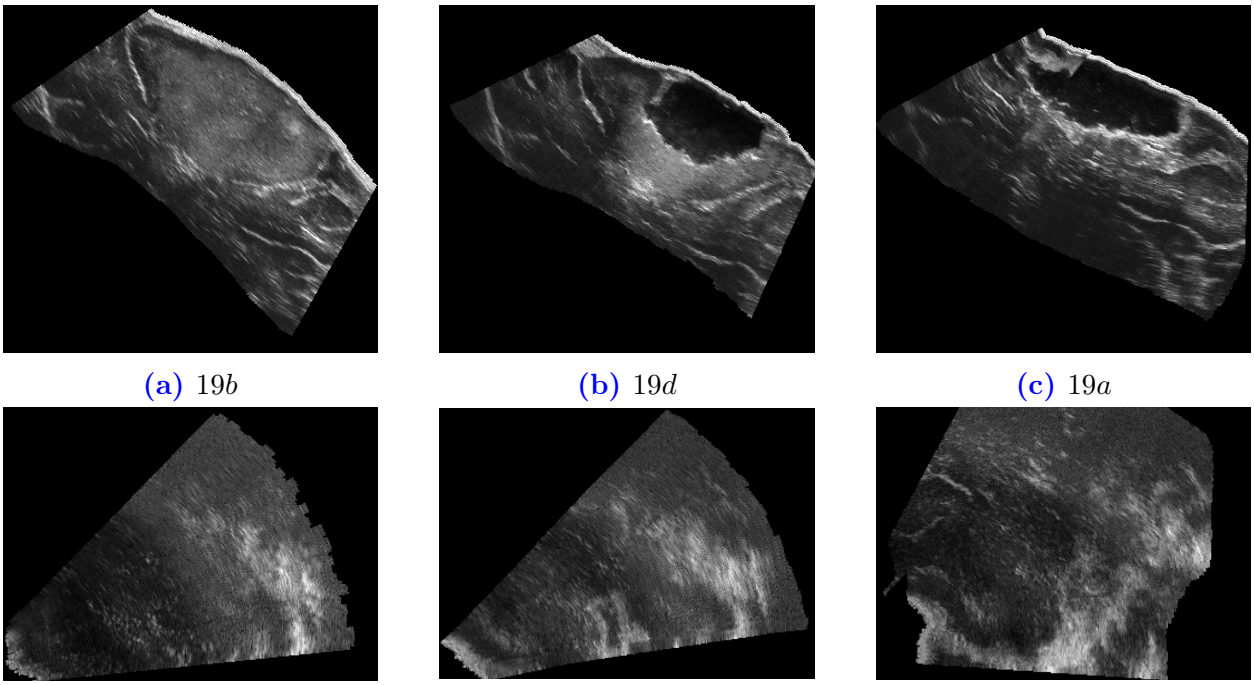
2.4.1 BITE

BITE [136] is a publicly available dataset of pre- and post-resection MR and US images. It contains images of fourteen patients, for which there are a pre- and post-resection MRI, as well as several iUS images. For each patient, there are two to five pre-resection iUS, labeled a to e, and for all but one patient, there are two to six post-resection iUS, labeled u to z. In the following, we refer to each individual iUS volume by the patient number followed by the acquisition letter, e.g. $13u$. Figure 2.7 shows iUS slices from the BITE dataset.

The MR volumes are T1w MRIs that were acquired with a 1.5T scanner.

The iUS volumes were acquired using a tracked 2D US probe, which was a P7-4 MHz phased array transducer. The individual 2D slices are available, as well as reconstructed 3D volumes. The mean 3D volume size is 316 voxels, ranging from 159 to 516. All volumes are isotropic with a voxel size of 0.3mm.

Since the field of view of US images is limited, several acquisitions with different angles were made so that as much anatomy as possible was covered. Pre-resection US images were



(a) 19*b* **(b)** 19*d* **(c)** 19*a*

(d) 3*b* **(e)** 3*d* **(f)** 3*a*

Figure 2.8 – iUS volumes before, during, and after resection from RESECT [137] dataset.

acquired with the probe on the *dura matter* when possible, otherwise directly on the cerebral cortex or on a dural repair patch. Post-resection US images were acquired either with the probe inserted into the resection cavity, which was filled with saline, or with the probe outside the resection cavity which was filled with saline and then closed with Dura-Guard.

For each patient, a set of ten homologous points in a pair of pre- and post-resection volumes were manually selected by a radiologist. There are also sets of nineteen to fourty homologous point in pre-resection pMR and iUS volumes. In addition, segmentation of tumor tissue in the pMR is provided in the database.

2.4.2 RESECT

RESECT [137] is a publicly available dataset more recent than BITE, with pMR and iUS volumes of twenty-three patients. For each patient, three iUS volumes were acquired, respectively before, during and after resection. In the following, we refer to each individual iUS volume by the patient number followed by b, d, or a for before, during and after resection, respectively. Figure 2.8 shows corresponding slices in before, during, and after resection iUS volumes.

T1w and FLAIR pMR are available for each patient. They were acquired with a 3T scanner for all but three patients, for which a 1.5T scanner was used. 3T scanners provide better contrast and lower noise than 1.5T scanners.

The iUS volumes were acquired with two types of tracked 2D linear US probes (12FLA-L and 12FLA), with a frequency range of 6-12 MHz. The mean volume size is 347 voxels, ranging from 221 to 492. All volumes are isotropic with a mean voxel size of 0.21mm (0.14-0.36mm). The volumes were acquired such that the whole tumor and/or resection cavity is covered.

A set of sixteen to twenty landmarks between the pMR and iUS before resection (iUSb) and between the pMR and iUS after resection (iUSd) were manually selected for most patients. In addition, a set of fourteen to twenty-two homologous landmarks between iUSb and iUS during resection (iUSd), and between iUSd and iUSa, are available for seventeen patients. No segmentations are provided with the database.

Chapter 3

Segmentation of the brain tumor resection cavity in iUS images

3.1 Introduction

As discussed in section 1.6, tumor tissue resection causes tissue displacement during brain tumor resection surgery, and thus is one of the causes of brain shift. This impairs navigation of the pMR, which no longer reflects tissues location accurately. To compensate for brain shift, several studies proposed methods to update the pMR with intraoperative data (see section 2.1). However, only a few methods compensate for resection-induced brain shift (see section 2.1.3).

The location of resected tissue is needed for accurate modeling of tissue displacement due to tissue resection. Both image-based and model-based methods would benefit from such information. Image-based methods could use the resection cavity segmentation to mask out voxels in that region, since they have no correspondence in the pMR. Model-based methods could integrate the resection cavity in their biomechanical model so that tissue displacement due to resection is modelled. In addition, the segmentation of the resection cavity may be used to assess the extent of resection and the presence of residual tumor tissue.

As such, we propose a method to automatically segment the resection cavity in iUS images, using deep learning. The deep learning models are evaluated on two datasets: BITE [136] and RESECT [137] (see sections 2.4.1 and 2.4.2). We compare different sampling methods, 2D and 3D models, and training strategies with the two datasets. Quantitative and qualitative results are then discussed in a detailed analysis. The methods and results presented in this chapter were presented at the SPIE 2019 medical imaging conference [2] and published in the Journal of Medical Imaging [1].

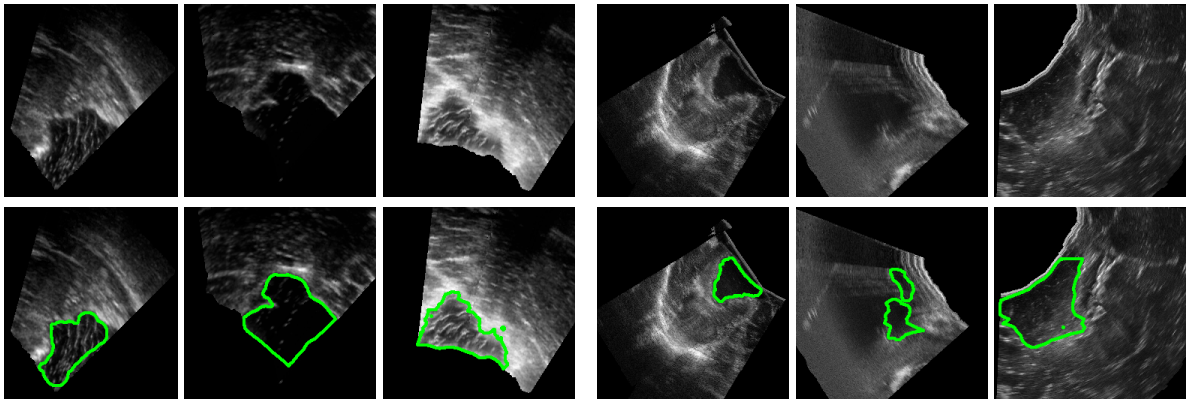


Figure 3.1 – Ground truth segmentations of resection cavities (BITE and RESECT).

3.2 Methods

3.2.1 Data and preprocessing

This study uses the iUS volumes from the BITE [136] and RESECT [137] databases (see sections 2.4.1 and 2.4.2). Since this study focuses on resection cavity, only post-resection volumes were considered (ie. post-resection volumes for BITE, and during and after resection volumes for RESECT).

It is worth noting that BITE and RESECT differ not only in the US equipment used but also in the operative strategy. RESECT is more recent than BITE, thus it uses more recent US probes that generate better images. RESECT iUS volumes were acquired such that they cover the whole tumor and/or resection cavity, whereas BITE has several acquisitions which may not individually cover the whole resection cavity. These are substantial differences to consider, especially with deep learning techniques.

No ground truth segmentations of the resection cavity was available for neither datasets at the time this study was conducted. Thus, two raters manually delineated the resection cavity in both datasets (see figure 3.1).

The RESECT cases were first segmented by myself (François-Xavier Carton). Ten cases were then segmented again, both by myself for intra-rater variability (about 1 month later), and by Matthieu Chabanas, for inter-rater variability. Dice score between the initial segmentations of the two raters are shown in figure 3.2. While both the mean intra- and inter-rater Dice score is 0.89, the variance is slightly higher for inter-rater variability. There was a high agreement for all cases, the differences being in the areas either lacking a clear contour or being difficult to interpret. These areas were then corrected using the guidance of a neu-

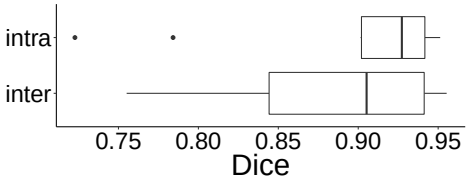
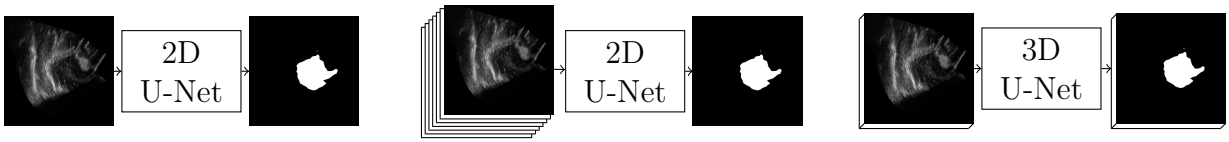


Figure 3.2 – Intra- and inter-rater variablity for initial RESECT resection cavity segmentations.



(a) 2D-1 **(b)** 2D-7 **(c)** 3D

Figure 3.3 – Representation of the three implemented segmentation models.

rosurgeon who reviewed the segmentation of all RESECT cases. The segmentations were edited if needed following the directions of the neurosurgeon. These final segmentations are the one used in this study.

The BITE volumes were segmented by the two raters, each segmenting half of the cases. The segmentations were then reviewed by the two raters together, and edited accordingly.

All volumes were pre-processed by normalizing the intensities. The normalization consists in subtracting the mean and dividing by the standard deviation for each volume.

During training, data was augmented with random transformations, to prevent overfitting. All training volumes were transformed at the beginning of every epoch with the affine, grid, and scale transformations. Each of the three types of transformations was applied with a probability of 0.5. Affine transformations consisted in a translation of up to 16 voxels in a random direction and a rotation around a random axis with an angle between -10 and 10 degrees. Grid transformations were perturbations of the volume sampling grid, computed on a $32 \times 32 \times 32$ grid for which a normal displacement was added to the regular coordinates. Scale transformations used a single scale factor for all dimensions, ranging from 0.75 to 1.25.

3.2.2 Model architecture and training parameters

We compare three versions of a U-Net-like [110] network architecture (figure 3.3): 2D models with a single input slice (named 2D-1) and with seven adjacent input slices (named 2D-7), and a 3D model (named 3D).

The input of the 2D models are $c_i \times 256 \times 256$ patches, where the number of slices c_i is one or seven. The slices are extracted from the original volumes by sampling all slices in one direction (sampling direction). Since the original volumes are larger than the input size, the slices also need to be adapted to fit the input size. To this end, three sampling methods are implemented and compared (figure 3.4): downsampling (DS) with linear interpolation, extracting patches with a sliding window (SW) with a stride of 64 voxels, and cropping to a region of interest (ROI) that is estimated using the result from the DS method. A segmentation volume is reconstructed from concatenating the output slices on the sampling direction. The slices are upsampled for the DS method, averaged over the patches for the SW method, or zero-padded for the ROI method.

The input of the 3D model are $128 \times 128 \times 128$ patches, which are extracted with a sliding window. During training, patches are extracted from a $256 \times 256 \times 256$ region of interest centered around the resection cavity, to avoid sampling too many patches containing no resection cavity voxels. The patches are extracted with a stride of 32 voxels. During testing, the sliding window covers the whole volume, with a stride of 64 voxels.

All models (2D and 3D) have the same network architecture, except that the 3D model has 3D convolutions and max-pooling operations rather than 2D. The architecture is based on the original U-Net [110], except that convolutions are padded, the layer sizes are different (as noted before), and the last layer activation is sigmoid instead of soft-max. The output layer values are between 0 and 1, and are then converted to binary masks with a threshold of 0.5. Only the largest connected component is kept in the final segmentation mask, as to remove possible small components around the resection cavity. Figure 3.5 shows an overview of the network architecture.

The Dice loss function (equation 2.4) is used for optimization of the model parameters, to mitigate class imbalance [113]. Other loss functions for segmentation, such as binary cross-entropy (equation 2.2) and weighted binary cross-entropy (equation 2.3), had convergence issues and did not yield better results, and thus were not considered for this study. Models trained with the binary cross-entropy loss tended to converge to a solution that would always return empty segmentations (null solution). This is due to class imbalance, as there are about 95% of background voxels. Weighted binary cross-entropy mitigates this issue by introducing class weights w_b for the background voxels and $w_f = 1 - w_b$ for foreground voxels. While this avoided convergence to a null solution, the lower weight w_b increased false positives. Such increase was observed with $w_b = 0.25$ and $w_b = 0.05$. Since the Dice loss function have neither of the aforementioned issues, it is the loss function that was retained in this work.

All models were trained using the Adam optimizer [107], with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and with a learning rate of 10^{-5}. The 2D-1 and 2D-7 models were trained for 100 epochs, while the 3D model was trained for 20 epochs. The best epoch was then selected as the one that had the minimal loss function over the validation set.

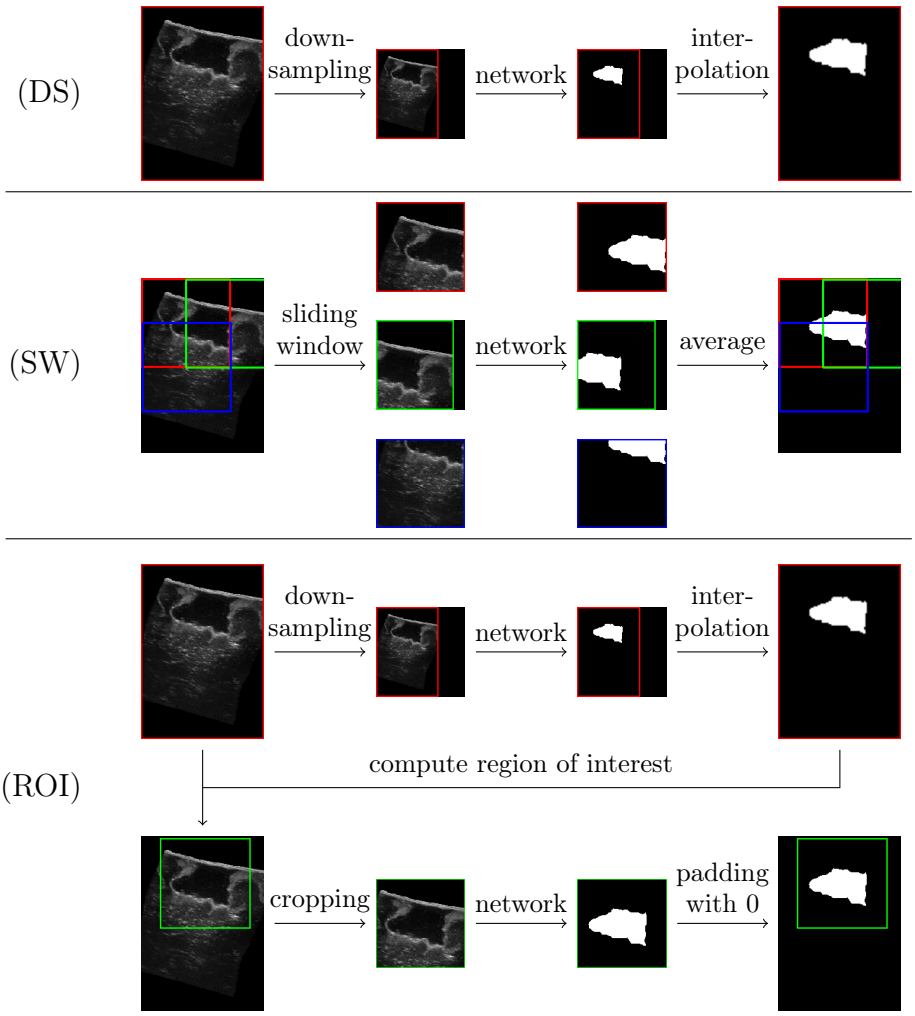
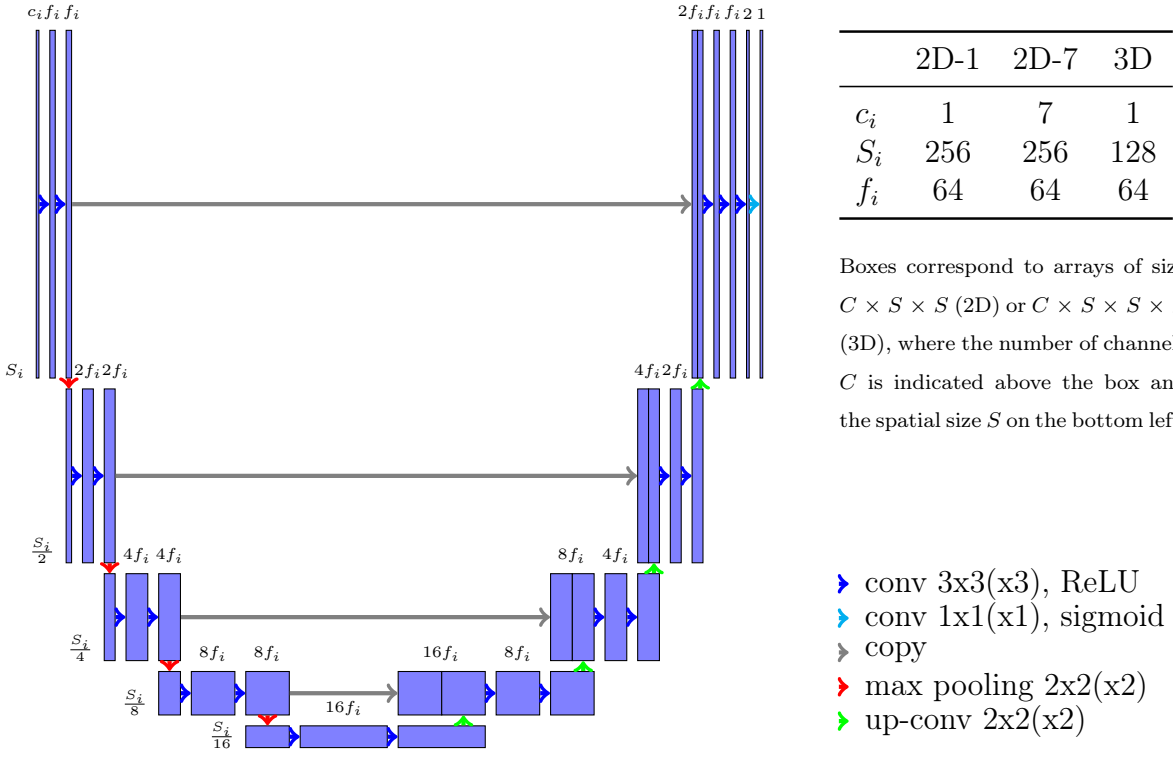


Figure 3.4 – 2D sampling methods.



	2D-1	2D-7	3D
c_i	1	7	1
S_i	256	256	128
f_i	64	64	64

Boxes correspond to arrays of size $C \times S \times S$ (2D) or $C \times S \times S \times S$ (3D), where the number of channels C is indicated above the box and the spatial size S on the bottom left.

Figure 3.5 – U-Net architecture.

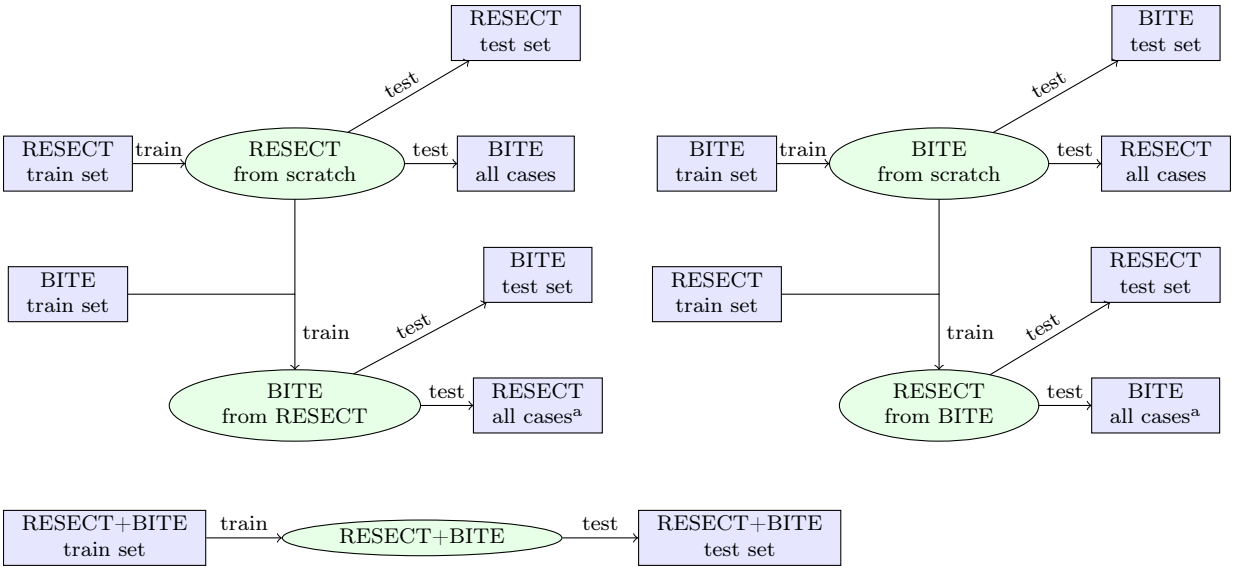
3.2.3 Training strategies and validation studies

We perform cross-validation to evaluate the robustness of the trained models. For RESECT, four folds are used with 23 training volumes, 4 validation volumes, and 10 test volumes. For BITE, two folds are used with 8 training volumes, 1 validation volume, and 4 test volumes.

We first train independent models on a single dataset. These models are evaluated on the test set, as well as on the other dataset, to assess whether generalization from one dataset to another is possible. For example, we evaluate the models trained with the RESECT training set on all BITE cases in addition to the RESECT testing set.

The models are then fine-tuned using the training set of the other dataset. The training is resumed starting with the weights at the selected best epoch of the initial training, using the other dataset for the fine-tuning. As such the models trained with RESECT were fine-tuned with BITE. All parameters are always updated, ie. no layers are frozen during the fine-tuning training.

Finally, a single model is trained with both dataset, with volumes from both datasets split into training, validation, and test sets. Figure 3.6 shows flowcharts with the different



a. Here we show that the model "forgot" about cases it was first trained with, leading to substancially worse results. This is the reason we included all cases, including the ones in the previous training set.

Figure 3.6 – Training and fine-tuning strategies with the two datasets.

training and fine-tuning strategies that were evaluated in this study.

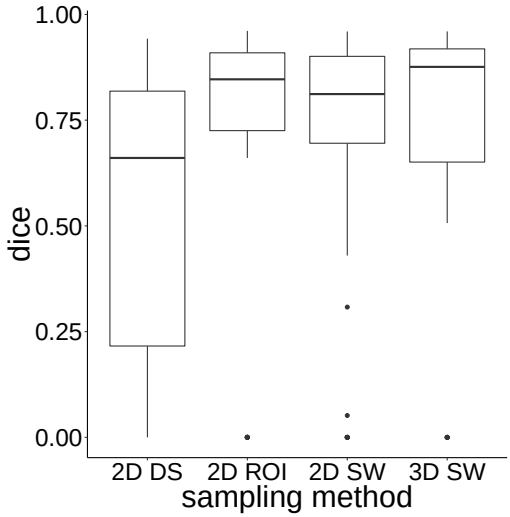
The three sampling methods (DS, SW, and ROI) are compared based on the results of the 2D-1 and 2D-7 trained on RESECT. These results are the segmentation outputs for the RESECT test cases on the four folds chosen for cross-validation. Then the three model architectures (2D-1, 2D-7, and 3D) trained on RESECT are compared. Finally, we compare the five training strategies for the two datasets: training from scratch (BITE and RESECT), fine-tuning (BITE and RESECT), and using both datasets.

3.3 Results and discussion

3.3.1 Sampling methods

In this section, the three 2D sampling methods are compared. 2D SW and ROI have the best results, with a mean Dice score of 0.73 and 0.74, respectively. Figure 3.7 shows the Dice scores and several other metrics for the three 2D sampling methods, as well as the 3D SW model.

The downsampling method have substantially worse results than the SW and ROI methods. There are both under-segmentation (figure 3.8a) and over-segmentation (figure 3.8b)



	2D DS	2D ROI	2D SW	3D SW
mean DSC	0.54	0.74	0.73	0.72
median DSC	0.66	0.85	0.81	0.88
sensitivity	0.46	0.71	0.69	0.68
specificity	0.999	0.999	0.999	0.999
precision	0.72	0.79	0.82	0.81
mean ΔV^*	35.9%	16.6%	21.6%	20.4%
median ΔV^*	29.2%	12.5%	14.8%	11.3%

*volume difference percentage, excluding volumes with no overlap with the GT

Figure 3.7 – Evaluation metrics for the three 2D sampling methods and for the 3D SW method, for the models trained with RESECT.

errors. There are 11 volumes out of 37 for which the output segmentation with the DS method have no overlap with the ground truth segmentation (figure 3.8c). In contrast, there are only 4 such volumes with the SW and ROI method. These correspond to resection cavities with the smallest volumes in the dataset. All but two of them have resection cavities smaller than the median resection cavity volume (figure 3.9). Thus, these errors are likely due to the downsampling, which further reduced the size of the resection cavities.

Overall, the SW and ROI methods have similar results. The ROI method have better segmentations than the SW method in a few cases, as it did not suffer from patch errors (figures 3.8d and 3.8e). These patch errors correspond to adjacent patches for which the resection cavity is missing in one of the patches. This could have been mitigated by using more patches (ie. increasing the overlap between patches by decreasing the sliding window stride), at the expense of longer computing times. On the other hand, the SW method had fewer segmentations with no overlap with the ground truth (figure 3.8f). The SW may be more reliable than the ROI because it does not depend on the output of the DS model, and because the sliding window covers the volumes entirely while the ROI is restricted to the selected region. The SW method have longer runtimes than the ROI method, because the model is evaluated on several patches, in contrast to the ROI method which evaluates only two models (DS and ROI). On a NVIDIA® GeForce GTX TITAN X, the SW method takes about 1 minute, whereas the ROI method takes about 15 seconds. Thus, one approach or the other may be preferred depending on accuracy and time constraints.

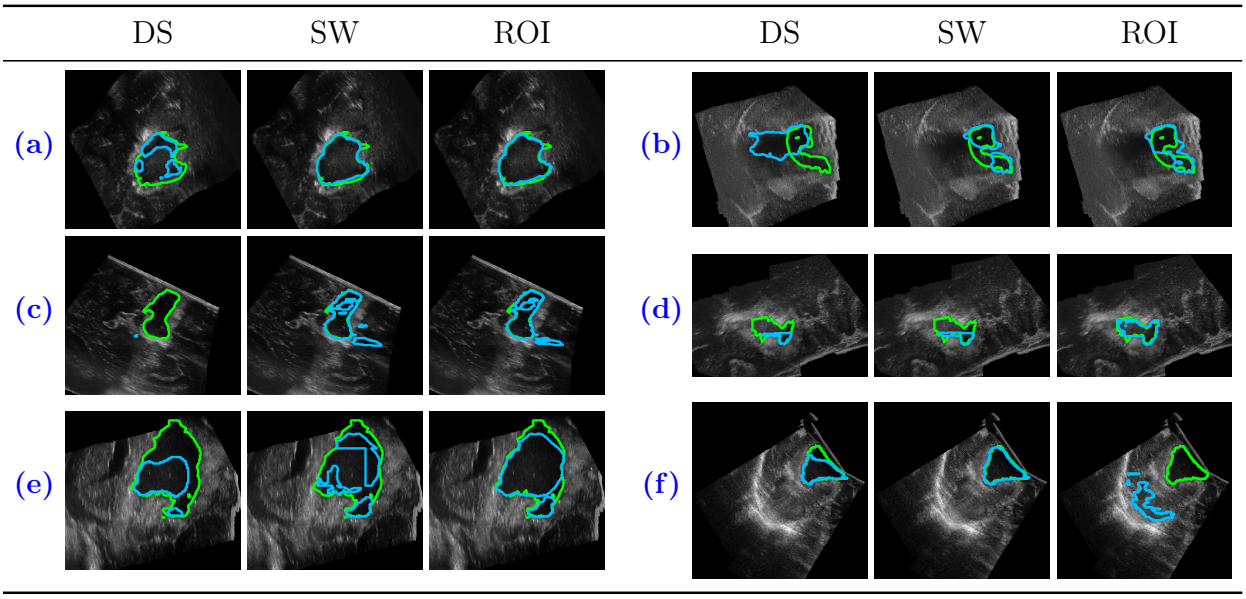


Figure 3.8 – Differences between the three 2D sampling methods. Green: ground truth segmentation. Blue: output segmentation.

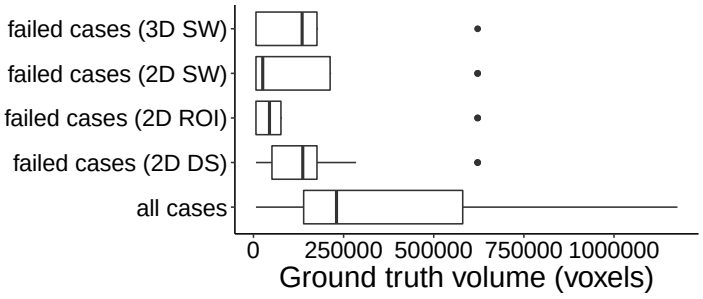
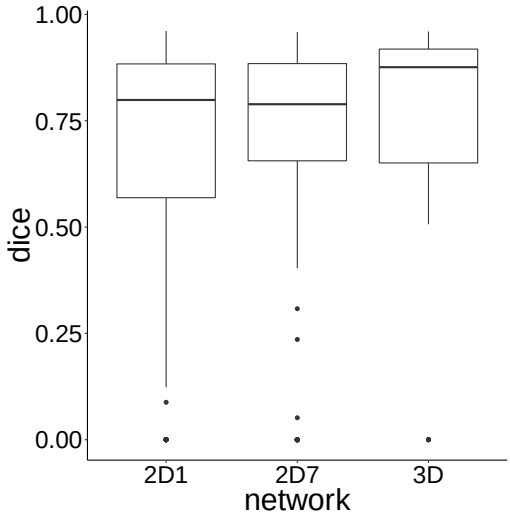


Figure 3.9 – Volumes of the resection cavity in all and failed volumes (volumes for which the model output segmentation have no overlap with the ground truth segmentation).



	2D1	2D7	3D
mean DSC	0.67	0.67	0.72
median DSC	0.80	0.79	0.88
sensitivity	0.62	0.62	0.68
specificity	0.999	0.999	0.999
precision	0.78	0.78	0.81
mean ΔV^*	24.5%	23.9%	20.4%
median ΔV^*	17.2%	17.1%	11.3%

*volume difference percentage, excluding volumes with no overlap with the GT

Figure 3.10 – Evaluation metrics for the three network architectures 2D-1, 2D-7, and 3D, for the models trained with RESECT.

3.3.2 Network architecture

Overall, the 3D model had better results than the 2D models, with a mean Dice score of 0.72 for 3D and 0.67 for 2D. 2D-7 had slightly better results than 2D-1, mainly with the DS method. Figure 3.10 shows the Dice scores and other metrics for the three architectures.

While the results of the 3D model are better than the results of 2D models, the runtime was substantially longer due to the large number of patches that are processed for one volume. Since the 3D patch size is smaller (128 instead of 256) due to memory constraints, more patches are needed to cover one volume. The average runtime of the 3D model to evaluate one case was about 5 minutes (1.5 second per patch), on a NVIDIA® GeForce GTX TITAN X. This may be reduced by decreasing the amount of patches, thus reducing the overlap between patches, which may decrease accuracy. Thus, 2D models may be favored in clinical application thanks to their shorter runtime.

2D-7 have slightly better results than 2D-1, however this is most visible with the DS method. The differences between 2D-1 and 2D-7 are minimal for the SW and ROI methods. With the DS method, the segmentations are better with 2D-7 in the areas for which the resection cavity was not labelled, which is often at boundaries and where blood produced artifacts in the US signal.

3.3.3 Datasets and training strategies

This section discusses the results of the different training and fine-tuning strategies, presented in section 3.2.3 and represented in figure 3.6. Results for models trained with one dataset from scratch and for fine-tuned models are shown on figure 3.11, and results for the model trained on both datasets are shown on figure 3.12.

3.3.3.1 Models trained on a single dataset

With the best method (3D), the models trained on the RESECT (respectively BITE) training set has a mean Dice score of 0.72 (respectively 0.75) over the corresponding test set. The median Dice score is 0.88 for RESECT and 0.81 for BITE. While the were four cases with no overlap with the ground truth, the remaining ones were successfully segmented, with resulting segmentations very close to the ground truth.

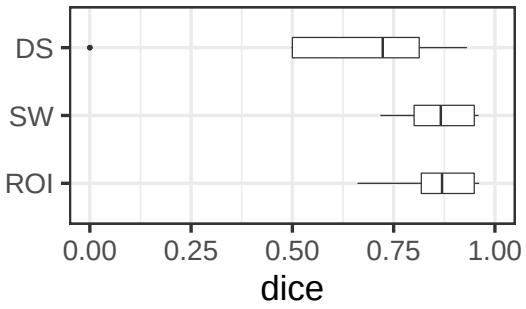
RESECT cases in particular are discussed in this paragraph. The results of one fold obtained with the 3D model are shown in figure 3.13. The model is robust to noise, as shown in case 8d in figure 3.13. It can also be noted that the resection cavity is correctly selected in volumes in which several cavities are present, such as the ventricle in case 26a.

Except with the 2D DS method, the segmentations with no overlap with the ground truth correspond to only four volumes (5d, 14d, 15a, 18d) regardless of the model. The resection cavity in volumes 14d and 15a was among the smallest in RESECT, and thus difficult to locate. In particular, the 2D DS method is the most sensitive to small volumes. In volume 5d, the *dura matter* is visible and appears as a high intensity signal (figure 3.1), which may confuse the model. In volume 18d, the bottom of the resection cavity have a different signal, possibly due to remaining blood, which may explain why it was not considered as the resection cavity. A larger training dataset would improve the training, allowing better generalization and preventing such errors.

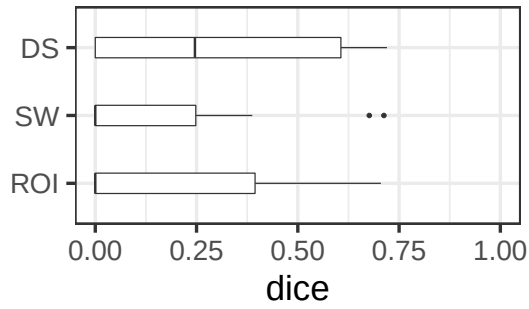
BITE results are similar to the RESECT ones. There are no failed case with the SW and ROI methods, and only one with the DS method.

3.3.3.2 Generalization to the other dataset and impact of fine-tuning

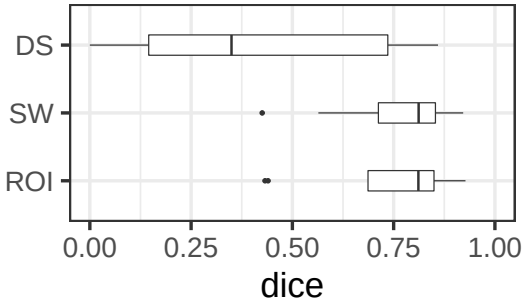
The models trained on RESECT only and BITE only are also evaluated on all cases of the other dataset (not seen for training), to determine whether these models can generalize to other datasets. This is of interest since the two datasets differ substantially in acquisition



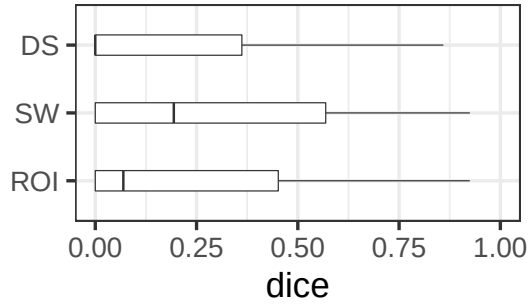
(a) Model trained from scratch with RESECT evaluated on RESECT test set.



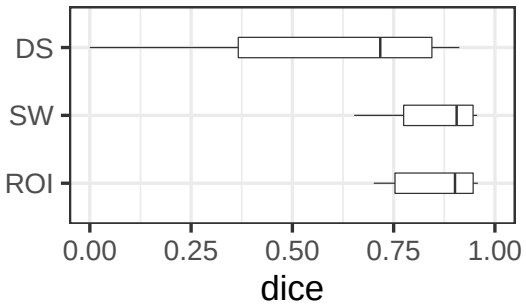
(b) Model trained from scratch with RESECT evaluated on all BITE dataset.



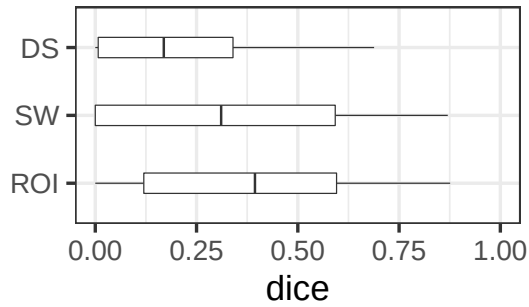
(c) Model trained from scratch with BITE evaluated on BITE test set.



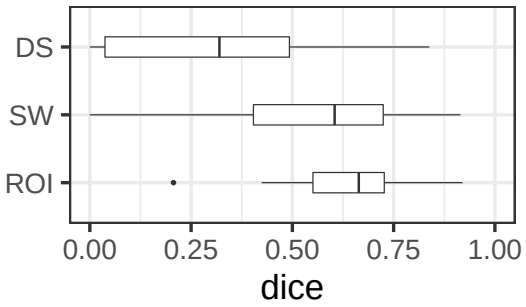
(d) Model trained from scratch with BITE evaluated on all RESECT dataset.



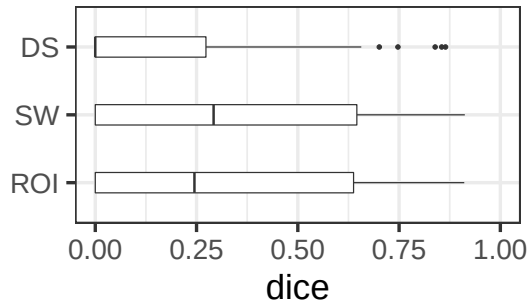
(e) Model trained with BITE then fine-tuned with RESECT evaluated on RESECT test set.



(f) Model trained with BITE then fine-tuned with RESECT evaluated on all BITE dataset.



(g) Model trained with RESECT then fine-tuned with BITE evaluated on BITE test set.



(h) Model trained with RESECT then fine-tuned with BITE evaluated on all RESECT.

Figure 3.11 – Dice scores for models trained from scratch and fine-tuned models.

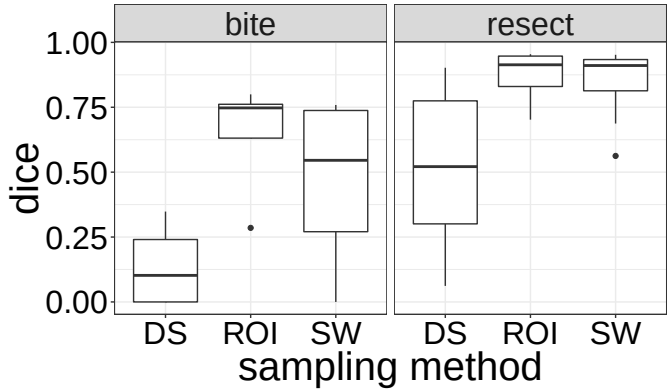


Figure 3.12 – Dice scores for the model trained with both RESECT and BITE. The test set contains both RESECT and BITE volumes, which are shown in seperate plots.

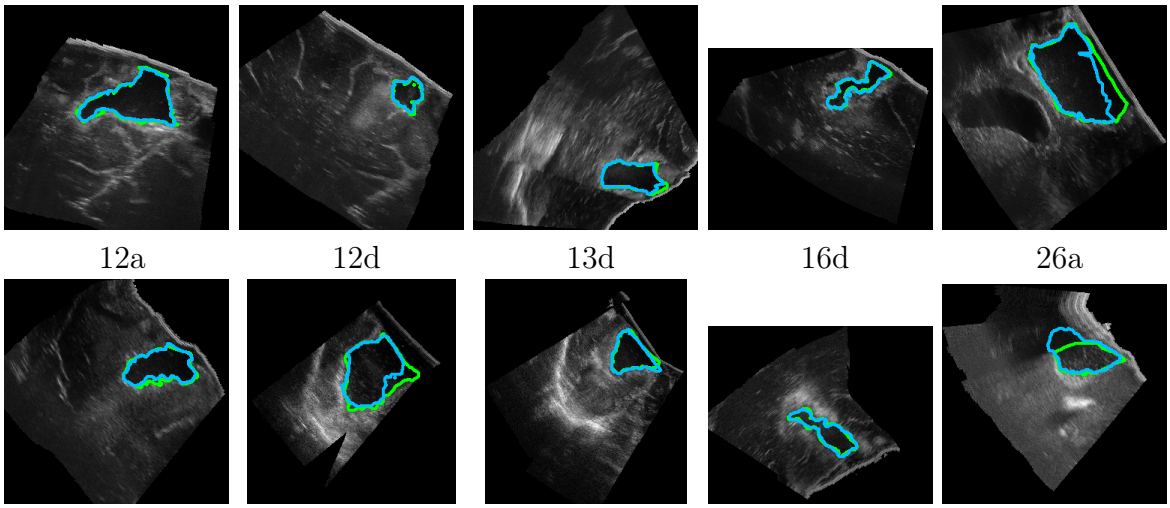


Figure 3.13 – Segmentations obtained with the 3D model for RESECT test set. Green: ground truth segmentation. Blue: model segmentation.

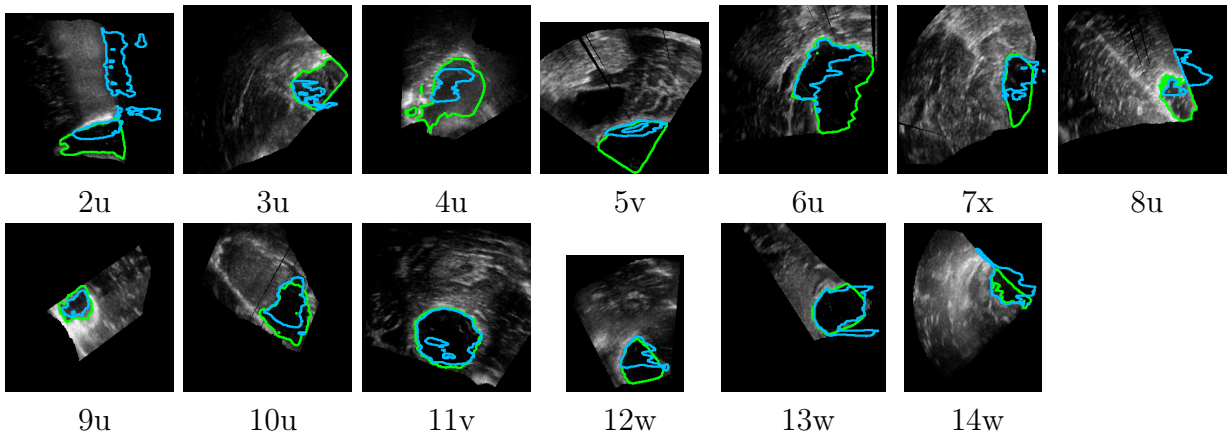


Figure 3.14 – BITE segmentations obtained with the 2D-7 model trained only with RESECT volumes. Green: ground truth segmentation. Blue: model segmentation.

equipment and protocol, as discussed in section 3.2.1. Results for the RESECT model evaluated on BITE are shown on figure 3.11b, while results for the BITE model evaluated on RESECT are shown on figure 3.11d.

While the models do not generalize well, with more cases with no overlap with the ground truth, there are several volumes that are correctly segmented. The BITE segmentations obtained with the 2D-7 model only trained on RESECT are shown in figure 3.14, in which all thirteen volumes are present. Most false negative errors are located on the footprint, which is often not visible on BITE volumes because the probe was inserted directly inside the resection cavity. This is for example the case for volumes 5v and 6u shown in figure 3.14. This is in contrast to RESECT volumes, which all cover the whole resection cavity. The ground truth segmentations includes all the voxels in the field of view in that area, whereas the model segmentations do not include voxels far from the visible boundaries of the resection cavities. False positives errors are mainly located outside the field of view, and thus could be easily removed with post-processing. Volumes for which there is no or little overlap with the ground truth (5v, 7x, and 14v) are all volumes covering only part of the resection cavity unlike RESECT. With the differences in acquisition methods between RESECT and BITE, it is not surprising that a model trained with one dataset does not generalize well to another dataset. The results are nonetheless promising, given that several volumes of the other dataset are correctly segmented although no volumes from that dataset were used in the training process. With such substantial differences between the two datasets, dataset-specific training remains the best solution.

With the fine-tuned models, good results are obtained on the dataset used for fine-tuning, but the results on the original dataset are worse than before the fine-tuning (see figure 3.11e compared to figure 3.11f, and figure 3.11g compared to figure 3.11h). This shows that the model somewhat "forgot" the first dataset it was trained with. This is likely due to the

differences between the two datasets being too important, so that the training focused on the current dataset in a way that is incompatible with the previous one. The fine-tuned models still have slightly better results than the models trained from scratch, on the other dataset. This shows that not all was forgotten about the first dataset. It appears that training a model to have good results on both datasets is difficult. This again shows that dataset-specific training is best suited to datasets that are substantially different.

Generalization from one dataset to the other has not proven successful with or without fine-tuning. The best results are obtained with models trained and evaluated on the same dataset. The last option for training a model that generalizes to both datasets is to train the model directly with cases from both datasets. This is investigated in the following paragraph.

3.3.3.3 Model trained and evaluated on both datasets

This paragraph discusses the model trained on both datasets (bottom model in figure 3.6), whose results are shown in figure 3.12. While the results are better than the ones for models trained on one dataset and evaluated on the other (with or without fine-tuning), the model does not outperform models trained on a single dataset and evaluated on that same dataset. The obtained model is however the model that generalizes the best to both datasets combined. The model have better results on RESECT than on BITE, despite having an equal number of volumes from each dataset in the training set.

These results confirm that training a specific model for each dataset yields the best results. This is most likely because the datasets are too different.

3.3.4 Comparison with intra- and inter-rater variability

The Dice scores of the models with the ground truth can be compared to the Dice scores between the manual segmentations of two raters. Figure 3.15 shows boxplots as such Dice scores, comparing the model with intra- and inter-rater variability. While there is a larger variance for the Dice scores of the model than for the ones of the raters, we observe that many cases have Dice scores close to rater variability. In particular, the median Dice scores are very close.

3.3.5 Comparison with other methods

At the time that these results were first published, no comparable study was conducted. Since then, Canalini et al. proposed a similar method for use with registration of the iUS before resection with the iUS during or after resection [67]. They train and evaluate their

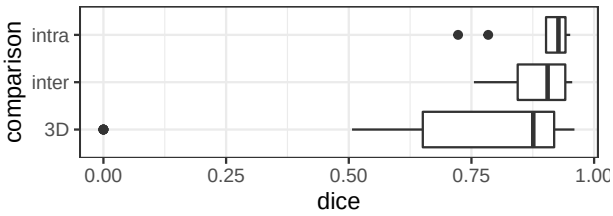


Figure 3.15 – Dice score of best model (3D) with ground truth (40 test cases), compared to intra- and inter-rater Dice scores (10 cases).

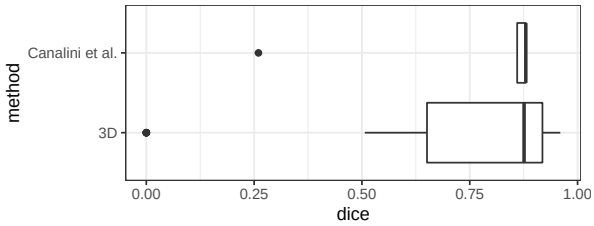


Figure 3.16 – Dice score of best model (3D) with ground truth (40 test cases), compared to Canalini et al. [67] (5 test cases).

model on RESECT, and report a mean Dice score of 0.84 over the 27 cases. However, this cannot be interpreted reliably, since it is computed on all cases, including training cases. They use a test set with five volumes, over which the mean Dice score is 0.75, which is very similar to what we report in this study.

Figure 3.16 shows these Dice scores, compared with the ones with obtained with our 3D model. There is a much smaller variance for the Dice scores of Canalini et al.; however, their test set only have 5 cases while we report our results for 40 cases (4 folds). They have an outlier case (volume 18d), which is the same case as one of the outliers we reported. This highlights how difficult segmenting this case is, and the need for robust segmentation methods.

3.4 Conclusions

In this chapter, we proposed an automatic method to segment the resection cavity in iUS of the brain. This was the first work published on this topic. We explored different variations of the methods (2D and 3D models, with different sampling methods) and evaluated generalization of the models between two datasets with different acquisition protocols. We analyzed the results extensively and discussed the benefits and drawbacks of the method variations that we proposed.

We found that better results were obtained with 3D models, but with a higher runtime. We also showed that for the two datasets used in this study, training specific models independently yields better results. This is due to the substantial differences between the two datasets, and in particular because the resection cavity is only partially visible in many BITE volumes.

Overall, the automatic segmentations obtained with the models are close to the manual ground truth segmentations. The obtained Dice scores for many cases are comparable to intra- and inter-rater variability. At the time we conducted this study, there was no other

study on segmentation of the resection cavity in brain US images, to the best of our knowledge. Recently, a similar study of Canalini et al. [67] showed that comparable results were obtained on RESECT volumes.

While for most cases the model segmentation was close to the ground truth, there are a few outlier cases. We expect that larger datasets would improve the generalizability of the models, as the two datasets are very small for deep learning techniques. Promising results were shown nonetheless, which motivates further research with larger datasets. In particular, the segmentations obtained with the models are accurate enough to be used in pMR-iUS registration, so that tissue displacement due to tissue resection can be modelled.

Chapter 4

Segmentation of brain tumors in iUS images

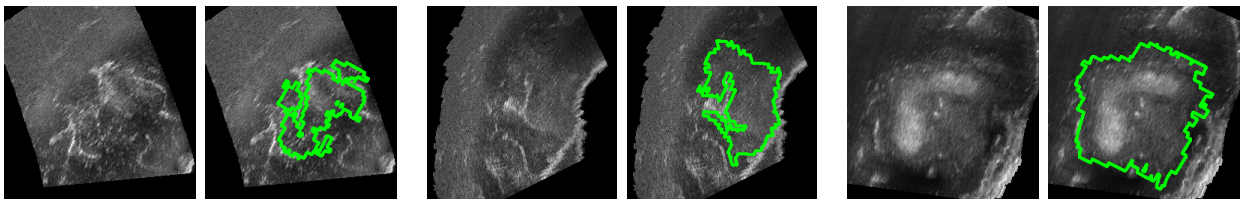
4.1 Introduction

US imaging can be used during brain tumor resection surgery to visualize tumor tissue and the surrounding anatomical structures. This can help surgeons to decide whether tissue resection is complete. However, US images have several limitations (see section 1.5.5) including noise, artifacts, and limited field of view. The contrast between tumor and healthy tissue is usually poor, and thus delineating the tumor accurately is challenging. In addition, it is difficult to use the pMR to better understand the iUS because of brain shift (see section 1.6).

One approach to mitigate these problems is to register the pMR with intraoperative data (see section 2.1). While a challenging task, it has been widely studied and several methods have been proposed. Another approach is to analyze iUS images directly, as an alternative or complement to registration. One benefit of such approach is that tracking the US probe would not be required, which would simplify considerably the surgical setup. In this chapter, we propose a method to segment tumor tissue directly in iUS images. We focused on pre-resection iUS images.

While brain tumor segmentation in MR images has been widely studied (see section 2.2.1 and reference [70]), few methods have been proposed for iUS images [105, 106]. These two methods [105, 106] are for high grade tumors only. To the best of our knowledge, no segmentation method for low grade brain tumors in iUS has been proposed. Low grade tumors usually have a lower contrast than high grade tumors, making them even more difficult to delineate. Other iUS segmentation methods focused on more salient structures like the *falx cerebri* and *sulci* [38, 39], or resection cavities [1, 67] (chapter 3).

In this chapter, we present a deep learning model for segmentation of brain tumors in iUS images acquired just after dura opening and before tissue resection. Preliminary results for this study were presented at the SPIE 2020 medical imaging conference [4].



RESECT 1*b* RESECT 2*b* RESECT 12*b*

Figure 4.1 – RESECT iUS images and tumor ground truth segmentations from [15].

4.2 Methods

4.2.1 Data and preprocessing

In this study, we have only used data from RESECT [137], a database a 23 patients with low grade gliomas. Section 2.4.2 presents this dataset with more details. Ground truth segmentations of tumor tissues were manually delineated in 17 volumes in both the pMR and pre-resection iUS for a previous study from SINTEF [15]. These segmentations are used in this work for training and evaluation of the models. Figure 4.1 shows three examples of ground truth segmentations.

The volumes are preprocessed by normalizing them the same way as in chapter 3. During training, data was augmented by using the affine, grid, and scale transformations described in chapter 3.

4.2.2 Model architecture and training parameters

Five models based on the U-Net models presented in chapter 3 are trained to segment tumor tissue, as illustrated figure 4.2.

The first two models are 2D models 2D-1 and 2D-9, whose input consists of one and nine adjacent slices, respectively. These two models are the same ones as the resection cavity models in chapter 3, except than 2D-9 have nine slices of context, whereas 2D-7 had seven. The size of input slices for these models is 256×256, and a 2D sliding window with a stride of 64 voxels was used, covering the entire original volumes.

The other three models are 3D models similar to the 3D model presented in chapter 3, with $128 \times 128 \times 128$ input patches and a 3D sliding window with a stride of 32 voxels. The three 3D models are referred to as 3D-1, 3D-2, and 3D-3.

The input of 3D-1 is the iUS volume only. It is thus similar to the 3D model of chapter 3,

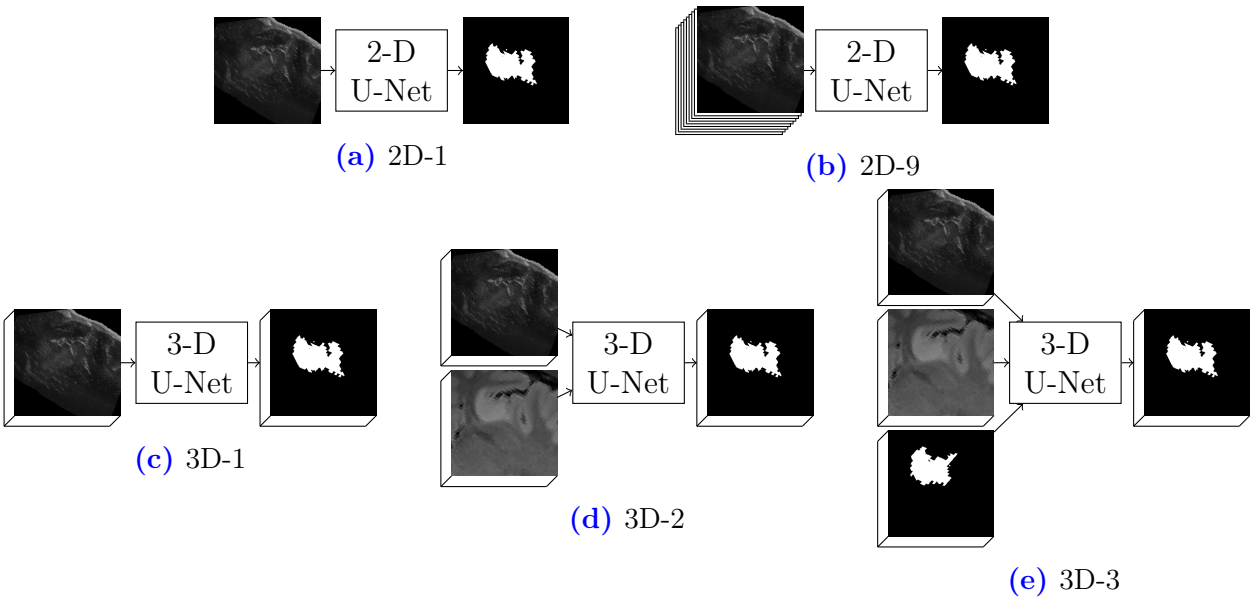


Figure 4.2 – Representation of the five segmentation models.

with a single 3D input channel.

The input of 3D-2 consists of both the iUS and pMR, in two separate 3D input channels. The signal is very different in the two volumes, not only due to different imaging modalities, but also because of brain shift (see section 1.6 and figure 4.3) which causes tissue displacement between the pMR and iUS. This is especially important for tumor delineation, because the tumor location will be different in the two volumes. In addition, it has been shown than the extent of tumor tissue differs between the pMR and iUS [15]. We hypothesised nonetheless that the pMR is valuable input for the segmentation model, since the volumes are still somewhat consistent and considerably helpful to humans for interpretation of the iUS volumes.

3D-3 is like 3D-2, with the addition of a third input channel with the pMR tumor segmentation. We assume that segmentation of the pMR is available, as several automatic methods have been proposed, including for delineation of the tumor. In this study, we used the MRI tumor segmentations from [15]. Although the pMR segmentation also do not account for the brain deformation compared to the iUS, the model might still benefit from that additional information to improve the segmentation of the iUS volume.

The pMR and iUS in RESECT are co-registered, thus can be provided in the same space to the segmentation model (yet brain shift is not corrected, as figure 4.3 shows). In the case that the US probe is not tracked, this method could still be used by estimating the initial transformation [138].

The Dice loss function (equation 2.4) is optimized with the Adam optimizer to train the

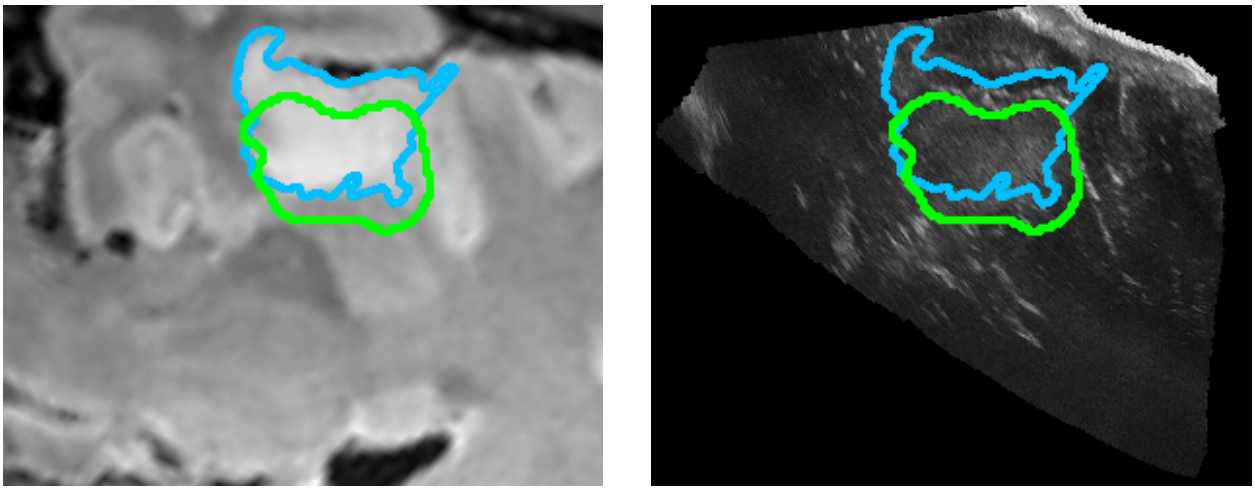


Figure 4.3 – Brain shift in a co-registered pMR (left) and pre-resection iUS (right). Blue contour: pMR tumor segmentation; green contour: iUS tumor segmentation.

models, as in chapter 3.

4.2.3 Validation studies

We use cross-validation with three folds of 10 training volumes, 2 validation volumes and 5 test volumes. We first compare 2D-1, 2D-9, and 3D-1, with the results on the three folds (15 volumes). We then evaluate the effect of adding the pMR to the model inputs by comparing 3D-1, 3D-2, and 3D-3 on fold 1 (5 volumes).

4.3 Results and discussion

4.3.1 Comparison of 2D and 3D models

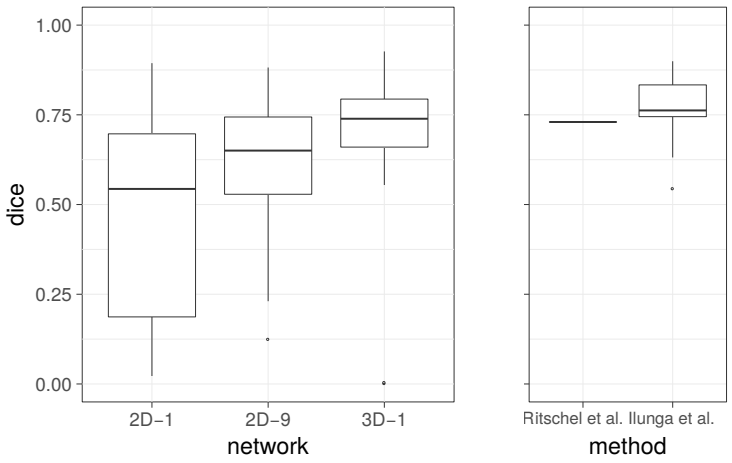
Both qualitatively and quantitatively (with Dice scores), 3D-1 have better results than the 2D models, and 2D-9 have better results than 2D-1. This demonstrates the importance of surrounding context for tumor tissue segmentation. Wilcoxon signed rank tests between 2D-1 and 3D-1 (p-value of 0.011) and between 2D-9 and 3D-1 (p-value of 0.004) show statistical significance between 2D and 3D models. Dice scores for 2D-1, 2D-9, and 3D-1 on the three folds are presented in the first three rows of table 4.1. Boxplots of these Dice scores are plotted in figure 4.4a.

While 3D-1 have the best results, the runtime was longer (5 minutes on a NVIDIA®

Method	Cases	Mean Dice	Median Dice	Precision	Sensitivity	Specificity
2D-1	15[a]	0.47 (0.28)	0.54	0.68 (0.31)	0.44 (0.28)	0.99 (0.01)
2D-9	15[a]	0.61 (0.21)	0.65	0.62 (0.25)	0.70 (0.22)	0.98 (0.02)
3D-1	15[a]	0.65 (0.28)	0.74	0.65 (0.30)	0.68 (0.29)	0.99 (0.01)
3D-1	5[b]	0.65 (0.37)	0.76	0.73 (0.41)	0.59 (0.35)	0.99 (0.01)
3D-2	5[b]	0.67 (0.17)	0.69	0.67 (0.17)	0.68 (0.21)	0.99 (0.01)
3D-3	5[b]	0.56 (0.27)	0.60	0.55 (0.21)	0.64 (0.34)	0.99 (0.01)
R. [105][c]	13	0.73[d]		0.71 (0.13)	0.76 (0.15)	0.94 (0.05)
I. [106][c]	14	0.77 (0.09)				

a. all three folds
b. first fold only
c. for high grade tumors, in contrast to this study
d. computed from precision and sensitivity values

Table 4.1 – Evaluation metrics for tumor segmentation methods (models presented in this study and two previous works).



(a) with our three models (2D-1, 2D-9, and 3D-1) on 15 low grade tumor cases

(b) with existing methods on 13 (R. [105]) and 14 (I. [106]) high grade tumor cases

Figure 4.4 – Dice scores for tumor segmentation methods.

Fold	Case	Dice score	GT volume	Intra-rater difference	Prediction difference
1	1	0.00	6.1 mL	3.5 mL (57%)	non overlapping
1	2	0.76	6.2 mL	1.7 mL (27%)	1.4 mL (22%)
1	12	0.81	12.9 mL	2.7 mL (20%)	3.7 mL (28%)
1	16	0.74	6.5 mL	1.6 mL (24%)	1.8 mL (27%)
1	19	0.93	23.2 mL	-1.6 mL (-6%)	-1.5 mL (-6%)
2	6	0.66	12.5 mL	3.2 mL (25%)	-4.0 mL (-31%)
2	8	0.70	33.9 mL	0.9 mL (2%)	3.7 mL (10%)
2	15	0.77	13.3 mL	1.5 mL (11%)	-3.6 mL (-27%)
2	16	0.66	6.5 mL	1.6 mL (24%)	-3.4 mL (-51%)
2	18	0.72	23.4 mL	6.5 mL (27%)	1.3 mL (5%)
3	3	0.00	1.1 mL	0.0 mL (0%)	non overlapping
3	5	0.55	21.2 mL	1.9 mL (9%)	-9.4 mL (-44%)
3	7	0.77	21.6 mL	5.9 mL (27%)	-4.9 mL (-22%)
3	21	0.83	97.1 mL	12.2 mL (12%)	-26.1 mL (-26%)
3	23	0.88	19.6 mL	1.1 mL (5%)	-0.6 mL (-3%)

Table 4.2 – Evaluation metrics for the 3D-1 model.

GeForce GTX TITAN X) than the 2D models (about 15 seconds), because of the higher number of patches evaluated to process one volume. The runtime could be reduced by decreasing the number of patches (by increasing the stride of the sliding window). This would however reduce the overlap between patches, and thus may decrease accuracy.

4.3.2 Results with the 3D-1 model

Among the five models trained in this study, 3D-1 is the model with the best results. The model segmentations are very similar to the ground truth, except two outlier cases (1*b* and 3*b*). Excluding these two cases, the mean Dice is 0.75, with Dice scores ranging from 0.55 to 0.93. Table 4.2 presents quantitative results for 3D-1 and figure 4.5 shows the model segmentations for all 15 test cases. In particular, volumes 15*b* and 18*b* are examples of volumes in which the tumor is difficult to segment, yet successfully segmented with the model. In 15*d*, there is a darker area in the tumor whose intensity is similar to healthy tissue, which is correctly labeled as tumor tissue. In 18*d*, a *sulcus* splits the tumor in two components. The model segmentation for this volume matches the ground truth closely in most areas, save for a small under-segmented region on the left.

The two outlier cases are the cases for which the tumor have the smallest volumes. The tumor volumes are 1.1mL (case 3*b*) and 6.1mL (case 1*b*), while the mean tumor volume in

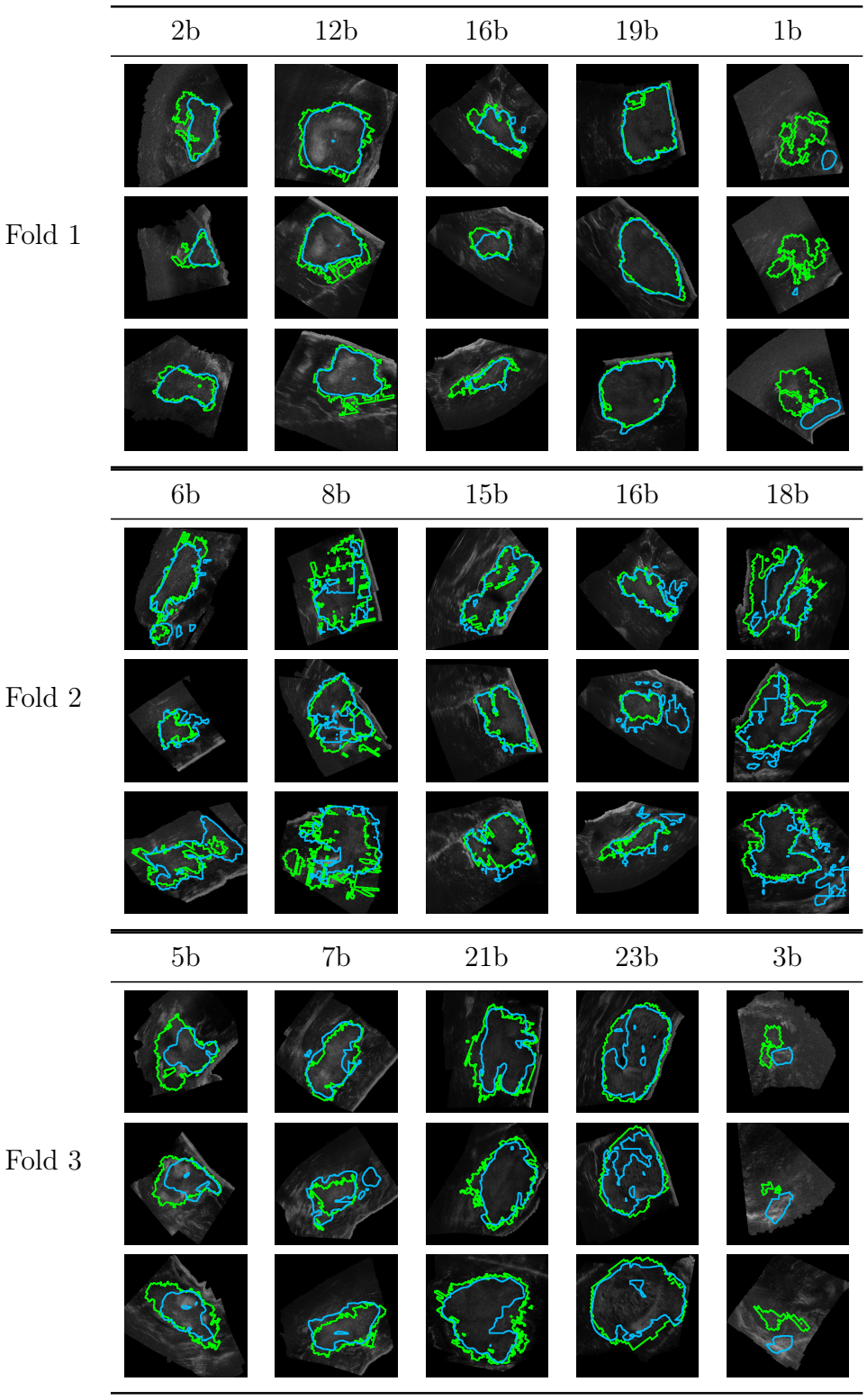


Figure 4.5 – Tumor segmentation with model 3D-1. Green: ground truth. Blue: model segmentation.

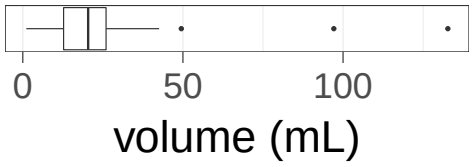


Figure 4.6 – Volume distribution of the 17 ground truth iUS tumor segmentations.

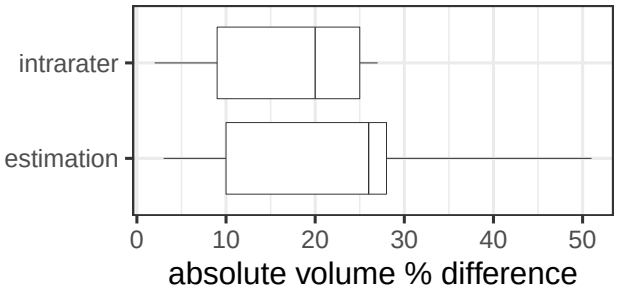


Figure 4.7 – Comparison of intra-rater (top) and ground-truth/model (bottom) tumor volume differences.

the dataset is 28.6mL. The distribution of tumor volumes is plotted in figure 4.6. The tumor is difficult to locate in these two volumes, because of the small tumor size. This is the main challenge associated with these volumes. In addition, there is less contrast between tumor tissue and the surrounding healthy tissue in these volumes. This is because the intensity of the surrounding tissue is brighter, which may be because the tumor is located next to the *tentorium*.

While these cases are inherently difficult to segment because the small tumors are difficult to locate, we hypothesize that the pMR might help localizing the tumor (which is investigated in this study with models 3D-2 and 3D-3, whose results are presented in the next section); and that a multi-class segmentation model may improve the accuracy by differentiating the different anatomical structures and using spatial relationships between them (which is investigated in chapter 5). In addition, we expect that training with a larger dataset will improve generalization of the model and robustness to signal variation such as brighter healthy tissue next to the *tentorium*.

4.3.3 Additional input channel with preoperative MRI

Adding the corresponding pMR volume to the input as a second channel slightly improved the reliability of the network. For case 1, although the error was still large (the Dice score was 0.46), the location of the segmentation for 3D-2 had overlap with the ground truth whereas the result for 3D-1 did not. However, the results were not improved on the other four cases where the tumors were successfully segmented with 3D-1. In some cases, the pMR in the input appeared to distract the network from the tumor location in the iUS (see for example case 12 in figure 4.8, where the estimated segmentation closely matches the intersection between the iUS and pMR tumor location).

Adding the segmentation of the pMR as a third input channel tended to skew the es-

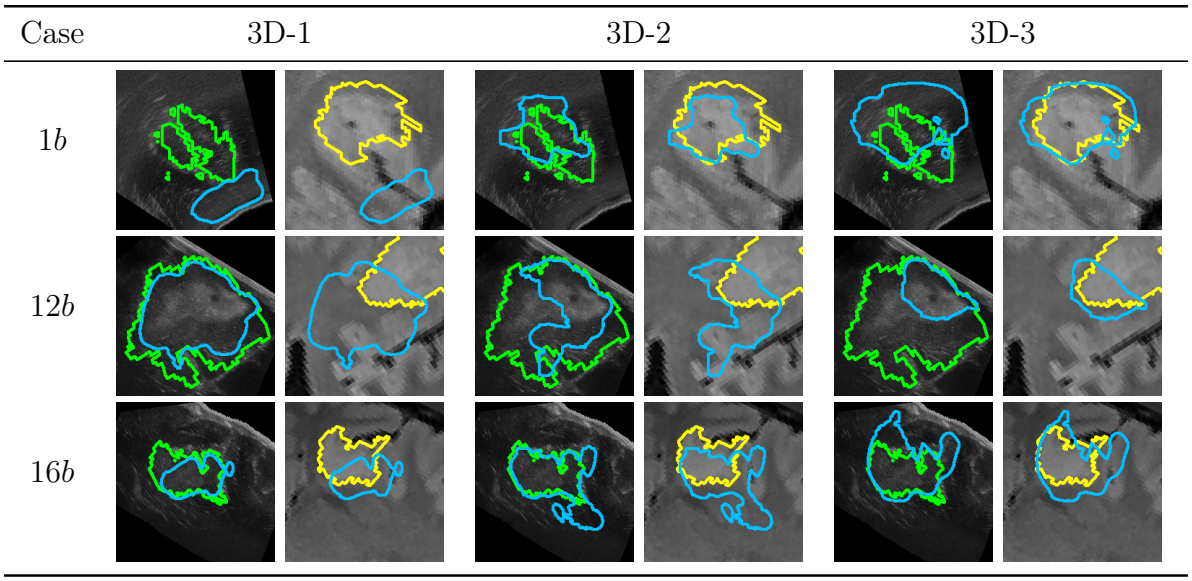


Figure 4.8 – Results with 3D-1 (iUS only), 3D-2 (iUS and pMR), and 3D-3 (iUS, pMR, and pMR segmentation). Green: iUS ground truth. Yellow: pMR ground truth. Blue: model segmentation.

timated segmentations towards the pMR tumor segmentation. This resulted in lower Dice scores between the estimated segmentations and the ground truth iUS segmentations.

As such, it appears that using the pMR image and/or segmentation without registration does not lead to improved results.

4.3.4 Comparison with intra-rater variability

The previous study for which the ground truth segmentation were made [15] reported the intra-rater tumor volume difference. We computed the same volume difference metric between the ground truth segmentations and the segmentations obtained with 3D-1. We excluded the two outliers because there is no overlap between the ground truth and model segmentations, and comparison of the volume difference is meaningless.

The intra-rater and difference between the ground truth for the remaining 13 test cases are shown in table 4.2 and figure 4.7. The absolute percentage difference ranges from 6% to 27% for the intra-rater variability, and from 5% to 59% for the difference between ground truth and model segmentations. The volume difference between the model and ground truth is similar to the intra-rater variability for many cases, especially from fold 1. Also, it is worth noting that the highest intra-rater difference is 57% for volume 1*b*, which is one of the two outliers.

These results are very encouraging, given the difficulty of tumor segmentation in iUS. Except for the two outlier cases, the output segmentations are comparable to the manual segmentations.

4.3.5 Comparison with other methods

Few studies addressed tumor segmentation in iUS images. Ritschel et al. [105] and Ilunga-Mbuyamba et al. [106] proposed methods for high grade tumor segmentation in iUS. We report their results on table 4.1 (last rows) and figure 4.4b. Ritschel et al. evaluated their method on 13 cases, with a mean Dice score of 0.73. Ilunga-Mbuyamba et al. obtained a mean Dice score of 0.77 over 14 cases, with their best method (rigid registration).

These two methods are evaluated on high grade tumors, whereas our models are evaluated on low grade tumors, which are more difficult to segment, because they are substantially less contrasted in iUS images. Nevertheless, the Dice scores obtained with 3D-1 are comparable to the two existing methods, despite this higher complexity of low grade tumor segmentation.

4.4 Conclusions

In this chapter, we proposed and evaluated 2D and 3D segmentation model for low grade brain tumors in iUS volumes. Promising results were obtained, despite the complexity of tumor tissue delineation in ultrasound images. These tumor segmentations are valuable information for surgeons, as well as for other processing methods such as registration.

We show that the 3D models have better results, at the cost of longer computation times. We also evaluated two basic strategies to include pMR information in the model inputs. Although this helps localizing small tumors, and thus improves reliability of the models, the models tend to output segmentations close to the tumor in the pMR, which is not corrected for brain shift.

The obtained model segmentations were very similar to the ground truth segmentations, except for the two outlier cases. Quantitatively, the volume differences for the model and ground truth is close to intra-rater variability, and the obtained median Dice score is 0.72. The results are comparable to previous methods proposed for high grade tumor segmentation. While a very good accuracy is obtained for most cases, the two outlier cases highlight the need of increasing the robustness of the models.

While promising, the results presented in this chapter are preliminary. This is in particular because the dataset size is very small, and especially for training deep learning models. Larger datasets would enable better training of the models, as well as a better validation of

the models and more thorough analysis of the results. Nonetheless, these preliminary results indicate that deep learning may be relevant for iUS processing, including challenging tasks such as tumor tissue segmentation, and motivates further research with larger datasets.

In this chapter, the models were trained and evaluated on pre-resection iUS only. Generalization to post-resection iUS (or iUS during resection) has not been investigated, with or without training with such cases. This would be interesting to evaluate, although the ground truth determination would be even more challenging than in pre-resection iUS.

Chapter 5

Multi-class segmentation of iUS images

5.1 Introduction

Segmentation of brain iUS images is valuable in a number of applications (section 2.2.2). In particular, the segmentation of structures in brain iUS can be used for registration of the pMR with iUS, which is one of the goals of this thesis. Registering the pMR with iUS allows updating the pMR to compensate for the brain shift (section 1.6) and enables navigation of the MR images during surgery to guide surgeons.

Registration methods for brain shift compensation are reviewed in section 2.1. Recent methods using iUS include image-based [23, 24, 36], model-based [16] and structure-based [39, 139] methods. Image-based techniques estimate the mapping between the two images with image intensity, for example by minimizing dissimilarity metrics or by extracting and coupling feature descriptors. Model-based methods estimate the tissue displacement with a biomechanical model of the brain. Structure-based algorithms match the delineation of structures in both images to estimate the mapping between the two images.

Existing structure-based registration methods [39, 139] use segmentations of *sulci*, *falx cerebri* and *tentorium cerebelli*. Nitsch et al. [39] proposed to segment these structures in the iUS with a random forest classifier, and then register the pMR with the iUS by minimizing the local cross correlation on the structures only, by masking the volumes with the segmentation masks. Canalini et al. [139] trained a U-Net [110] neural network to segment the *sulci* and *falx cerebri* in iUS volumes, and then registered iUS volumes at different stages of tissue resection by minimizing the sum of squared differences of the segmentation masks.

Previous iUS segmentation methods have focused on segmenting one structure, thus are single-class segmentation methods. Methods have been proposed for segmenting structures such as resection cavities [1, 67, chapter 3], tumor tissue [4, 105, 106, chapter 4], *falx cerebri* and *sulci* [5, 38, 39].

We hypothesize that multi-class segmentation models may have greater accuracy than

single class models. This is because the locations of the different structures are correlated. In particular, other studies suggested that multi-class segmentation models can lead to better results in other applications [140, 141]. In the case of brain iUS, the resection cavity is located next to tumor tissue in images during resection. Locating the resection cavity is often easier than classifying tumor tissue, thus may help locating tumor tissue, especially for small tumors. This motivated the training of a multi-class segmentation model, which segments the *sulci*, *falx cerebri*, tumor, resection cavity, and ventricles.

In this chapter, we compare single- and multi-class segmentation models for the aforementioned classes. Training the multi-class models is challenging because some classes may be missing on some samples. This further increase class imbalance, as there are far less voxels of the present classes than background voxels. Thus, we implemented several strategies to mitigate these limitations. In deep learning models, this is typically done by adapting data sampling (over- or under-sampling) and/or adapting the loss function. In this case, we use a combination of patches selection and class weighting in the loss function. Preliminary results for this work have been presented at the SPIE 2021 medical imaging conference [5].

5.2 Methods

5.2.1 Data and preprocessing

5.2.1.1 Dataset and ground truth segmentations

The models are trained and evaluated with the RESECT [137] database (see section 2.4.2), considering all iUS volumes (before, during and after resection). The ground truth segmentations of the resection cavity and tumor tissue are the ones used for the single-class models in chapters 3 and 4. The resection ground truth segmentations were made for that study in particular, with the segmentation process described in chapter 3. The tumor ground truth segmentations were made for a previous study on tumor volume [15]. The *sulci*, *falx cerebri*, and ventricles were segmented specifically for this study.

The *sulci*, *falx cerebri*, and ventricles were segmented using ITK-SNAP [1] [142]. The structures were segmented using the contour tool, every couple of slices (usually around one in five slices), then interpolated using the interpolation feature of ITK-SNAP. The *sulci* were segmented in the three slice directions, to increase smoothness and ease the segmentations of *sulci* that were difficult to segment in some directions. All segmentations were then reviewed by the two authors of the segmentations together, looking at slices in all three directions. The segmentation masks were finally edited as needed, following the discussions during the

1. http://www.itksnap.org

class	before resection	during resection	after resection	total	voxels per volume
resection	0	20	22	42	0.77%
tumor	23	11 (generated)	0	38	4.01%
sulci	23	22	22	67	1.47%
falx cerebri	8	8	8	24	0.33%
ventricles	8	9	9	26	0.68%

Table 5.1 – Number of ground truth volumes and mean voxel distribution per class.

review process. We are in the process of validating these segmentations with neurosurgeons, with the goal of publishing a complete set of segmentations for the RESECT database.

No ground truth segmentations of volumes acquired during resection were available. Since delineating tumor tissue manually is challenging, and especially during resection due to artifacts and displacements induced by the resection cavity, we generated these segmentation from the corresponding segmentation of the volume acquired before resection, using an automated process. This was accomplished by mapping the pre-resection segmentation into the volume space of the during-resection iUS volume, using the landmarks provided in the RESECT dataset, using thin plate splines [143]. Then, voxels in the resection cavity ground truth segmentation and out-of-field voxels were masked out from the tumor segmentation. The resulting segmentation were not perfectly accurate but provided a reasonable estimation that can be used for training a model. This allowed the models to learn from cases in which both the resection cavity and tumor tissue are visible and for which ground truth is available.

The classes are highly imbalanced, not only in the number of class voxels compared to background voxels, but also in the number of volumes in which the classes are present (see table 5.1). Generally, the structures that are visible in the iUS volumes depend on the tumor, since the limited US field of view covers a region around the tumor. *Sulci* are visible in all volumes, because the entire surface of the brain is folded. In contrast, the *falx cerebri* and ventricles are only visible in the few cases with a tumor next to these structures. The resection cavity is only present in volumes acquired during and after resection, and the tumor is mostly visible before and during resection, as there is generally little tumor tissue left after resection. As such, there is a high variability in the structures that are visible in the iUS volumes, which is one of the main challenge of this segmentation task.

5.2.1.2 Preprocessing and data augmentation

We use data augmentation to increase model robustness and prevent overfitting. Unlike the single-class models trained in chapters 3 and 4, the data augmentation was performed prior to training and not during. We generated three additional volumes for all volumes in the training set, using the random affine, grid, and scale transformations described in 3.2.1. The four training sets were used cyclically during training, each epoch corresponding to

one set. This is in contrast to the previous augmentation scheme which generated a new set at the beginning of each epoch. While this second option is preferable to a fixed data augmentation, we found that the multi-class case and with a larger training set increased memory requirements and slowed the training process substantially. The training set had more volumes than in the previous studies, because all three iUS volumes (before, during, and after resection) of each case was used. In addition, the transformations has to be applied to five segmentation maps instead of one, thus increasing computing time and memory requirements.

The volumes are normalized as in the previous chapters, by subtracting the mean intensity and dividing by the standard deviation.

5.2.1.3 Patch sampling

2D patches of size 256×256 are extracted using a sliding window with a stride of 64 pixels. Among these patches, the ones that contain only background voxels are discarded. This ensures that at least one class is present in all patches.

5.2.2 Model architecture and loss function

Five models are compared in this study. All models are 2D U-Nets as presented in chapter 3. In the previous chapters on single-class models, we showed that 3D models could obtain better results than 2D models. However, smaller patches were used due to GPU memory constraints. In the multi-class case, more memory is needed for the additional classes, which would have further reduced the input patch size. Thus, we only evaluate 2D models for this multi-class study.

Karimi et al. recently presented a method to estimate the Hausdorff distance that is suitable for backpropagation [114]. We evaluate the effect of such terms by training models with and without this term. The motivation for this choice is that the distance term captures the error for the surfaces, while the Dice score is a volume-based metric. Since *sulci* generally have the shape of thin surfaces, it can be hypothesized that surface-based metrics can be more relevant than volume-based metrics.

The first model is a single-class model which comprises three models trained independently for the resection cavity, tumor, and *sulci*, respectively. The single class models were trained with the Dice loss (equation 2.4). These three submodels are then grouped in one model capable of segmenting these three classes, taking the output of each individual submodel.

model name	type	w_{bg}	w_{HD}
`single`	single-class		
`multi000`	multi-class	0	0
`multi001`	multi-class	0.01	0
`multi000hd`	multi-class	0	1
`multi001hd`	multi-class	0.01	1

Table 5.2 – The five models presented in this study.

The four other models are multi-class models with different coefficients in the loss function. The loss function is given in equation 5.1, in which y_t is the ground truth binary mask, y_p is the output segmentation, w_c is a class weight, w_{HD} is a constant weight for the Hausdorff term, and HD_{CV} is the convolution-based Hausdorff distance presented by Karimi et al. [114]. It consists in a Dice loss term and a Hausdorff distance term, which are both weighted by a class weight w_c.

$$loss(y_t, y_p) = \sum_{c \in \text{classes}} w_c \left[\left(1 - \frac{\sum_i 2 y_{ti} y_{pi} + \epsilon}{\sum_i y_{ti} + y_{pi} + \epsilon} \right) + w_{HD} HD_{CV}(y_t, y_p) \right] \tag{5.1}$$

The class weight depends on the patch considered, and is computed as follows. If the patch contained at least one voxel of class c, then $w_c = w_{fg} = 1$. Otherwise, if the patch did not contain any voxel of class c, but if class c was present in other patches of the volume from which the patch was extracted, then $wc = w_{bg}$, where w_{bg} is a constant. We chose two possible values for w_{bg}: 0 and 0.01, the latter being the order of magnitude of the foreground voxel ratio (see table 5.1). Otherwise, ie. if class c is not present in the entire volume, then $w_c = 0$.

To evaluate the impact of the Hausdorff distance term, we trained models with and without this term, ie. $w_{HD} \in \{0, 1\}$.

Table 5.2 summarizes the five models presented in this chapter. The first model is the single-class model trained separately for each class, referred to as `single`; Next, there are multi-class models trained without the Hausdorff term, referred to as `multi000` ($w_{bg} = 0$) and `multi001` ($w_{bg} = 0.01$). Last, there are multi-class models trained with the Hausdorff term, referred to as `multi000hd` and `multi001hd`.

5.2.3 Post-processing

A segmentation volume is reconstructed from all the corresponding patches, using an average (the patches overlap). This volume is then thresholded at 0.5 to obtain a binary

segmentation. All connected components are labeled in the binary segmentation. This is done for all classes, each class having its own segmentation volume.

For the resection cavity and tumor, the largest connected component is selected and the others are discarded. For the *sulci*, the components with a mean voxel Euclidian distance to the resection cavity that is less than 10 are discarded. This eliminates potential misclassification of hyper-echoic regions (which could be artifacts or hemostasis) on the border of the resection cavity.

Finally, the class of a voxel was chosen as follows. If the voxel was foreground in the *sulci* binary mask, the voxel was labelled as *sulci*. Otherwise, if the voxel was foreground in both the resection cavity and the tumor masks, it was labelled as background. Otherwise, if the voxel was foreground in either the resection cavity and the tumor masks, it was labelled as such in the final segmentation. Otherwise, the voxel was labelled as background. In particular, a voxel that was labelled both resection and tumor is classified as background, as undersegmenting is preferred over labeling a voxel with high uncertainty.

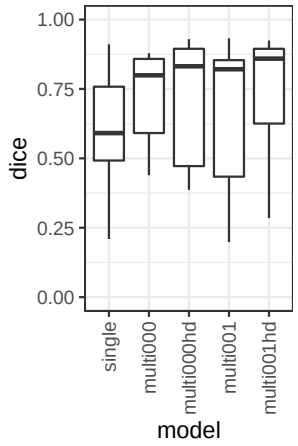
5.3 Results and discussion

In the following, we evaluate the models on the resection cavity, tumor, and *sulci* classes. No model was able to reliably segment the remaining two classes, *falx cerebri* and ventricles, although the multi-class models were able to detect the *falx* in a few cases. This is most likely because there are very few cases in which these two classes are present. In addition, the signal of the *falx cerebri* in iUS volumes is similar to the one of *sulci*, because this signal correspond to CSF in both cases. This makes it difficult to identify the *falx cerebri*, and especially in 2D slices (as opposed to complete 3D volumes).

We separate volumes for which ground truth is available from volumes for which it is not. Volumes with no ground truth are volumes before resection for the resection cavity, and volumes during and after resection for the tumor. Except for volumes during resection, which can contain tumor, foreground voxels can be labeled as false positives. Thus, specificity is relevant for these volumes. All models have specificity values greater than 0.999, indicating that very few voxels were labeled as foreground in these cases. In the following, we only consider the volumes with ground truth.

5.3.1 Comparison of single-class and multi-class models

Overall, the multi-class models have slightly better results than the single-class model. In particular, there is less undersegmentation for resection cavities and less false positives for



model	sensitivity	specificity	precision	dice
`single`	0.48	0.999	0.93	0.59
`multi000`	0.60	0.999	0.95	0.71
`multi000hd`	0.61	0.999	0.96	0.71
`multi001`	0.55	0.999	0.98	0.65
`multi001hd`	0.62	0.999	0.95	0.71

Figure 5.1 – Result metrics for the resection cavity segmentation (5 cases, 9 volumes).

tumors. Both quantatively and qualitatively, the best results are obtained with `multi000hd`.

For the resection cavity, all the multi-class models have better results than the single-class models (figure 5.1). The improvement is a higher sensitivity, while precision is similar. This means there is less undersegmentation with multi-class models (see figure 5.2a for example).

In contrast, the single-class model has a higher sensitivity but lower precision for the tumor segmentations than the multi-class models (figure 5.3). This means that there is more undersegmentation but less false positives in multi-class models, which is desirable in many applications. Quantitatively, the single-class model has higher Dice scores for the tumor segmentations, however the multi-class segmentations are qualitatively better because they have less false positives (see figure 5.2b for example).

The models produce similar *sulci* segmentations, with the exception of `multi001hd` which misses many *sulci* (figure 5.4). The best performing model is `multi000hd`, which misses less *sulci* and has less false positives.

5.3.2 Effect of the weight parameter

With the Hausdorff distance enabled in the loss function, the best w_{bg} value is 0. With $w_{bg} = 0.01$, the model has substantially worse results for the tumor and *sulci*. However, without the Hausdorff distance, setting $w_{bg} = 0.01$ slightly improves the results over $w_{bg} = 0$.

This improvement is minimal compared to the decreased accuracy with the Hausdorff distance enabled. In addition, the combination of $w_{bg} = 0$ and $w_{HD} = 1$ yields the best results. Thus, $w_{bg} = 0$ seems to be the best option.

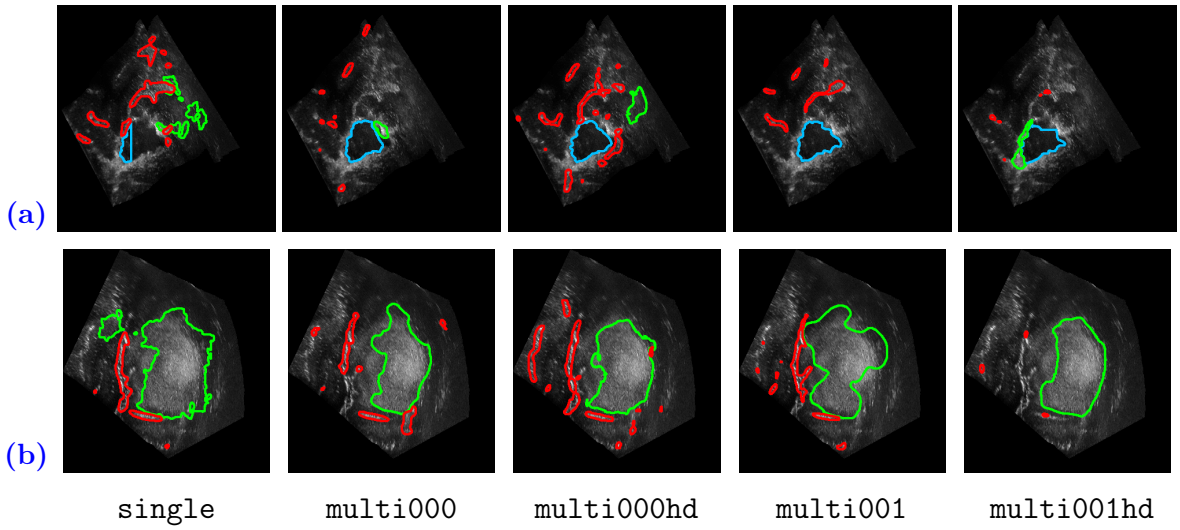
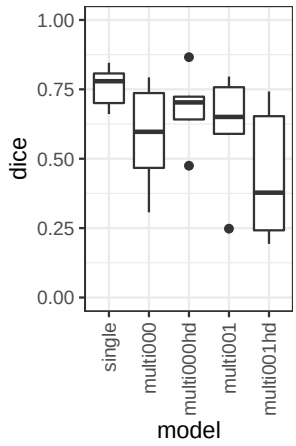
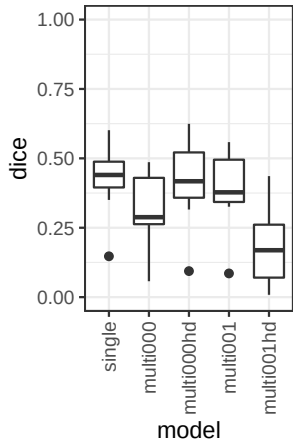


Figure 5.2 – Examples of multi-class segmentations (blue: resection cavity; green: tumor; red: *sulci*).



model	sensitivity	specificity	precision	dice
`single`	0.78	0.999	0.75	0.76
`multi000`	0.48	0.999	0.76	0.58
`multi000hd`	0.55	0.999	0.95	0.68
`multi001`	0.52	0.999	0.83	0.61
`multi001hd`	0.32	0.999	0.95	0.44

Figure 5.3 – Result metrics for the tumor segmentation (5 cases, 5 volumes).



model	sensitivity	specificity	precision	dice
`single`	0.41	0.999	0.52	0.44
`multi000`	0.36	0.999	0.36	0.32
`multi000hd`	0.46	0.999	0.47	0.42
`multi001`	0.43	0.999	0.46	0.40
`multi001hd`	0.11	0.999	0.69	0.18

Figure 5.4 – Result metrics for *sulci* segmentation (5 cases, 15 volumes).

5.3.3 Hausdorff distance loss function

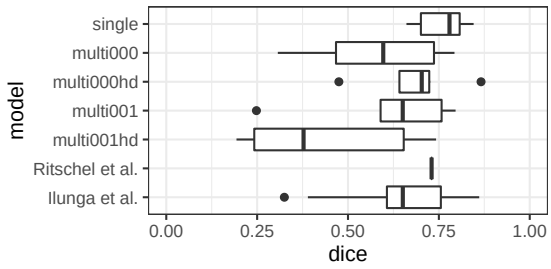
With $w_{bg} = 0$, the model trained with the Hausdorff term have better than the model trained without it. The tumor and *sulci* segmentations are improved with the Hausdorff term enabled, while results for the resection cavity are similar.

The best results are obtained with `multi000hd`, which suggests that the training process may benefit from the Hausdorff term.

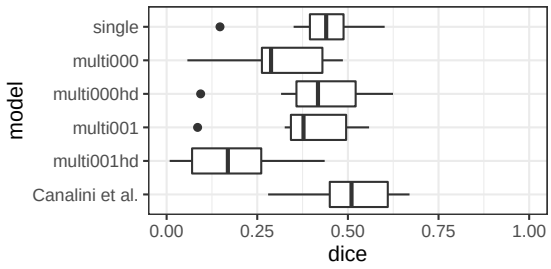
5.3.4 Comparison with other methods

The results for tumor segmentation can be compared to previous methods [105,106], even though these methods were applied to high grade tumors in contrast to this study. Overall, Dice scores obtained with the models presented in this study are similar to the ones reported in the aforementioned studies, with the exception of `multi001hd` which has lower scores than any other method. Figure 5.5a shows boxplots of the Dice scores for the different models.

Canalini et al. [38] proposed a segmentation method for hyperechogenic structures such as *sulci* and the *falx cerebri*. This is not strictly comparable to the *sulci* segmentation models presented in this chapter, because we distinguish the *sulci* and *falx cerebri*. We trained the models to segment only *sulci* and the *falx cerebri* specifically, as separate classes, which is more specific than hyperechogenic structures. Nonetheless, we provide a comparison of the Dice scores in figure 5.5b. Note that Canalini et al. [38] report both normal Dice scores and Dice scores after removing non-overlapping components. For a fair comparison with the other models, we plot the normal Dice scores, as this metric was also used for the other models. The results obtained with the best model `multi000hd` are similar to the results



(a) tumor segmentation methods.



(b) sulci segmentation methods.

Figure 5.5 – Comparison with other brain ultrasound segmentation methods.

reported by Canalini et al.

Nitsch et al. [39] proposed a segmentation method for the *falx cerebri* and *tentorium cerebelli*. Their method is able to reliably segment such structures, with a mean Dice score of 0.74. While the models presented here were not able to reliably segment the *falx cerebri*, it is worth noting two major differences in the datasets. First, the cases used by Nitsch et al. all contains the *falx cerebri*, in contrast to this study that uses all RESECT cases, most of which do not contain the *falx cerebri*. Second, the dataset used in Nitsch et al. is more recent than RESECT, and thus, the iUS image quality is better.

5.4 Conclusions

In this chapter, we compared single- and multi-class segmentation models for iUS brain images. The results suggest that multi-class models may benefit from learning all classes together, and obtain more accurate segmentations than single-class models trained independently. While the Dice scores allow a quantitative analysis of the results, they do not reflect completely the differences between the proposed models, and especially for the *sulci*. In particular, we found substantial differences by looking at the resulting segmentations, as illustrated by the examples in figure 5.2.

The `multi000hd` model obtains the best results, quantitatively and especially qualitatively. This suggests that the Hausdorff term may be beneficial to the training process, and that it is better to ignore patches with only background, even if the class is present elsewhere in the corresponding volume.

The training data is challenging because of a high class imbalance, with some classes missing in certain volumes. Discarding patches with only background, as well as class weighting in the loss function allowed successful training of the models, despite the limitations of the training data.

The models were successfully trained to segment the resection cavity, tumor, and *sulci*. However, they could not segment the *falx cerebri* and ventricles, because there were not enough volumes with these classes. The motivation for having separate classes for the *sulci* and *falx cerebri* is that the *falx cerebri* can provide a valuable landmark for registration. Nonetheless, it could be interesting to merge these two classes, since their ultrasound signals both correspond to CSF.

The models presented in this chapter are 2D models, due to GPU memory constraints. As GPU cards with more memory become available, it would be interesting to train 3D models. We expect that 3D models would have better results, as experimented with the single-class models in the previous chapter. In particular, it is easier to classify the *falx cerebri* with the full 3D volume than with a single or a few slices.

Chapter 6

Registration of preoperative MRI with intraoperative ultrasound images for brain shift compensation using deep learning

6.1 Introduction

In this chapter, we explore the use of deep learning models for registration of the pMR with iUS images, for brain shift compensation. More specifically, we aim to use the segmentations of structures in the pMR and iUS to guide to registration model. We assume that these segmentations are available, since we proposed segmentation models in the previous chapters.

An overview of deep learning registration methods is given in section 2.3.3. Very few deep learning methods have been proposed for image registration, and most of them are applied to monomodal registration. Registration of brain pMR and iUS is challenging, not only because this is a multimodal registration problem, but also because the field of view of the iUS is very limited compared to the pMR and varies substantially across cases, depending on the tumor location. In addition, there is no complete and accurate ground truth for the tissue displacement, which would simplify the training of a deep learning model.

Hu et al [124] proposed the use of labels as a weak supervision. In their method, the labels are used in the loss function, with a Dice score between the registered labels of the moving space and the labels of the fixed space. The benefit of such approach is that labels are not required for registering unseen images. In our case, we can assume that the segmentations are available, since we developed segmentation models for various structures. Thus, we propose to add the labels to the model inputs, to provide more information to the model. In addition, we adapt the LocalNet architecture proposed bu Hu et al. [124] to multimodal registration by having two encoder paths for each modality.

6.2 Methods

6.2.1 Notations

For a physical point p, let $p^{(\mathrm{US})}$ be the corresponding point in the iUS volume, and $p^{(\mathrm{MR})}$ the corresponding point in the pMR volume. When not specified, p can be either in iUS or pMR space. For a volume V and a point $p_V = (x, y, z)$, $V[p] = V[p_V] = V[x, y, z]$ is the voxel in V at point p.

There are several coordinate systems that can be used to express a point p, so specify the coordinate system where it is relevant by using p_W for world coordinate, p_V for volume coordinates, and p_n for normalized volume coordinates.

The world coordinate system is a reference coordinate system common to the pMR and iUS volumes, which correspond to an affine transformation of volume coordinates. This affine transform is computed during the calibration process. It is stored in the metadata of the RESECT volume files as a 4×4 affine matrix in homogeneous coordinates $A_{\mathrm{US}\to W}$, which map volume coordinates to world coordinates (equation 6.1). Note that equation 6.1 stands not only for $p^{(\mathrm{US})}$ but also for $p^{(\mathrm{MR})}$ since the pMR volumes are resampled in iUS volume space. This resampling is done by mapping a point in iUS volume space to the corresponding point in the original pMR space with the transform $A^{-1}_{\mathrm{MR}\to W} \cdot A_{\mathrm{US}\to W}$.

$$p_W = A_{\mathrm{US}\to W} \cdot p_V \tag{6.1}$$

The volume coordinates are the integer indices of voxels, ranging from 0 to $N - 1$ where N is the size of the volume. Note that we only consider one volume coordinate system, which is the one of the iUS volume. Because the pMR volumes are resampled in iUS volume space, they have the same volume coordinate systems as the corresponding iUS volumes.

Normalized volume coordinates are volume coordinates that were scaled to the range $]-1, 1[$. The benefit of this coordinate system is that it does not depend on the volume size, and thus can be used at various scales. Conversion between volume coordinates and normalized volume coordinates is done using a scale matrix S (equation 6.2). In our implementation, we also flip the x and z coordinates to match the semantics of Pytorch's [1] `grid_sample`, which indexes the grid as z, y, x. Thus, S is actually defined as the product of a matrix that flips the coordinates and a scale matrix (equation 6.3). This is an implementation detail, however, which does not introduce any conceptual change.

1. https://pytorch.org/

$$p_n = S \cdot p_V \tag{6.2}$$

$$S = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{2}{N_x} & 0 & 0 & -1 + \frac{1}{N_x} \\ 0 & \frac{2}{N_y} & 0 & -1 + \frac{1}{N_y} \\ 0 & 0 & \frac{2}{N_z} & -1 + \frac{1}{N_z} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.3}$$

We define the identity grid I which gives an identity mapping. In particular, the identity grid in normalized coordinates is defined as:

$$I_n[p] = p_n \tag{6.4}$$

6.2.2 Ground truth

6.2.2.1 Labels

We used the same ground truth segmentations as in the previous chapters. In this chapter, we focus on registration as this problem is challenging in itself. Thus, we train the models using the ground truth segmentation as opposed to the outputs of the segmentation models.

The MRI segmentations were created using freesurfer [2]. The `recon-all` scripts outputs a segmentation map, which we post-processed to create segmentations for the same classes as the US segmentations. The ventricles segmentation were used as is. The *sulci* segmentation was created by selecting all background voxels that were included in the brain mask. This brain mask was obtained with the binary closing of the brain tissue classes, with a sphere of radius 7 as the structural element. Then, the *falx cerebri* was located using the *corpus callosum* class, and extracted from the sulci mask.

6.2.2.2 Landmarks

Obtaining accurate ground truth mappings is often challenging in image registration problems. In this work, we rely on the landmarks provided in the RESECT database to compute an estimate of the pMR-iUS mapping. These landmarks are a set of homologous points in the pMR and iUS, which we refer to as lm_{MR} and lm_{US}, respectively (see figure 6.1. The pMR-iUS mapping is estimated using thin plate splines [143] with the landmarks. This

2. https://surfer.nmr.mgh.harvard.edu/

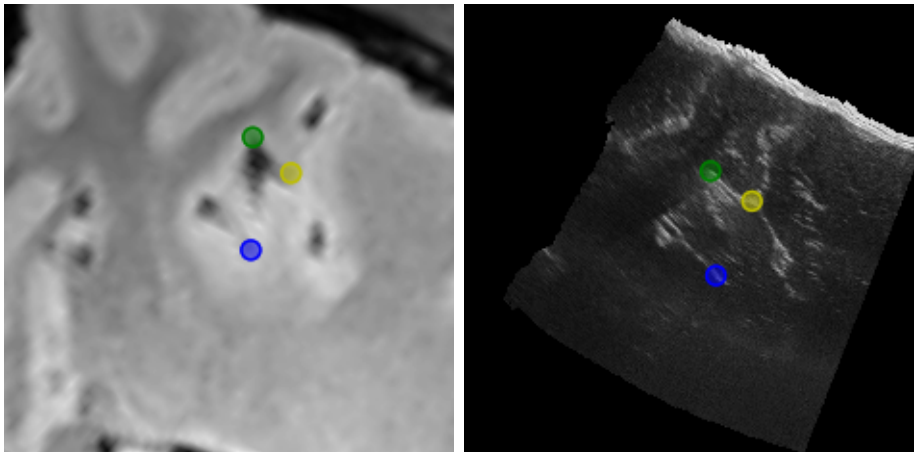


Figure 6.1 – Corresponding 2D slices of a pMR and iUS in RESECT, with landmarks represented as colored spheres.

approach has several limitations, thus, the estimated mappings are not completely accurate. The main limitation is that the number of landmarks is small (around 15 per volume) and the displacements at other points is interpolated. This guarantees that the displacement field is smooth, but may be inaccurate if some displacement is not captured by the landmarks. An example of this is shown in figure 6.2, in which the *falx cerebri* in the bottom of the left image and the *sulci* in the right image are shifted. In addition, the boundary of the ventricle is visible in the top right part of the image on the right, and is also shifted compared to the ventricle shown in the MRI. In the former case (left image), this is because the landmarks are located around the tumor, and there are no landmarks on the *falx cerebri*. More generally, the landmarks are located on points for which a correspondence could be found (such as *sulci* bifurcation points and ventricle corners), thus, they do not cover the whole volumes. Another limitation is that it is extremely challenging to precisely select corresponding points, especially in US images, so there may be a small error in these correspondences.

The thin plate splines mappings were computed prior to training, in world coordinates. This allowed using the mappings with augmented data, as the transformed coordinates could be mapped to world coordinates.

6.2.3 Registration model

In this chapter, we build on the LocalNet model proposed by Hu et al. [124]. In addition to the standard LocalNet, we also propose a modified LocalNet with two encoder paths and with segmentation labels and field of view mask added to the inputs.

LocalNet is a convolutional network designed to output a dense displacement field, from a pair of images. The input includes the images, labels, and a mask corresponding to the field

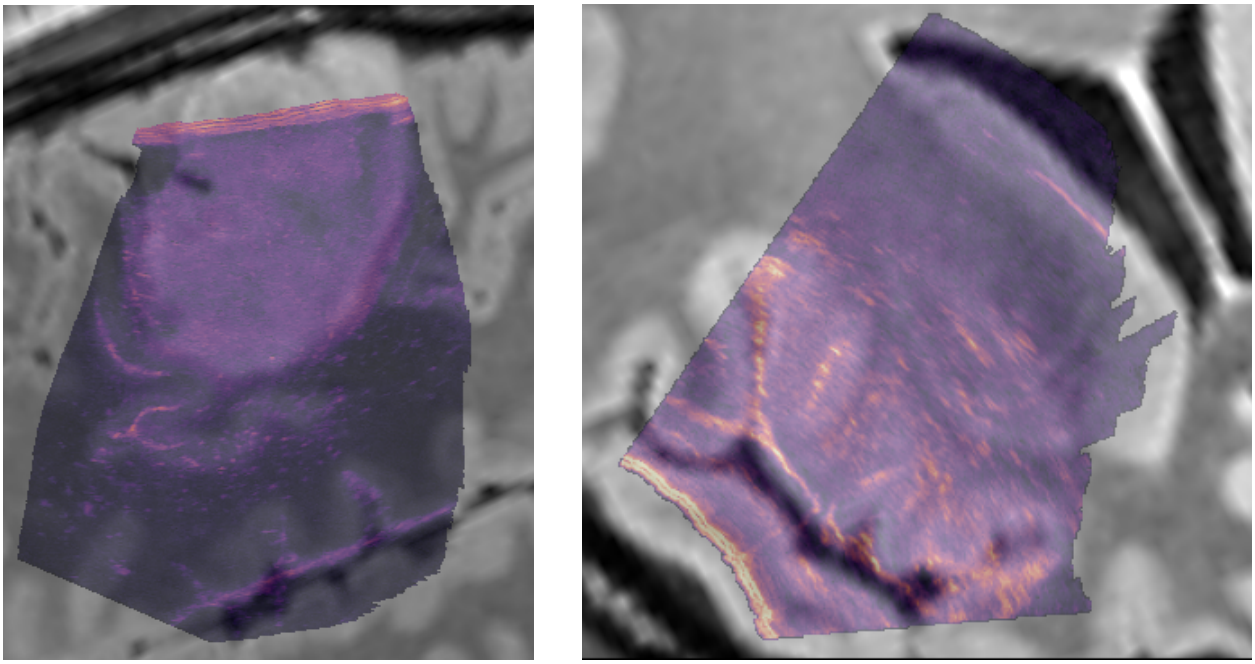


Figure 6.2 – Example of error in the ground truth displacement field: the *falx cerebri* (left image) and *sulci* (right image) in the overlaid iUS do not match the registered MR.

of view of the iUS volume. We defined that the fixed image is the iUS and the moving image is the pMR. This means that the output mapping f transforms $p^{(\mathrm{US})}$ to $p^{(\mathrm{MR})}$, which allows resampling the pMR in iUS space (hence the pMR is the moving image). This registered (updated) pMR volume is denoted $\widetilde{\mathrm{MR}}$ and is obtained by sampling the original pMR volume (MR):

$$\forall p_V^{(\mathrm{US})}; \widetilde{\mathrm{MR}}[p_V^{(\mathrm{US})}] = \mathrm{MR}[p_V^{(\mathrm{MR})}] = \mathrm{MR}[f(p_V^{(\mathrm{US})})] \tag{6.5}$$

The output of the network is a dense displacement field volume DDF, which gives the normalized displacement of all voxels in the volume:

$$\mathrm{DDF}[p_V^{(\mathrm{US})}] = p_n^{(\mathrm{MR})} - p_n^{(\mathrm{US})} \tag{6.6}$$

This displacement field gives a discretized version of the iUS to pMR mapping:

$$p_n^{(\mathrm{MR})} = (\mathrm{DDF} + I_n)[p_n^{(\mathrm{US})}] \tag{6.7}$$

Since this is a multi-modal registration problem, we propose a modified LocalNet architecture so that there are two encoder paths. This allows separate processing of the MR and

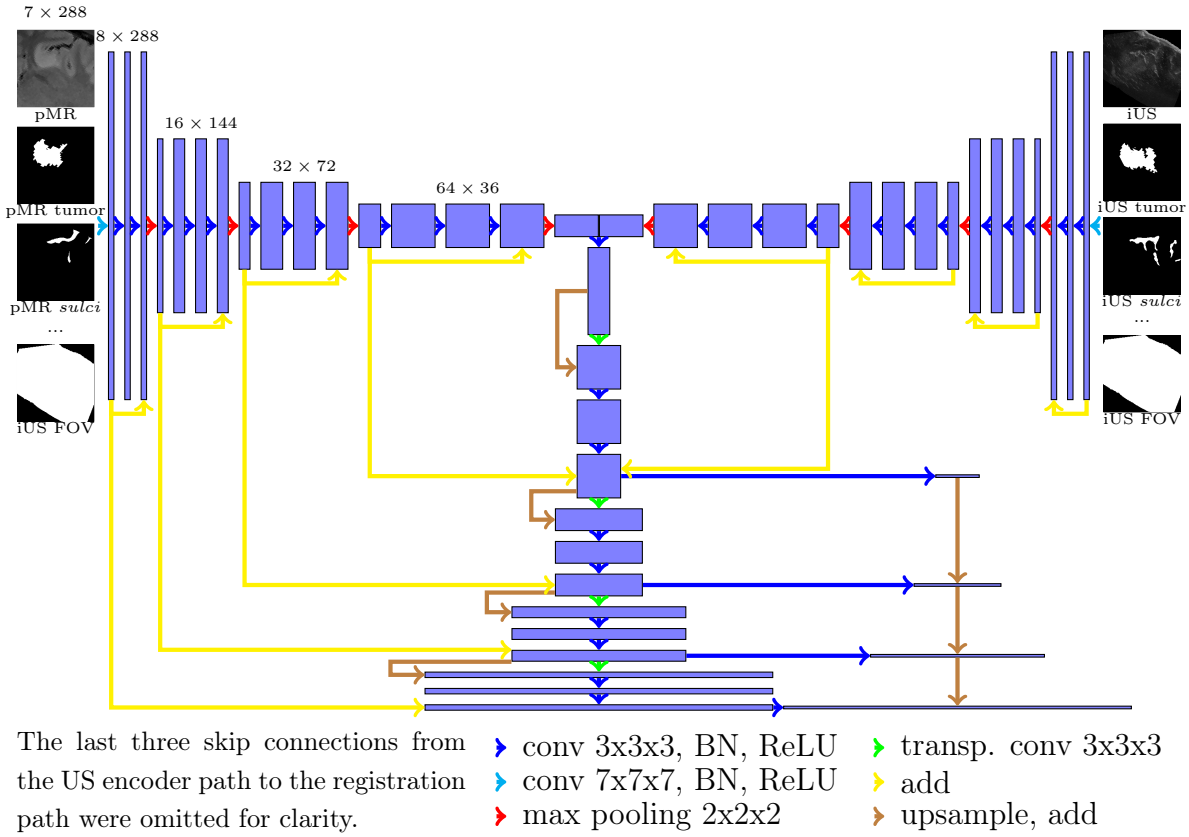


Figure 6.3 – Modified LocalNet architecture with two encoder paths (MR, US).

US volumes. This is motivated by previous studies on other clinical applications that suggested that encoder-decoder models could benefit from two encoder paths in the multimodal case [144]. Figure 6.3 illustrates this architecture.

6.2.4 Loss function

The loss function consists of several terms, which are given in equation 6.8. In this equation, w_{TRE}, w_{L1}, w_E, and w_{labels} are the fixed weights assigned to the loss terms. We experimented with several weight values, including disabling terms with a weight of 0.

$$\begin{aligned} loss = & \; w_{TRE} \cdot \text{mTRE}(lm_{MR}, \widetilde{lm_{US}}) \\ & + w_{L1} \cdot \text{l1}(DDF_t, DDF_p) \\ & + w_E \cdot \text{energy}(DDF_p) \\ & + w_{labels} \cdot \sum_{c \in \text{classes}} \text{diceloss}(\widetilde{L_{cMR}} \cdot fov, L_{cUS} \cdot fov) \end{aligned} \tag{6.8}$$

The first term is the landmarks registration error. It is the mean absolute error between the landmark points in the MR space lm_{MR} and the landmark points in the US space after registration $\widetilde{lm_{US}}$. This is computed by adding the displacement specified in the output displacement field DDF_p (equation 6.9). The points (and thus distances) are specified in normalized volume coordinates.

The second term is an L1 comparison of the output displacement field DDF_p with the TPS-generated ground truth displacement field DDF_t.

The third term is a regularization term for the output displacement field. It is the bending energy function as was used in Hu et al. [124] (equation 6.10).

The last term is the label term, which compares the US labels L_{cUS} with the registered MR labels $\widetilde{L_{cMR}}$, for each class c. The MR labels are registered by resampling the original MR labels with the grid $I_n + DDF_p$. Since the field of view of the US volume does not cover the entire volume, the label maps are masked with the field of view mask fov. This removes MR labels outside the US field of view. The labels are compared with the dice loss, as defined in equation 2.4.

$$\widetilde{lm_{US}} = lm_{US} + DDF_p[lm_{US}] \tag{6.9}$$

$$\text{energy}(V) = \left(\frac{\partial V}{\partial x^2}\right)^2 + \left(\frac{\partial V}{\partial y^2}\right)^2 + \left(\frac{\partial V}{\partial z^2}\right)^2 + 2\left(\frac{\partial V}{\partial xy}\right)^2 + 2\left(\frac{\partial V}{\partial yz}\right)^2 + 2\left(\frac{\partial V}{\partial xz}\right)^2 \tag{6.10}$$

In this study, we did not include a term based on the image intensities, because finding such term for multimodal registration is difficult. In particular, metrics that are traditionally used for multimodal registration are not directly implementable in a loss function, as they are not differentiable.

6.2.5 Data preprocessing

The data was preprocessed the same way as the previous chapters, including data augmentation.

6.3 Discussion

Training the models presented in this chapter has proven difficult, and has not led to satisfying results. Depending on the parameters, the training process would either diverge or converge to a non working solution. In this section, we discuss several findings learned from experimenting with these models, and possible improvements.

6.3.1 Dataset size and complexity

The dataset size is small, in particular for training deep learning models, and especially registration models. Other studies that successfully trained deep learning registration models have larger datasets.

Another challenge is the high variability of the data. Different structures may be visible depending of the tumor location.

In addition, multimodal registration is challenging, and especially with ultrasound images. The pMR and iUS do not image the same physical properties, and the iUS has a limited field of view covering only part of the pMR.

6.3.2 Loss function

The weights of the different loss terms is critical to the convergence of the training process, and affects greatly the outcome. It is therefore essential to chose the weights with care.

The mean landmarks registration error term leads to training instability. Thus, we removed this term entirely in the following experiments. In addition, the usefulness of this term is not clear in the presence of the l1 comparison with the ground truth. An alternative could be to weight the l1 term per voxel, using the distance to landmarks. This would put more weight on voxels close to the landmarks points, for which the ground truth is more reliable since these points were used for estimating ground truth. Conversely, it would assign less importance to voxels further to the landmarks, where the ground truth displacements are interpolated and less accurate.

We found that assigning larger weights to the l1 term did not reduce the l1 error further. This is true regardless of the training stage, and in particular in the end of the training process. We hypothesize that it may be because the ground truth DDF was generated from the landmarks, and thus only accurate at these points.

The regularization term is essential to the convergence of the training process. Its weight should be large enough such that the energy of the DDF volume decreases during training.

6.3.3 Quality of the ground truth DDF

As discussed in the previous sections, the ground truth DDF is an estimation based on the landmarks. Improving the accuracy of the ground truth may allow a better training of the models, since it is the main measure of how accurate the output DDF is. While the label terms give another measure of the DDF accuracy, it is limited to the location of the labels. Thus, the l1 term is the only term that fully evaluates the DDF for every voxel.

The ground truth DDF could be improved iteratively by adding landmarks in areas where the errors are the largest. This could be done by visually inspecting the registered volumes and estimating the registration error. However, this process would be time consuming and error-prone. An alternative would be to use the segmentation labels as an additional constraint for the DDF estimation.

6.3.4 Output of the registration network

The choice of a DDF for the registration network is motivated by the fact that brain shift may include non rigid displacements, especially after resection started. However, computing such displacements is extremely challenging, not only for registration models but also for determining a ground truth.

For this reason, many previous pMR-iUS registration methods for brain shift limited the output to affine transforms. Other studies first estimate a global affine transform, and then a DDF refining the affine transform. Both approaches are worth exploring in the context of deep learning models. Another option would be coarser DDF outputs, which would simplify registration and ground truth estimation while retaining the precision of DDFs. Pdd-net is an example of such output [44, 131].

6.3.5 Network architecture

In this work, we followed and build on the LocalNet architecture that was designed for label-driven registration. Other alternatives could be explored, such as combining the segmentation and registration problems in one model, since they are inter-dependent. End-to-end training of several tasks simultaneously has been shown to increase model accuracy, however training such models generally need large training sets.

6.4 Conclusions

In this chapter, we proposed a deep learning model using segmentation labels to drive registration of the pMR with iUS. While training of such models proved unsuccessful, we present the findings learned when training the models, discuss the limitations of the proposed method, and suggest areas in which improvements can be made.

One of the major challenges the limited amount of data available, and the high variability of the input, which highly depends on the tumor location. The number of cases used in this study is substantially lower than in other deep learning registration studies. Thus, we expect that a larger dataset will help the training of deep learning registration models for pMR-iUS brain shift compensation.

Another challenge is the lack of accurate ground truth, which is difficult to estimate. This could be mitigated by using a simpler representation such as affine transforms, either as the final output or a first registration step.

Chapter 7

Conclusions

7.1 Summary

The work presented in this thesis evaluates deep learning methods for brain iUS segmentation and pMR-iUS registration. We first proposed single-class segmentation models for the resection cavity and tumor tissue. Next, we adapt these models for multi-class segmentation of several structures, including the *sulci* and *falx cerebri*. Finally, we propose a registration model for registering the pMR with iUS volumes.

We show that single-class models based on U-Net can successfully segment the resection cavity and tumor in iUS images. 3D models are more accurate than 2D models, but have a higher runtime. The sliding window approach was the most robust, and downsampling should be avoided for US images, as it lead to substantially worse results. For tumor segmentation, adding the pMR as an additional input could help localizing the tumor in the iUS volume, but this would require registration.

Our results suggest that multi-class models can be more accurate than independent single-class models. We successfully trained a multi-class segmentation model for the resection cavity, tumor and *sulci*. The dataset did not contain enough volumes in which the *falx cerebri* and ventricles were in the field of view of the iUS volume to obtain reliable segmentation models for these structures. Adding a Hausdorff distance term in the loss function seems to improve the results, in particular for the *sulci* which are surface-like structures. For these structures, the Dice score has limitations, since it is a volume-based metric.

We were not able to successfully train a pMR-iUS registration deep learning model. We believe this is due to the limited dataset and incomplete ground truth, in addition to the higher complexity of image registration. Building larger brain iUS datasets will be instrumental in the development of deep learning pMR-iUS registration models.

For the segmentation and registration studies presented in this thesis, we have created a set of ground truth segmentations. The resection cavity segmentations were checked by

a neurosurgeon, and we are in the process of validating the other segmentation masks. We plan to publish these segmentations once they are validated, so that further studies may benefit from them.

7.2 Discussion

In general, our segmentation results show that deep learning models can produce very accurate results for brain iUS images. The results obtained are comparable to other methods (see sections 3.3.5, 4.3.5 and 5.3.4) and close to intra-rater variability (see sections 3.3.4 and 4.3.4).

However, they also highlight that one of the main limitation of such models is robustness. This is shown by the outlier cases, for which the segmentation have little to no overlap with the ground truth. This is also evidenced by the generalizability study we performed in chapter 3. The model trained on one dataset could not generalize well to the other.

The robustness of the models depends on the dataset size. This is a limitation of the models presented in this thesis. We mainly used the 23 cases of the RESECT database, which is the most recent public dataset of brain iUS volumes available. We also evaluated the resection cavity models on the BITE dataset, an older database with 14 cases. This is substantially lower than the number of cases used in other deep learning studies in medical image processing, which commonly use hundreds or even thousands of cases. The limited number of cases impairs not only training but also evaluation. A larger dataset would enable a better training of the models, as well as a better evaluation of their generalization to unseen cases.

With the limited data available, we were able to train segmentation models. Registration models likely need more data as well as more research, and thus are still a work in progress.

7.3 Future works

7.3.1 Validation of the manual segmentations

The manual segmentations created as part of the work in this thesis are valuable data, which other studies may benefit from. As such, we are in the process of creating a validated set of publicly available segmentations for the RESECT database.

7.3.2 Multi-class segmentation

While the *falx cerebri* can represent valuable information for registration processes, it would be interesting to compare the models presented here (with the *sulci* and *falx cerebri* separate), with models trained with these two classes merged.

It would also be valuable to train 3D multi-class models, as newer GPU card with more memory allow larger 3D inputs. In particular, 3D models may be able to segment the *falx cerebri*, which is otherwise difficult with only a single 2D slice.

7.3.3 Registration model

There are several areas of possible improvement for the registration model. In particular, it is difficult to define a ground truth, and the proposed estimation using TPS with the provided landmarks can be refined. For example, the ground truth segmentations could be used in addition to the landmarks to estimate a deformation that matches both. In addition, changes to the model such as using an affine transform or a coarser DDF, either as a first registration step or as a global output, are worth exploring. Finally, it would be interesting to study deep learning models for monomodal US-US registration. This would be useful for US images acquired at different times during surgery, and can be less challenging than multimodal MR-US registration.

7.3.4 Larger datasets

The proposed models would benefit from larger datasets, for training and validation. This would require collecting more data, since the only public brain iUS datasets are RESECT and BITE, which are used in this thesis. While collecting clinical data takes time and is complicated, an simpler alternative is to synthesise such data. There are many approaches for data augmentation, including deep learning with generative adversarial networks, which is still an active research domain. These methods could help in our context, to build a larger dataset from the limited number of cases we have. Of course, the longer term goal is to acquire actual clinical data.

Bibliography

[1] F.-X. Carton, M. Chabanas, F. Le Lann, and J. H. Noble. **Automatic segmentation of brain tumor resections in intraoperative ultrasound images using U-Net.** In *Journal of Medical Imaging*, vol. 7, no. 3, pp. 1–15 (2020). DOI: 10.1117/1.JMI.7.3.031503.

[2] F.-X. Carton, J. H. Noble, and M. Chabanas. **Automatic segmentation of brain tumor resections in intraoperative ultrasound images.** In *Proceedings of SPIE*, vol. 10951 (2019). DOI: 10.1117/12.2513613.

[3] F.-X. Carton, J. H. Noble, B. K. R. Munkvold, I. Reinertsen, and M. Chabanas. **Automatic segmentation of intraoperative ultrasound images of the brain using U-Net.** In *Surgetica 2019 - Computer-Assisted Medical Interventions: scientific problems, tools and clinical applications*, pp. 66–68 (2019).

[4] F.-X. Carton, M. Chabanas, B. K. R. Munkvold, I. Reinertsen, and J. H. Noble. **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 (2020). DOI: 10.1117/12.2549516.

[5] F.-X. Carton, J. H. Noble, F. Le Lann, B. K. R. Munkvold, I. Reinertsen, and M. Chabanas. **Multiclass segmentation of brain intraoperative ultrasound images with limited data.** In *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11598, pp. 149–154 (2021). DOI: 10.1117/12.2581861.

[6] M. Donnez, F.-X. Carton, F. Le Lann, E. De Schlichting, and M. Chabanas. **Realistic synthesis of brain tumor resection ultrasound images with a generative adversarial network.** In *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11598, pp. 637–642 (2021). DOI: 10.1117/12.2581911.

[7] M. L. Goodenberger, and R. B. Jenkins. **Genetics of adult glioma.** In *Cancer Genetics*, vol. 205, no. 12, pp. 613–621 (2012). DOI: 10.1016/j.cancergen.2012.10.009.

[8] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues. **The 2007 WHO classification of tumours of the central nervous system.** In *Acta Neuropathologica*, vol. 114, pp. 97–109 (2007). DOI: 10.1007/s00401-007-0243-4.

[9] A. H. Jacobs, L. W. Kracht, A. Gossmann, M. A. Rüger, A. V. Thomas, A. Thiel, and K. Herholz. **Imaging in neurooncology.** In *NeuroRX*, vol. 2, no. 2, pp. 333–347 (Apr. 2005). DOI: 10.1602/neurorx.2.2.333.

[10] A. Perkins, and G. Liu. **Primary brain tumors in adults: Diagnosis and treatment.** In *American family physician*, vol. 93, no. 3, pp. 211–217 (Feb. 2016).

[11] S. J. Hentschel, and R. Sawaya. **Optimizing outcomes with maximal surgical resection of malignant gliomas.** In *Cancer Control*, vol. 10, no. 2, pp. 109–114 (2003). DOI: 10.1177/107327480301000202.

[12] N. Sanai, and M. S. Berger. **Glioma extent of resection and its impact on patient outcome.** In *Neurosurgery*, vol. 62, no. 4, pp. 753–766 (Apr. 2008). DOI: 10.1227/01.neu.0000318159.21731.cf.

[13] G. Unsgaard, S. Ommedal, T. Muller, A. Gronningsaeter, and T. A. Nagelhus Hernes. **Neuronavigation by intraoperative three-dimensional ultrasound: Initial experience during brain tumor resection.** In *Neurosurgery*, vol. 50, no. 4, pp. 804–812 (Apr. 2002). DOI: 10.1097/00006123-200204000-00022.

[14] G. Unsgaard, O. M. Rygh, T. Selbekk, T. B. Müller, F. Kolstad, F. Lindseth, and T. A. N. Hernes. **Intra-operative 3D ultrasound in neurosurgery.** In *Acta Neurochirurgica*, vol. 148, pp. 235–253 (2006). DOI: 10.1007/s00701-005-0688-y.

[15] B. K. R. Munkvold, H. K. Bø, A. S. Jakola, I. Reinertsen, E. M. Berntsen, G. Unsgård, S. H. Torp, and O. Solheim. **Tumor volume assessment in low-grade gliomas: A comparison of preoperative magnetic resonance imaging to coregistered intraoperative 3-dimensional ultrasound recordings.** In *Neurosurgery*, vol. 83, no. 2, pp. 288–296 (2018). DOI: 10.1093/neuros/nyx392.

[16] X. Fan, D. W. Roberts, J. D. Olson, S. Ji, T. J. Schaewe, D. A. Simon, and K. D. Paulsen. **Image updating for brain shift compensation during resection.** In *Operative Neurosurgery*, vol. 14, no. 4, pp. 402–411 (2018). DOI: 10.1093/ons/opx123.

[17] M. I. Miga, K. Sun, I. Chen, L. W. Clements, T. S. Pheiffer, A. L. Simpson, and R. C. Thompson. **Clinical evaluation of a model-updated image-guidance approach to brain shift compensation: experience in 16 cases.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 8, pp. 1467–1474 (Aug. 2016). DOI: 10.1007/s11548-015-1295-x.

[18] I. J. Gerard, M. Kersten-Oertel, K. Petrecca, D. Sirhan, J. A. Hall, and D. L. Collins. **Brain shift in neuronavigation of brain tumors: A review.** In *Medical Image Analysis*, vol. 35, pp. 403–420 (2017). DOI: 10.1016/j.media.2016.08.007.

[19] S. Bayer, A. Maier, M. Ostermeier, and R. Fahrig. **Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery.** In *International Journal of Biomedical Imaging*, vol. 2017 (2017). DOI: 10.1155/2017/6028645.

[20] I. J. Gerard, M. Kersten-Oertel, J. A. Hall, D. Sirhan, and D. L. Collins. **Brain shift in neuronavigation of brain tumors: An updated review of intra-operative ultrasound applications.** In *Frontiers in Oncology*, vol. 10, pp. 3390–3401 (2021). DOI: 10.3389/fonc.2020.618837.

[21] A. Sotiras, C. Davatzikos, and N. Paragios. **Deformable medical image registration: A survey.** In *IEEE transactions on medical imaging*, vol. 32 (May. 2013). DOI: 10.1109/TMI.2013.2265603.

[22] H. Rivaz, and D. L. Collins. **Deformable registration of preoperative MR, pre-resection ultrasound, and post-resection ultrasound images of neurosurgery.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 7, pp. 1017–1028 (Jul. 2015). DOI: 10.1007/s11548-014-1099-4.

[23] M. Riva, C. Hennersperger, F. Milletari, A. Katouzian, F. Pessina, B. Gutierrez-Becker, A. Castellano, N. Navab, and L. Bello. **3D intra-operative ultrasound and MR image guidance: pursuing an ultrasound-based management of brainshift to enhance neuronavigation.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 10, pp. 1711–1725 (Oct. 2017). DOI: 10.1007/s11548-017-1578-5.

[24] D. H. Iversen, W. Wein, F. Lindseth, G. Unsgård, and I. Reinertsen. **Automatic intraoperative correction of brain shift for accurate neuronavigation.** In *World Neurosurgery*, vol. 120, pp. e1071–e1078 (2018). DOI: 10.1016/j.wneu.2018.09.012.

[25] W. Wein. **Brain-shift correction with image-based registration and landmark accuracy evaluation.** In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation : International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018*, vol. 11042 (2018). DOI: 10.1007/978-3-030-01045-4_17.

[26] R. Shams, M.-A. Boucher, and S. Kadoury. **Intra-operative brain shift correction with weighted locally linear correlations of 3DUS and MRI.** In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation : International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018*, vol. 11042 (2018). DOI: 10.1007/978-3-030-01045-4_22.

[27] M. Heinrich. **Intra-operative ultrasound to MRI fusion with a public multimodal discrete registration tool.** In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation : International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018*, vol. 11042 (2018). DOI: 10.1007/978-3-030-01045-4_19.

[28] Y. Xiao, L. Eikenes, I. Reinertsen, and H. Rivaz. **Nonlinear deformation of tractography in ultrasound-guided low-grade gliomas resection.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 3, pp. 457–467 (Mar. 2018). DOI: 10.1007/s11548-017-1699-x.

[29] N. Masoumi, Y. Xiao, and H. Rivaz. **Arena: Inter-modality affine registration using evolutionary strategy.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 3, pp. 441–450 (Mar. 2019). DOI: 10.1007/s11548-018-1897-1.

[30] P. Coupé, P. Hellier, X. Morandi, and C. Barillot. **3D rigid registration of intraoperative ultrasound and preoperative MR brain images based on hyperechogenic structures.** In *International Journal of Biomedical Imaging*, vol. 2012, pp. 1–14 (Jan. 2012). DOI: 10.1155/2012/531319.

[31] I. Reinertsen, F. Lindseth, C. Askeland, D. H. Iversen, and G. Unsgård. **Intra-operative correction of brain-shift.** In *Acta Neurochirurgica*, vol. 156, no. 7, pp. 1301–1310 (Jul. 2014). DOI: 10.1007/s00701-014-2052-6.

[32] E. Ilunga-Mbuyamba, J. G. Avina-Cervantes, D. Lindner, I. Cruz-Aceves, F. Arlt, and C. Chalopin. **Vascular structure identification in intraoperative 3D contrast-enhanced ultrasound data.** In *Sensors*, vol. 16, no. 4 (2016). DOI: 10.3390/s16040497.

[33] P. Farnia, B. Makkiabadi, A. Ahmadian, and J. Alirezaie. **Curvelet based residual complexity objective function for non-rigid registration of pre-operative MRI with intra-operative ultrasound images.** In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1167–1170 (2016). DOI: 10.1109/EMBC.2016.7590912.

[34] P. Farnia, E. Najafzadeh, A. Ahmadian, B. Makkiabadi, M. Alimohamadi, and J. Alirezaie. **Co-sparse analysis model based image registration to compensate brain shift by using intra-operative ultrasound imaging.** In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4 (2018). DOI: 10.1109/EMBC.2018.8512375.

[35] D. Drobny, T. Vercauteren, S. Ourselin, and M. Modat. **Registration of MRI and iUS data to compensate brain shift using a symmetric block-matching based approach.** In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation : International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018*, vol. 11042 (2018). DOI: 10.1007/978-3-030-01045-4_21.

[36] I. Machado, M. Toews, J. Luo, P. Unadkat, W. Essayed, E. George, P. Teodoro, H. Carvalho, J. Martins, P. Golland, S. Pieper, S. Frisken, A. Golby, and W. Wells. **Non-rigid registration of 3D ultrasound for neurosurgery using automatic feature detection and matching.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 10, pp. 1525–1538 (Oct. 2018). DOI: 10.1007/s11548-018-1786-7.

[37] I. Machado, M. Toews, E. George, P. Unadkat, W. Essayed, J. Luo, P. Teodoro, H. Carvalho, J. Martins, P. Golland, S. Pieper, S. Frisken, A. Golby, W. Wells III, and Y. Ou. **Deformable MRI-ultrasound registration using correlation-based attribute matching for brain shift correction: Accuracy and generality in multi-site data.** In *NeuroImage*, vol. 202, no. 116094 (2019). DOI: j.neuroimage.2019.116094.

[38] L. Canalini, J. Klein, D. Miller, and R. Kikinis. **Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 10, pp. 1697–1713 (Oct. 2019). DOI: 10.1007/s11548-019-02045-6.

[39] J. Nitsch, J. Klein, P. Dammann, K. Wrede, O. Gembruch, J. Moltz, H. Meine, U. Sure, R. Kikinis, and D. Miller. **Automatic and efficient MRI-US segmentations for improving intraoperative image fusion in image-guided neurosurgery.** In *NeuroImage: Clinical*, vol. 22, pp. 101766 (2019). DOI: 10.1016/j.nicl.2019.101766.

[40] T. Arbel, X. Morandi, R. M. Comeau, and D. L. Collins. **Automatic nonlinear MRI-ultrasound registration for the correction of intra-operative brain deformations.** In *Computer Aided Surgery*, vol. 9, no. 4, pp. 123–136 (2004). DOI: 10.3109/10929080500079248.

[41] T. C. W. Mok, and A. C. S. Chung. **Large deformation image registration with anatomy-aware laplacian pyramid networks.** In *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data : MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020*, vol. 12587 (2021). DOI: 10.1007/978-3-030-71827-5_7.

[42] S. Häger, S. Heldmann, A. Hering, S. Kuckertz, and A. Lange. **Variable fraunhofer mevis reglib comprehensively applied to learn2reg challenge.** In *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data : MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020*, vol. 12587 (2021). DOI: 10.1007/978-3-030-71827-5_9.

[43] W. Wein, A. Ladikos, B. Fuerst, A. Shah, K. Sharma, and N. Navab. **Global registration of ultrasound to MRI using the LC2 metric for enabling neurosurgical guidance.** In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pp. 34–41 (2013). DOI: 10.1007/978-3-642-40811-3_5.

[44] L. Hansen, and M. P. Heinrich. **Discrete unsupervised 3D registration methods for the learn2reg challenge.** In *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data : MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020*, vol. 12587 (2021). DOI: 10.1007/978-3-030-71827-5_8.

[45] F. Morin, H. Courtecuisse, I. Reinertsen, F. Le Lann, O. Palombi, Y. Payan, and M. Chabanas. **Brain-shift compensation using intraoperative ultrasound and constraint-based biomechanical simulation.** In *Medical Image Analysis*, vol. 40, pp. 133–153 (2017). DOI: 10.1016/j.media.2017.06.003.

[46] A. Myronenko, and X. Song. **Intensity-based image registration by minimizing residual complexity.** In *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, pp. 1882–1891 (2010). DOI: 10.1109/TMI.2010.2053043.

[47] S. Frisken, M. Luo, P. Juvekar, A. Bunevicius, I. Machado, P. Unadkat, M. M. Bertotti, M. Toews, W. M. Wells, M. I. Miga, and A. J. Golby. **A comparison of thin-plate spline deformation and finite element modeling to compensate for brain shift during tumor resection.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 1, pp. 75–85 (Jan. 2020). DOI: 10.1007/s11548-019-02057-2.

[48] X. Zhong, S. Bayer, N. Ravikumar, N. Strobel, A. Birkhold, M. Kowarschik, R. Fahrig, and A. Maier. **Resolve intraoperative brain shift as imitation game.** In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation : International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018*, vol. 11042 (2018). DOI: 10.1007/978-3-030-01045-4_15.

[49] L. Sun, and S. Zhang. **Deformable MRI-ultrasound registration using 3D convolutional neural network.** In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation : International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018*, vol. 11042 (2018). DOI: 10.1007/978-3-030-01045-4_15.

[50] N. Gunnarsson, J. Sjölund, and T. B. Schön. **Learning a deformable registration pyramid.** In *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data : MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020*, vol. 12587 (2021). DOI: 10.1007/978-3-030-71827-5_10.

[51] M. I. Miga, D. W. Roberts, F. E. Kennedy, L. A. Platenik, A. Hartov, K. E. Lunn, and K. D. Paulsen. **Modeling of retraction and resection for intraoperative updating of images.** In *Neurosurgery*, vol. 49, no. 1, pp. 75–85 (2001). DOI: 10.1097/00006123-200107000-00012.

[52] M. Ferrant, A. Nabavi, B. Macq, P. Black, F. A. Jolesz, R. Kikinis, and S. K. Warfield. **Serial registration of intraoperative MR images of the brain.** In *Medical Image Analysis*, vol. 6, no. 4, pp. 337–359 (2002). DOI: 10.1016/S1361-8415(02)00060-9.

[53] G. Soza, R. Grosso, U. Labsik, C. Nimsky, R. Fahlbusch, G. Greiner, and P. Hastreiter. **Fast and adaptive finite element approach for modeling brain shift.** In *Computer Aided Surgery*, vol. 8, no. 5, pp. 241–246 (2003). DOI: 10.3109/10929080309146059.

[54] H. Sun, K. E. Lunn, H. Farid, Z. Wu, D. W. Roberts, A. Hartov, and K. D. Paulsen. **Stereopsis-guided brain shift compensation.** In *IEEE Transactions on Medical Imaging*, vol. 24, no. 8, pp. 1039–1052 (Aug. 2005). DOI: 10.1109/TMI.2005.852075.

[55] O. Clatz, H. Delingette, I.-F. Talos, A. J. Golby, R. Kikinis, F. A. Jolesz, N. Ayache, and S. K. Warfield. **Robust nonrigid registration to capture brain shift from intraoperative MRI.** In *IEEE transactions on medical imaging*, vol. 24, no. 11, pp. 1417–1427 (Nov. 2005). DOI: 10.1109/TMI.2005.856734.

[56] N. Archip, O. Clatz, S. Whalen, D. Kacher, A. Fedorov, A. Kot, N. Chrisochoides, F. Jolesz, A. Golby, P. M. Black, and S. K. Warfield. **Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery.** In *NeuroImage*, vol. 35, no. 2, pp. 609–624 (Apr. 2007). DOI: 10.1016/j.neuroimage.2006.11.060.

[57] A. Wittek, T. Hawkins, and K. Miller. **On the unimportance of constitutive models in computing brain deformation for image-guided surgery.** In *Biomechanics and modeling in mechanobiology*, vol. 8, no. 1, pp. 77–84 (Feb. 2009). DOI: 10.1007/s10237-008-0118-1.

[58] P. Dumpuri, R. C. Thompson, A. Cao, S. Ding, I. Garg, B. M. Dawant, and M. I. Miga. **A fast and efficient method to compensate for brain shift for tumor resection therapies measured between preoperative and postoperative tomograms.** In *IEEE transactions on bio-medical engineering*, vol. 57, no. 6, pp. 1285–1296 (Jun. 2010). DOI: 10.1109/TBME.2009.2039643.

[59] I. Chen, A. M. Coffey, S. Ding, P. Dumpuri, B. M. Dawant, R. C. Thompson, and M. I. Miga. **Intraoperative brain shift compensation: Accounting for dural septa.** In *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 499–508 (2011). DOI: 10.1109/TBME.2010.2093896.

[60] M. Bucki, O. Palombi, M. Bailet, and Y. Payan. **Doppler ultrasound driven biomechanical model of the brain for intraoperative brain-shift compensation: A proof of**

concept in clinical conditions. In book *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*, pp. 135–165 (2012). DOI: 10.1007/8415_2012_119.

[61] L. M. Vigneron, L. Noels, S. K. Warfield, J. G. Verly, and P. A. Robe. **Serial fem/xfem-based update of preoperative brain images using intraoperative MRI.** In *International Journal of Biomedical Imaging*, vol. 2012 (2012). DOI: 10.1155/2012/872783.

[62] C. DeLorenzo, X. Papademetris, L. H. Staib, K. P. Vives, D. D. Spencer, and J. S. Duncan. **Volumetric intraoperative brain deformation compensation: model development and phantom validation.** In *IEEE transactions on medical imaging*, vol. 31, no. 8, pp. 1607–1619 (Aug. 2012). DOI: 10.1109/TMI.2012.2197407.

[63] K. Sun, T. S. Pheiffer, A. L. Simpson, J. A. Weis, R. C. Thompson, and M. I. Miga. **Near real-time computer assisted surgery for brain shift correction using biomechanical models.** In *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 2, pp. 1–13 (2014). DOI: 10.1109/JTEHM.2014.2327628.

[64] A. Mohammadi, A. Ahmadian, A. D. Azar, A. D. Sheykh, F. Amiri, and J. Alirezaie. **Estimation of intraoperative brain shift by combination of stereovision and doppler ultrasound: phantom and animal model study.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 11, pp. 1753–1764 (Nov. 2015). DOI: 10.1007/s11548-015-1216-z.

[65] A. Wittek, R. Kikinis, S. K. Warfield, and K. Miller. **Brain shift computation using a fully nonlinear biomechanical model.** In *Medical image computing and computer-assisted intervention (MICCAI)*, vol. 8, no. pt 2, pp. 583–590 (2005). DOI: 10.1007/11566489_72.

[66] M. Kochan, M. Modat, T. Vercauteren, M. White, L. Mancini, G. P. Winston, A. W. McEvoy, J. S. Thornton, T. Yousry, J. S. Duncan, S. Ourselin, and D. Stoyanov. **Bilateral weighted adaptive local similarity measure for registration in neurosurgery.** In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*, pp. 81–88 (2016). DOI: 10.1007/978-3-319-46726-9_10.

[67] L. Canalini, J. Klein, D. Miller, and R. Kikinis. **Enhanced registration of ultrasound volumes by segmentation of resection cavity in neurosurgical procedures.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 12, pp. 1963–1974 (Dec. 2020). DOI: 10.1007/s11548-020-02273-1.

[68] I. Despotović, B. Goossens, and W. Philips. **MRI segmentation of the human brain: Challenges, methods, and applications.** In *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 450341 (Mar. 2015). DOI: 10.1155/2015/450341.

[69] G. Mirzaei, and H. Adeli. **Segmentation and clustering in brain MRI imaging.** In *Reviews in the Neurosciences*, vol. 30, no. 1, pp. 31–44 (2019). DOI: doi:10.1515/revneuro-2018-0050.

[70] A. Wadhwa, A. Bhardwaj, and V. Singh Verma. **A review on brain tumor segmentation of MRI images.** In *Magnetic Resonance Imaging*, vol. 61, pp. 247–259 (2019). DOI: 10.1016/j.mri.2019.05.043.

[71] B. H. Menze, et al. **The multimodal brain tumor image segmentation benchmark (BRATS).** In *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024 (Oct. 2015). DOI: 10.1109/TMI.2014.2377694.

[72] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. **A comparison of retrospective intensity non-uniformity correction methods for MRI.** In *Information Processing in Medical Imaging*, pp. 459–464 (1997). DOI: 10.1007/3-540-63046-5_43.

[73] J.-W. Liu, and L. Guo. **Selection of initial parameters of K-means clustering algorithm for MRI brain image segmentation.** In *2015 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 1, pp. 123–127 (2015). DOI: 10.1109/ICMLC.2015.7340909.

[74] E. Abdel-Maksoud, M. Elmogy, and R. Al-Awadi. **Brain tumor segmentation based on a hybrid clustering technique.** In *Egyptian Informatics Journal*, vol. 16, no. 1, pp. 71–81 (2015). DOI: 10.1016/j.eij.2015.01.003.

[75] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal. **Fuzzy models and algorithms for pattern recognition and image processing.** The Handbooks of Fuzzy Sets Series (1999). DOI: 10.1007/b106267.

[76] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, and T. Moriarty. **A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data.** In *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199 (2002). DOI: 10.1109/42.996338.

[77] D. Rivest-Hénault, and M. Cheriet. **Unsupervised MRI segmentation of brain tissues using a local linear model and level set.** In *Magnetic Resonance Imaging*, vol. 29, no. 2, pp. 243–259 (2011). DOI: 10.1016/j.mri.2010.08.007.

[78] A.Rajendran, and R. Dhanasekaran. **Fuzzy clustering and deformable model for tumor segmentation on MRI brain image: A combined approach.** In *Procedia Engineering*, vol. 30, pp. 327–333 (2012). DOI: 10.1016/j.proeng.2012.01.868.

[79] B. Fischl. **FreeSurfer.** In *NeuroImage*, vol. 62, no. 2, pp. 774–781 (2012). DOI: 10.1016/j.neuroimage.2012.01.021.

[80] D. C. Van Essen. **Cortical cartography and Caret software.** In *NeuroImage*, vol. 62, no. 2, pp. 757–764 (2012). DOI: 10.1016/j.neuroimage.2011.10.077.

[81] J. Zhong, D. Y. L. Phua, and A. Qiu. **Quantitative evaluation of LDDMM, FreeSurfer, and CARET for cortical surface mapping.** In *NeuroImage*, vol. 52, no. 1, pp. 131–141 (2010). DOI: 10.1016/j.neuroimage.2010.03.085.

[82] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore. **A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI.** In *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2007–2016 (2011). DOI: 10.1109/TIP.2011.2146190.

[83] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans. **A fully automatic and robust brain MRI tissue classification method.** In *Medical image analysis*, vol. 7, pp. 513–527. DOI: 10.1016/s1361-8415(03)00037-9.

[84] M. D. Steenwijk, P. J. Pouwels, M. Daams, J. W. van Dalen, M. W. Caan, E. Richard, F. Barkhof, and H. Vrenken. **Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs).** In *NeuroImage: Clinical*, vol. 3, pp. 462–469 (2013). DOI: 10.1016/j.nicl.2013.10.003.

[85] S. Bauer, L.-P. Nolte, and M. Reyes. **Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization.** In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, pp. 354–361 (2011). DOI: 10.1007/978-3-642-23626-6_44.

[86] W. Wu, A. Y. C. Chen, L. Zhao, and J. J. Corso. **Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 9, no. 2, pp. 241–253 (Mar. 2014). DOI: 10.1007/s11548-013-0922-7.

[87] N. Zhang, S. Ruan, S. Lebonvallet, Q. Liao, and Y. Zhu. **Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation.** In *Computer Vision and Image Understanding*, vol. 115, no. 2, pp. 256–269 (2011). DOI: 10.1016/j.cviu.2010.09.007.

[88] N. J. Tustison, K. L. Shrinidhi, M. Wintermark, C. R. Durst, B. M. Kandel, J. C. Gee, M. C. Grossman, and B. B. Avants. **Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR.** In *Neuroinformatics*, vol. 13, no. 2, pp. 209–225 (Apr. 2015). DOI: 10.1007/s12021-014-9245-2.

[89] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye. **Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair MRI.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 2, pp. 183–203 (Feb. 2017). DOI: 10.1007/s11548-016-1483-3.

[90] V. Damerjian, O. Tankyevych, N. Souag, and E. Petit. **Speckle characterization methods in ultrasound images – a review.** In *IRBM*, vol. 35, no. 4, pp. 202–213 (2014). DOI: 10.1016/j.irbm.2014.05.003.

[91] C. A. Duarte-Salazar, A. E. Castro-Ospina, M. A. Becerra, and E. Delgado-Trejos. **Speckle noise reduction in ultrasound images for improving the metrological evaluation of biomedical applications: An overview.** In *IEEE Access*, vol. 8, pp. 15983–15999 (2020). DOI: 10.1109/ACCESS.2020.2967178.

[92] P. C. Tay, C. D. Garson, S. T. Acton, and J. A. Hossack. **Ultrasound despeckling for contrast enhancement.** In *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1847–1860 (2010). DOI: 10.1109/TIP.2010.2044962.

[93] Y. Yu, and S. Acton. **Speckle reducing anisotropic diffusion.** In *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1260–1270 (2002). DOI: 10.1109/TIP.2002.804276.

[94] K. Krissian, C.-F. Westin, R. Kikinis, and K. G. Vosburgh. **Oriented speckle reducing anisotropic diffusion.** In *IEEE Transactions on Image Processing*, vol. 16, no. 5, pp. 1412–1424 (2007). DOI: 10.1109/TIP.2007.891803.

[95] Y. Yue, M. Croitoru, A. Bidani, J. Zwischenberger, and J. Clark. **Nonlinear multiscale wavelet diffusion for speckle suppression and edge enhancement in ultrasound images.** In *IEEE Transactions on Medical Imaging*, vol. 25, no. 3, pp. 297–311 (2006). DOI: 10.1109/TMI.2005.862737.

[96] A. Khare, M. Khare, Y. Jeong, H. Kim, and M. Jeon. **Despeckling of medical ultrasound images using Daubechies complex wavelet transform.** In *Signal Processing*, vol. 90, no. 2, pp. 428–439 (2010). DOI: 10.1016/j.sigpro.2009.07.008.

[97] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot. **Nonlocal means-based speckle filtering for ultrasound images.** In *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2221–2229 (2009). DOI: 10.1109/TIP.2009.2024064.

[98] L. H. Breivik, S. R. Snare, E. N. Steen, and A. H. S. Solberg. **Real-time nonlocal means-based despeckling.** In *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 6, pp. 959–977 (2017). DOI: 10.1109/TUFFC.2017.2686326.

[99] D. Mishra, S. Chaudhury, M. Sarkar, and A. S. Soin. **Ultrasound image enhancement using structure oriented adversarial network.** In *IEEE Signal Processing Letters*, vol. 25, no. 9, pp. 1349–1353 (2018). DOI: 10.1109/LSP.2018.2858147.

[100] F. Dietrichson, E. Smistad, A. Ostvik, and L. Lovstakken. **Ultrasound speckle reduction using generative adversial networks.** In *2018 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4 (2018). DOI: 10.1109/ULTSYM.2018.8579764.

[101] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl. **Beamforming and speckle reduction using neural networks.** In *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 5, pp. 898–910 (2019). DOI: 10.1109/TUFFC.2019.2903795.

[102] J. A. Noble, and D. Boukerroui. **Ultrasound image segmentation: a survey.** In *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, pp. 987–1010 (Aug. 2006). DOI: 10.1109/TMI.2006.877092.

[103] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang. **Deep learning in medical ultrasound analysis: A review.** In *Engineering*, vol. 5, no. 2, pp. 261–275 (2019). DOI: 10.1016/j.eng.2018.11.020.

[104] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, and N. Navab. **Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound.** In *Computer Vision and Image Understanding*, vol. 164, pp. 92–102 (2017). DOI: 10.1016/j.cviu.2017.04.002.

[105] K. Ritschel, I. Pechlivanis, and S. Winter. **Brain tumor classification on intraoperative contrast-enhanced ultrasound.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 5, pp. 531–540 (May. 2015). DOI: 10.1007/s11548-014-1089-6.

[106] E. Ilunga-Mbuyamba, J. G. Avina-Cervantes, D. Lindner, F. Arlt, J. F. Ituna-Yudonago, and C. Chalopin. **Patient-specific model-based segmentation of brain tumors in 3D intraoperative ultrasound images.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 3, pp. 331–342 (Mar. 2018). DOI: 10.1007/s11548-018-1703-0.

[107] D. P. Kingma, and J. L. Ba. **Adam: A method for stochastic optimization.** In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (2015). DOI: 11245/1.505367.

[108] M. H. Hesamian, W. Jia, X. He, and P. Kennedy. **Deep learning techniques for medical image segmentation: Achievements and challenges.** In *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596 (Aug. 2019). DOI: 10.1007/s10278-019-00227-x.

[109] E. Shelhamer, J. Long, and T. Darrell. **Fully convolutional networks for semantic segmentation.** In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651 (2017). DOI: 10.1109/TPAMI.2016.2572683.

[110] O. Ronneberger, P. Fischer, and T. Brox. **U-Net: Convolutional networks for biomedical image segmentation.** In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241 (2015). DOI: 10.1007/978-3-319-24574-4_28.

[111] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel. **Loss odyssey in medical image segmentation.** In *Medical Image Analysis*, vol. 71, pp. 102035 (2021). DOI: 10.1016/j.media.2021.102035.

[112] L. R. Dice. **Measures of the amount of ecologic association between species.** In *Ecology*, vol. 26, no. 3, pp. 297–302. DOI: 10.2307/1932409.

[113] F. Milletari, N. Navab, and S. Ahmadi. **V-Net: Fully convolutional neural networks for volumetric medical image segmentation.** In *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571 (Oct. 2016). DOI: 10.1109/3DV.2016.79.

[114] D. Karimi, and S. E. Salcudean. **Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks.** In *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 499–513 (Feb. 2020). DOI: 10.1109/TMI.2019.2930068.

[115] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen. **Deformable image registration based on similarity-steered CNN regression.** In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pp. 300–308 (2017). DOI: 10.1007/978-3-319-66182-7_35.

[116] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen. **Robust non-rigid registration through agent-based action learning.** In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pp. 344–352 (2017). DOI: 10.1007/978-3-319-66182-7_40.

[117] H. Sokooti, B. de Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring. **Non-rigid image registration using multi-scale 3D convolutional neural networks.** In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pp. 232–239 (2017). DOI: 10.1007/978-3-319-66182-7_27.

[118] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum. **End-to-end unsupervised deformable image registration with a convolutional neural network.** In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 204–212 (2017). DOI: 10.1007/978-3-319-67558-9_24.

[119] H. Li, and Y. Fan. **Non-rigid image registration using self-supervised fully convolutional networks without training data.** In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1075–1078 (Apr. 2018). DOI: 10.1109/ISBI.2018.8363757.

[120] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum. **A deep learning framework for unsupervised affine and deformable image registration.** In *Medical Image Analysis*, vol. 52, pp. 128–143 (2019). DOI: 10.1016/j.media.2018.11.010.

[121] M. Blendowski, and M. P. Heinrich. **Combining MRF-based deformable registration and deep binary 3D-CNN descriptors for large lung motion estimation in COPD patients.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 1, pp. 43–52 (Jan. 2019). DOI: 10.1007/s11548-018-1888-2.

[122] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec. **SVF-Net: Learning deformable image registration using shape matching.** In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pp. 266–274 (2017). DOI: 10.1007/978-3-319-66182-7_31.

[123] X. Yang, R. Kwitt, M. Styner, and M. Niethammer. **Quicksilver: Fast predictive image registration – a deep learning approach.** In *NeuroImage*, vol. 158, pp. 378–396 (2017). DOI: 10.1016/j.neuroimage.2017.07.008.

[124] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, S. Ourselin, J. A. Noble, D. C. Barratt, and T. Vercauteren. **Weakly-supervised convolutional neural networks for multimodal image registration.** In *Medical Image Analysis*, vol. 49, pp. 1–13 (2018). DOI: 10.1016/j.media.2018.07.002.

[125] 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). DOI: 10.1109/TMI.2019.2897538.

[126] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi. **Learning a probabilistic model for diffeomorphic registration.** In *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2165–2176 (Sep. 2019). DOI: 10.1109/TMI.2019.2897112.

[127] J. Fan, X. Cao, P.-T. Yap, and D. Shen. **BIRNet: Brain image registration using dual-supervised fully convolutional networks.** In *Medical Image Analysis*, vol. 54, pp. 193–206 (2019). DOI: 10.1016/j.media.2019.03.006.

[128] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi. **Deformable medical image registration using generative adversarial networks.** In *2018 IEEE 15th Interna-*

tional Symposium on Biomedical Imaging (ISBI 2018), pp. 1449–1453 (Apr. 2018). DOI: 10.1109/ISBI.2018.8363845.

[129] J. Fan, X. Cao, Q. Wang, P.-T. Yap, and D. Shen. **Adversarial learning for mono- or multi-modal registration.** In *Medical Image Analysis*, vol. 58, pp. 101545 (2019). DOI: 10.1016/j.media.2019.101545.

[130] C. Qin, B. Shi, R. Liao, T. Mansi, D. Rueckert, and A. Kamen. **Unsupervised deformable registration for multi-modal images via disentangled representations.** In *Information Processing in Medical Imaging*, pp. 249–261 (2019). DOI: 10.1007/978-3-030-20351-1_19.

[131] M. Heinrich. **Closing the gap between deep and conventional image registration using probabilistic dense displacement networks.** In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pp. 50–58 (2017). DOI: 10.1007/978-3-030-32226-7_6.

[132] D. Stoyanov, Z. Taylor, S. Aylward, J. Tavares, Y. Xiao, A. Simpson, A. Martel, L. Maier-Hein, S. Li, H. Rivaz, I. Reinertsen, M. Chabanas, and F. K. **Simulation, image processing, and ultrasound systems for assisted diagnosis and navigation : International workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018, held in conjunction with MICCAI 2018, Granada, Spain.** In vol. 11042 of *Lecture Notes in Computer Science* (2018). DOI: 10.1007/978-3-030-01045-4.

[133] Y. Xiao, H. Rivaz, M. Chabanas, M. Fortin, I. Machado, Y. Ou, M. Heinrich, J. Schnabel, X. Zhong, A. Maier, W. Wein, R. Shams, S. Kadoury, D. Drobny, M. Modat, and I. Reinertsen. **Evaluation of MRI to ultrasound registration methods for brain shift correction: The CuRIOUS2018 challenge.** In *IEEE Transactions on Medical Imaging* (2020). DOI: 10.1109/TMI.2019.2935060.

[134] N. Shusharina, M. P. Heinrich, and R. Huang. **Segmentation, classification, and registration of multi-modality medical imaging data : MICCAI 2020 challenges, ABCs 2020, L2R 2020, TN-SCUI 2020, held in conjunction with MICCAI 2020, Lima, Peru.** In vol. 12587 of *Lecture Notes in Computer Science* (2021). DOI: 10.1007/978-3-030-71827-5.

[135] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz. **PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume.** In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8934–8943 (2018). DOI: 10.1109/CVPR.2018.00931.

[136] L. Mercier, R. F. Del 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, no. 6 part 1, pp. 3253–3261 (Jun. 2012). DOI: 10.1118/1.4709600.

[137] 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*, vol. 44, no. 7, pp. 3875–3882 (2017). DOI: 10.1002/mp.12268.

[138] J. Rackerseder, M. Baust, R. Göbl, N. Navab, and C. Hennersperger. **Initialize globally before acting locally: Enabling landmark-free 3D US to MRI registration.** In *Medical*

Image Computing and Computer Assisted Intervention – MICCAI 2018, pp. 827–835 (2018). DOI: 10.1007/978-3-030-00928-1_93.

[139] L. Canalini, J. Klein, D. Miller, and R. Kikinis. **Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery.** In *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 1697–1713 (2019). DOI: 10.1007/s11548-019-02045-6.

[140] L. Fidon, W. Li, L. C. Garcia-Peraza-Herrera, J. Ekanayake, N. Kitchen, S. Ourselin, and T. Vercauteren. **Generalised Wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks.** In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 64–76 (2018). DOI: 10.1007/978-3-319-75238-9_6.

[141] K. Dmitriev, and A. E. Kaufman. **Learning multi-class segmentations from single-class datasets.** In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9493–9503 (Jun. 2019). DOI: 10.1109/CVPR.2019.00973.

[142] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig. **User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability.** In *Neuroimage*, vol. 31, no. 3, pp. 1116–1128 (2006). DOI: 10.1016/j.neuroimage.2006.01.015.

[143] F. Bookstein. **Principal warps: thin-plate splines and the decomposition of deformations.** In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585 (1989). DOI: 10.1109/34.24792.

[144] J. Dolz, I. Ben Ayed, and C. Desrosiers. **Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities.** In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 271–282 (2019). DOI: 10.1007/978-3-030-11723-8_27.

Image segmentation and registration using machine learning for brain shift compensation in image-guided neurosurgery

Abstract

In brain tumor resection surgery, a preoperative magnetic resonance (MR) image is used for surgical planning and guidance. Brain tissue displacement during surgery, known as brain shift, severely limits image guidance with the preoperative MR (pMR), which no longer reflects the location of tissue and anatomical structures accurately. This can be mitigated by registering the pMR with intraoperative data, ie. updating the pMR so that it matches images acquired during surgery. Ultrasound (US) imaging has several benefits compared to other intraoperative imaging modalities such as MR. In particular, it has fast acquisition times and do not require moving the patients, which makes the surgery longer. Thus, there is growing interest in using US imaging in brain surgery, despite the low image quality.

While traditional registration methods have been proposed, deep learning (DL) are being actively researched as they have proven successful and achieve state-of-the-art results in many domains such as medical image segmentation, and more recently, image registration. DL models are complex models with parameters learned with data, with many applications. However, there are few studies proposing DL models for segmentation of intraoperative US (iUS) images of the brain, or for registration of the pMR image with iUS images.

We start with the segmentation of iUS images of the brain. This is can be used for several purposes, such as guiding a registration model, which is the main objective of this thesis. We first focus on the segmentation of the resection cavity, which is essential to model the tissue displacement due to tissue removal. Next, we address tumor tissue segmentation, which is more challenging but can be used not only for registration but also to determine if tissue resection is complete. The results show that DL models can successfully segment the resection cavity and tumor tissue in US images. Then, we train a multi-class segmentation model, that segments several structures at the same time. The results suggest that multi-class models can be more accurate than single-class models, and leverage inter-class relationships.

We then investigate DL models for registration of pMR with iUS images. Image registration is inherently difficult, and especially with different modalities such as MR and US. In addition, there is a high variability across cases, as different structures are visible in the limited US field of view, depending on tumor location. It is also difficult to estimate ground truth displacements for training DL models. While we were not able to train a registration model with the available data, we discuss our findings and possible areas of improvement.

DL models can successfully segment iUS images. We currently lack data to train registration models but expect that larger datasets will enable the training of such models.

Keywords : Brain-shift, image-guided neurosurgery, image segmentation, image registration, deep learning.

Résumé

Dans le cadre des chirurgies d'ablation de tumeurs cérébrales, un examen préopératoire par imagerie par résonance magnétique (IRM) est réalisé pour la planification chirurgicale et le guidage lors de la chirurgie. Le déplacement des tissus cérébraux pendant la chirurgie, appelé communément *brain shift*, limite considérablement le guidage par image utilisant l'IRM préopératoire (pMR), celui-ci ne représentant plus la position courante des tissus. Ces limitations peuvent être compensées en recalant le pMR avec des images peropératoires. L'échographie présente plusieurs avantages à d'autres modalités comme l'IRM, notamment sa rapidité d'acquisition, et la non-nécessité de déplacer les patients. L'échographie a donc un réel intérêt en neurochirurgie, malgré la moindre qualité de ces images.

Bien que des méthodes traditionnelles de recalage aient été proposées, l'apprentissage profond (AP) fait l'objet de nombreux travaux de recherche car il obtient les meilleurs résultats dans plusieurs domaines comme le traitement d'images médicales, notamment la segmentation et plus récemment, le recalage d'image. Les modèles d'AP sont des modèles complexes dont les paramètres sont appris à partir de données. Cependant, il y a peu d'études sur leur applicabilité à la segmentation d'images échographiques, ou au recalage avec un pMR.

L'étude porte tout d'abord sur la segmentation des images échographiques peropératoires (iUS) du cerveau. Ces segmentations peuvent notamment servir au recalage d'image, l'objectif principal de cette thèse. Nos travaux portent premièrement sur la segmentation des cavités de la résection, permettant de modéliser précisément le décalage des tissus lié à l'ablation de tissus. Nous proposons ensuite un modèle pour la segmentation des tumeurs, problème plus difficile mais utile non seulement pour le recalage d'image mais aussi pour déterminer si la résection est complète. Les résultats obtenus montrent que l'AP peut être utilisé pour la segmentation d'iUS. Nous avons ensuite entraîné un modèle multi-classes, segmentant directement plusieurs structures. Les résultats suggèrent que ces modèles multi-classes peuvent produire de meilleures segmentations, tenant compte des dépendances entre les classes.

Puis nous étudions les modèles de recalage des pMRs avec des iUSs. Ce problème est considérablement plus compliqué que la segmentation d'image, en particulier entre différentes modalités comme l'IRM et l'échographie. Il y a de plus une grande variabilité entre patients, les structures visibles dans le champ de vue limité des iUSs dépendant de la position de la tumeur. Il est aussi difficile de déterminer les vrais déplacements de tissus avec lesquels entraîner des modèles. Bien que n'étant pas parvenus à entraîner un modèle avec les données disponibles, nous analysons les résultats et identifions de potentielles pistes d'améliorations.

Les modèles d'AP peuvent segmenter les iUS. Les données sont actuellement insuffisantes pour des modèles de recalage, que plus de données permettraient d'entraîner.

Mots-clés : Brain-shift, neurochirurgie guidée par image, segmentation d'image, recalage d'image, apprentissage profond.