Numerical methods on medical images for per-operative dose evaluation in electroporation ablation

Eloïse Inacio

▶ **To cite this version:**

Eloïse Inacio. Numerical methods on medical images for per-operative dose evaluation in electroporation ablation. Medical Imaging. Université de Bordeaux, 2024. English. NNT : 2024BORD0474 . tel-04903951

HAL Id: tel-04903951

https://theses.hal.science/tel-04903951v1

Submitted on 21 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

université de BORDEAUX

Inria

Thèse présentée
pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

Ecole doctorale de Mathématiques et d'Informatique
Spécialité : Mathématiques Appliquées et Calcul Scientifique

Par Eloïse INACIO

Méthodes numériques en imagerie médicale pour l'évaluation de dose per-opératoire en ablation par électroporation

Sous la direction de :

Dr. Baudouin DENIS DE SENNEVILLE & Dr. Clair POIGNARD

Soutenue le 16 décembre 2024

Membres du jury :

Pr. François CORNÉLIS, Professor, Memorial Sloan Kettering Cancer Center, Rapporteur

Pr. Amandine CROMBÉ, PU-PH, CHU Pellegrin, Présidente du jury

Dr. Baudouin DENIS DE SENNEVILLE, DR, Université de Bordeaux, Directeur de thèse

Dr. Abdallah EL HAMIDI, Enseignant-Chercheur, La Rochelle Université, Rapporteur

Dr. Clair POIGNARD, DR, INRIA, Co-directeur de thèse

Dr. Damien VOYER, Enseignant-Chercheur, EIGSI, Examinateur

Dr Emeline RIBOT, CR, Université de Bordeaux, Invitée

Acknowledgment

I would like to thank both my supervisors, Dr. Baudouin Denis de Senneville and Dr. Clair Poignard, for guiding me through those enriching years. Not only have you advised me through the ups and downs of mathematics, but you have been of great help and support for my carreer choices.

I would also like to thank some of my fellow PhD students: Virginie Montalibet, Pedro Jaramillo Aguayo, Kylian Desier, Khaoula Chadi ; and more specifically those with whom I shared an office, the two Simons, Simone Nati Poltri and Simon Bihoreau. Thank you for lending me your best advice on numerical analysis, always an attentive ear, and sometimes even a shoulder to cry on. The PhD years can be a difficult experience but I believe we made the most of it together.

Thanks to my parents for supporting my studies, even when I chose to go abroad to learn English. More specifically, thanks to my mother for sharing her passion for mathematics and image processing.

I thank my dear friend Dean Gunning for all the emotional and technical support, especially with LateX... You have always found time to help me and listen to my rants. I wish you the best for your own PhD which I am convinced will be very successful.

Last but not least, a million thanks to Kylian for all the love and the support you have given me through the years. You carried me single handedly most of the way and I don't think I could have done this without you. I can only hope to be as good a partner as you are when you embark on your own professional journey.

Abstract

As life expectancy rises, cancer has tragically become one of the world's leading causes of death. Among the most challenging cancers are deep-seated tumors, which are difficult to treat due to their location in vital organs like the liver or the pancreas. A promising method to tackle these tumors is electroporation ablation, which uses electric fields to create pores in the cell membranes of tumor cells. When applied with high intensity, this results in irreversible electroporation, leading to cell death without damaging nearby structures.

However, electroporation requires precise planning and real-time adaptation due to its complexity. This involves numerical tools to analyze medical images and estimate the treatment area. The aim of this work is to provide such tools, analysing medical images, to per-operatively estimate the treatment area so that the interventional radiologists may adapt their approach as they are performing the procedure. More specifically, we tackle the localisation of the electrode by introducing deep learning in the existing pipeline, and the registration of the multiple scans captured during the intervention with novel auto-adaptive boundary conditions. Both computer vision tasks are crucial for a precise estimation of the electric field and need to be solved in near real time to be practical in clinical settings.

These advancements in computer vision and image processing contribute to more accurate electric field estimation and improve the overall effectiveness of the procedure, leading to better patient outcomes for those battling deep-seated cancers.

Keywords : Medical imaging, Variational methods, Image registration, Deep-learning, Segmentation, Dose computation

Résumé

Alors que l'espérance de vie augmente, le cancer est devenu l'une des principales causes de décès dans le monde. Parmi les cancers les plus difficiles à traiter figurent les tumeurs profondes, qui sont compliquées à soigner en raison de leur emplacement près de structures vitales dans des organes tels que le foie ou le pancréas. Une méthode prometteuse pour traiter ces tumeurs est l'ablation par électroporation, qui utilise des champs électriques pour créer des pores dans les membranes des cellules tumorales. Lorsqu'ils sont appliqués à haute intensité, cela entraîne une électroporation irréversible, conduisant à la mort des cellules sans endommager les structures avoisinantes.

Cependant, l'électroporation nécessite une planification précise et une adaptation en temps réel en raison de sa complexité. Cela implique des outils numériques pour analyser les images médicales et estimer la zone de traitement. L'objectif de ce travail est de fournir de tels outils permettant d'analyser les images médicales afin d'estimer, pendant l'opération, la zone de traitement, de manière à ce que le radiologue interventionnel puisse adapter son approche. Plus précisément, nous abordons la localisation des électrodes en introduisant l'apprentissage profond dans le programme existant, ainsi que le recalage des multiples images capturées durant l'intervention avec des conditions aux limites auto-adaptatives innovantes.

Ces deux tâches de vision par ordinateur sont cruciales pour une estimation précise du champ électrique et doivent être résolues en quasi-temps réel pour être praticables en contexte clinique. Ces avancées dans la vision par ordinateur et le traitement d'images permettent une estimation plus précise du champ électrique et améliorent l'efficacité globale de la procédure, conduisant à de meilleurs résultats pour les patients atteints de tumeurs profondes.

Mots clés : Imagerie médicale, Méthodes variationnelles, Recalage d'image, Apprentissage profond, Segmentation, Calcul de dose

Résumé étendu

Le cancer continue d'être un défi sanitaire redoutable, touchant un nombre croissant d'individus chaque année. Rien qu'aux États-Unis, plus d'un million de nouveaux cas sont prévus pour 2024, soulignant l'urgence de solutions médicales innovantes [87]. Parmi ces cas, les cancers impliquant des tumeurs profondes présentent des obstacles thérapeutiques uniques en raison de leur emplacement difficile dans le corps. Les thérapies traditionnelles peinent souvent à atteindre et à éradiquer efficacement ces tumeurs, ce qui rend indispensable le développement de nouvelles approches.

Une méthode prometteuse dans le traitement du cancer est l'ablation par électroporation irréversible (IRE). Cette technique innovante utilise un champ électrique pulsé délivré par des électrodes peu invasives pour induire la formation de nanopores dans la membrane cellulaire bilipidique. À mesure que ces pores se forment, les matériaux internes des cellules commencent à s'échapper, déclenchant finalement l'apoptose, ou mort cellulaire programmée. Contrairement aux méthodes ablatives traditionnelles, qui peuvent endommager les tissus environnants, l'IRE préserve efficacement l'intégrité structurelle de la matrice tissulaire. Cette préservation est cruciale, car elle réduit considérablement les effets secondaires souvent associés à d'autres options de traitement, permettant ainsi une récupération plus tolérable pour les patients.

Nos efforts collaboratifs avec les radiologues interventionnels de l'Hôpital Avicenne AP-HP ont été essentiels pour intégrer de manière fluide notre flux de travail numérique dans leurs pratiques cliniques, créant ainsi une synergie bénéfique pour les équipes médicales et les patients. En effet, tout au long de la procédure, l'imagerie médicale joue un rôle crucial, car elle permet non seulement de cartographier la tumeur et les structures anatomiques environnantes avec une grande précision, mais elle contribue également à la planification et à l'exécution des interventions. Cela inclut le guidage pour l'insertion précise des électrodes, ce qui est fondamental pour maximiser l'efficacité des traitements tout en minimisant les risques pour le patient. De plus, l'évaluation de la zone de traitement effective après l'ablation est un aspect essentiel pour s'assurer que la tumeur a été correctement ciblée. Cette thèse se concentre spécifiquement sur ce dernier aspect, proposant

de nouvelles méthodes pour l'estimation en ligne, soit en quelques minutes, de la zone d'ablation, ce qui pourrait améliorer significativement les résultats cliniques en permettant une adaptation de l'approche pendant l'opération-même.

Initialement, nous réalisons une tomodensitométrie à faisceau conique (CBCT) au début de l'ablation afin de faciliter la segmentation précise de la tumeur. Cette analyse est effectuée de manière semi-automatique par les radiologues à l'aide d'un algorithme avancé d'optimisation des contours. Ensuite, les aiguilles sont insérées sous guidage échographique et fluoroscopique en temps réel, garantissant un placement optimal.

Pour confirmer la localisation des aiguilles, un deuxième CBCT est réalisé. Cette répétition est cruciale, car les électrodes peuvent introduire d'importants artefacts en raison de leur densité, rendant la tumeur difficile à discerner. Après l'insertion des électrodes, des impulsions électriques sont délivrées et leurs chronogrammes sont enregistrés. Combinées aux données de l'imagerie médicale, ces informations sont essentielles pour modéliser le champ électrique délivré pendant le traitement. Si la tumeur n'est pas entièrement incluse dans la zone de traitement, les radiologues interventionnels peuvent adapter leur approche en temps réel — soit en ajoutant une séquence d'impulsions électriques, soit en ajustant la profondeur des aiguilles par une technique de retrait progressif.

Le processus de collecte des données nécessaires pour la simulation du champ électrique englobe une série de tâches de vision par ordinateur. Dans un premier temps, une segmentation précise de la tumeur est effectuée pour localiser avec exactitude les cellules cancéreuses. Ensuite, la localisation des électrodes est réalisée. Étant donné que ces électrodes servent de points d'entrée pour les impulsions électriques, leur placement précis est essentiel pour estimer correctement la distribution du champ électrique. Enfin, plusieurs scans d'imagerie doivent être recalés dans un référentiel commun.

L'intégration de méthodes numériques est essentielle pour une analyse complète et une exécution efficace du traitement. Dans ce contexte, nous présentons des solutions innovantes conçues pour améliorer le programme existant, en nous concentrant particulièrement sur l'amélioration de la localisation des aiguilles et du processus de recalage. En perfectionnant ces tâches, nous visons à optimiser la précision et la fiabilité des simulations du champ électrique, contribuant ainsi à des interventions thérapeutiques plus efficaces.

***

D'une part, nous introduisons des méthodes d'apprentissage profond, désormais reconnues comme particulièrement efficaces dans le domaine de la vision par

ordinateur, au sein de l'algorithme original adoptant une stratégie « coarse-to-fine » pour la localisation des électrodes [47].

Une architecture U-Net est personnalisée pour gérer la segmentation grossière d'objets fins dans des données ayant un faible rapport signal/bruit. Pour traiter le déséquilibre de classes inhérent à la nature des aiguilles, des ajustements sont apportés à la fois à l'architecture du réseau et à la stratégie d'apprentissage. Nous introduisons une asymétrie dans le réseau pour accorder plus d'importance aux informations de haut niveau provenant du décodeur, ce qui aide à capturer plus efficacement les détails complexes. De plus, des ensembles de patchs, sous-parties de l'image, sont créés pour contrer la sous-représentation des électrodes, avec des pourcentages variés de patchs contenant des informations pertinentes sur les aiguilles. Lorsqu'un équilibre est atteint avec 50% de patchs aiguille, la qualité de segmentation se retrouve améliorée, sans imposer une charge computationnelle importante. Cependant, le déséquilibre des classes reste un défi, et pour affiner davantage les résultats, nous optimisons le seuil d'inférence en fonction de la tendance du réseau à prédire avec une confiance accrue les voxels d'arrière-plan. Notre approche s'avère plus efficace que le célèbre nnU-Net [48], probablement en raison de la nature hautement spécialisée de notre stratégie d'entraînement adaptée à ce problème.

Le masque de segmentation généré par les méthodes d'apprentissage profond est ensuite affiné à l'aide de la transformation de Hough pour obtenir une représentation analytique de la partie active des aiguilles, invisible sur le CBCT à cause des artefacts. Les électrodes sont modélisées comme des lignes, en utilisant la paramétrisation optimale de ligne de Robert. Bien que le processus d'insertion puisse introduire une légère courbure dans les aiguilles, ce modèle reste une approximation valable pour localiser avec précision les pointes des aiguilles, d'où les impulsions électriques sont appliquées.

Pour améliorer la détection, une procédure de vote identifie les lignes pertinentes représentant les aiguilles. Lorsqu'elle est combinée avec l'apprentissage profond, cette approche hybride montre d'excellentes performances, offrant une précision et une fiabilité supérieures par rapport aux méthodes basées sur le seuillage utilisées auparavant. Notre technique détecte avec succès des aiguilles qui avaient été précédemment manquées, ce qui la rend particulièrement efficace dans des cas difficiles.

***

D'autre part, en abordant les défis des conditions aux limites dans les tâches de recalage, nous soulignons un problème critique souvent négligé : l'impact d'une information insuffisante aux bords de l'image, qui peut entraîner des erreurs de

recalage se propageant vers l'intérieur. Traditionnellement, les solutions se sont appuyées sur des conditions aux limites standards—spécifiquement les conditions de Dirichlet homogène (valeurs nulles aux bords) ou de Neumann homogène (tenseur de cisaillement nul)—ou sur des méthodes spécifiques à la tâche.

Pour faire progresser ce domaine, nous proposons une condition aux limites de type Robin automatiquement adaptable, qui exploite des cartes de champ de flux dérivées des images à recaler [46]. Notre approche simplifie l'espace des hyper-paramètres, le réduisant à deux paramètres clés. Le premier paramètre joue un rôle crucial en équilibrant les conditions de Neumann et de Dirichlet, tandis que le second ajuste le poids du terme source. Pour déterminer les valeurs optimales de ces hyper-paramètres, nous avons recours à une recherche en grille, visant à minimiser l'énergie de recalage. Nous démontrons que cette minimisation est étroitement liée à une maximisation de la qualité du recalage, en confrontant nos résultats aux données de réalité terrain disponibles. Cependant, il est important de souligner que cette méthode, bien que prometteuse, présente des limitations dans un contexte clinique. Les contraintes temporelles souvent imposées par les environnements médicaux rendent son application difficile, ce qui soulève la nécessité d'explorer des alternatives plus adaptées aux exigences du milieu clinique.

Par exemple, nous avons observé que des conditions de capture analogues—par la modalité et la zone anatomique—permettent une réutilisation des mêmes hyper-paramètres. Cette approche simplifie non seulement l'optimisation initiale, mais ouvre également la porte à d'éventuels raffinements lorsque le temps le permet. Autrement, l'adoption de méthodes d'apprentissage profond, connues pour leur efficacité dans les tâches de recalage, pourrait réduire considérablement le temps d'inférence. Bien que nous ayons validé notre méthode de conditions aux limites à l'aide d'une méthode variationnelle, elle est facilement généralisable, ouvrant la voie à une efficacité et à une précision accrues dans les applications cliniques.

Additionnellement, nous montrons que les conditions aux bords imposées lors de la tâche de recalage ont un impact direct sur l'évaluation du champs électrique dans le cadre d'une ablation de tumeur par IRE. En effet, différentes conditions aux bords mènent à des zones de traitement variant de manière significante cliniquement parlant. Entre autre, les volumes traités diffèrent au moins de 33mm^3. Un traitement inadéquat ou incomplet de la zone cible pourrait ainsi permettre la persistance ou la reformation de la tumeur après l'intervention.

***

Les techniques proposées dans cette thèse pour améliorer la simulation du champ électrique reposent sur l'intégration de l'apprentissage profond pour la segmentation des aiguilles et des conditions aux bords auto-adaptatives pour la tâche de recalage.

Ces technologies permettent de générer des estimations rapides et précises, en ligne. Grâce à ces algorithmes, il est possible de calculer en environ deux minutes sur un ordinateur standard des résultats cliniquement exploitables.

Le recours à l'apprentissage profond apporte des solutions robustes pour traiter la segmentation d'image, tout en s'adaptant dynamiquement aux conditions variables. Cela garantit non seulement une précision accrue, mais aussi une conformité aux contraintes cliniques imposées, notamment en matière de rapidité, de fiabilité et d'accessibilité des outils utilisés dans des environnements opératoires.

Cependant, l'apprentissage supervisé présente souvent une limitation en termes de généralisation. Un modèle formé sur un ensemble de données spécifique, comme des images de CBCT, peut avoir du mal à s'adapter à de nouvelles données, comme celles provenant d'un scanner tomodensitométrie classique (CT), sans entraîner un nouveau modèle.

Pour éviter la tâche coûteuse de recréer des jeux de données étiquetés et de former un modèle entièrement nouveau, le « transfer learning » serait une solution efficace. Cette approche consiste à utiliser les connaissances apprises par un modèle sur une tâche initiale et à les réutiliser pour une nouvelle tâche. Cette technique permettrait ainsi d'accélérer le processus d'entraînement et d'améliorer la performance des modèles en capitalisant sur des représentations déjà apprises, tout en offrant une grande flexibilité face à de nouvelles données.

Dans nos travaux futurs, l'intégration des réseaux de neurones profonds dans le processus de recalage pourrait ouvrir la voie à des améliorations significatives, notamment en matière de performance et de réduction du temps d'inférence. En comparaison avec les algorithmes variationnels employés dans cette thèse pour aborder les conditions aux bords, les réseaux de neurones profonds ont le potentiel de produire des résultats tout aussi précis, mais avec une rapidité remarquable. Cette efficacité accrue s'explique par la capacité des réseaux neuronaux à apprendre de manière autonome les caractéristiques des données, éliminant ainsi la nécessité de formuler explicitement les équations complexes associées à l'optimisation de l'énergie de recalage. Cette approche novatrice pourrait transformer notre manière d'aborder le recalage, en alliant précision et rapidité, tout en facilitant une adaptation dynamique aux variations des données.

La réduction du temps d'inférence permettrait d'optimiser le processus en milieu clinique ou opérationnel, rendant possible une adaptation rapide aux situations spécifiques de chaque patient ou cas d'usage. Par exemple, avec une analyse plus rapide des bords, le recalage peut être ajusté en temps réel ou quasi-temps réel, tout en garantissant une évaluation précise et sur mesure.

Cette approche serait donc particulièrement adaptée aux besoins des radiologues interventionnels qui doivent prendre des décisions rapides et précises basées sur des simulations réalistes des champs électriques dans le cadre d'une ablation de

tumeur profonde par IRE.

Contents

List of Figures

List of Tables

List of Acronyms

- **ADAM:** ADAptive Moment
- **ADMM:** Alternating Direction Method of Multipliers
- **APRIL:** Assisted Point Registration of Internal Landmarks
- **CBCT:** Cone Beam Computed Tomography
- **CT:** Computed Tomography
- **CRF:** Conditional Random Field
- **ECG:** Electro-CardioGram
- **FISTA:** Fast Iterative Shrinkage-Thresholding Algorithm
- **GAN:** Generative Adversarial Network
- **GET:** Gene Electro-Transfer
- **GPU:** Graphical Processing Unit
- **HCC:** HepatoCellular Carcinoma
- **HQS:** Half Quadratic Splitting
- **IRE:** IReversible Electroporation
- **MRF:** Markov Random Field

- **MRI:** Magnetic Resonance Imaging
- **PET:** Positron Emission Tomography
- **RAG:** Region Adjancy Graph
- **RAM:** Random Access Memory
- **RANSAC:** RANdom SAmple Concensus
- **ReLU:** Rectified Linear Unit
- **RMSP:** Root Mean Square Propagation
- **SIFT:** Scale Invariante Feature Transform
- **TRE:** Target Registration Error
- **TV:** Total Variation
- **TTF:** Tumor Treating Field

Chapter 1

Introduction

Contents

1.1 Medical context

Oncology is the branch of medicine dedicated to the comprehensive study and treatment of cancer. This field encompasses a wide variety of pathologies that are contingent upon the tumor's location and specific characteristics. At the core of the disease, cancerous cells in healthy tissue disrupt the body's delicate equilibrium, where the harmony of cellular function is overtaken by rogue elements. In a normally functioning tissue, cells work in a regulated environment, adhering to strict controls that dictate their growth, division, and death. This balance ensures that tissues maintain their structure and function effectively, contributing to the overall well-being of the organism.

However, cancerous cells defy these biological norms. They arise from mutations that alter their DNA, effectively overriding the natural checkpoints that prevent unchecked growth. Unlike healthy cells, which communicate and cooperate with their neighbors, cancerous cells prioritize their survival and proliferation. They disregard signals that would normally indicate to a cell to stop dividing or to initiate programmed cell death, known as apoptosis.

In this chaotic environment, cancerous cells begin to dominate, rapidly multiplying and invading the surrounding healthy tissue. They disrupt the architecture and function of organs, creating a growing mass, called tumor, that consumes resources and energy at the expense of the body's health. This invasion is not just a physical encroachment; it is a biological insurgency, where the fundamental rules of cellular behavior are rewritten to favor survival over harmony. Illustrated in Figure 1.1, the resulting tumors can vary greatly in size and malignancy.

A critical aspect of cancer's danger lies in its ability to metastasize. This process involves cancerous cells breaking away from the primary tumor and traveling through the bloodstream or lymphatic system to colonize distant organs and tissues. These secondary tumors, or metastases, complicate treatment and significantly impact the prognosis of the patient.

The extent of the disease is summarised in a stage category: a number between I and IV. It defines prognosis and treatment plans available, including inclusion in a clinical trial. Briefly, stage I corresponds to less invasive cancer, stage II and III have increasing tumor extent, and stage IV present with metastases. Historically, staging takes into account anatomical criteria such as the tumor characteristics (morphology, histology, etc), the involvement of regional nodes and the presence or absence of metastases [34]. However, research showed that patient related factors (gender, age, etc) also influence the prognosis. Though accounting for those new criteria may prove valuable, it increases the complexity of the staging system and makes comparison between patient more difficult.

Oncology is a critical and ever-evolving field of research, profoundly shaping

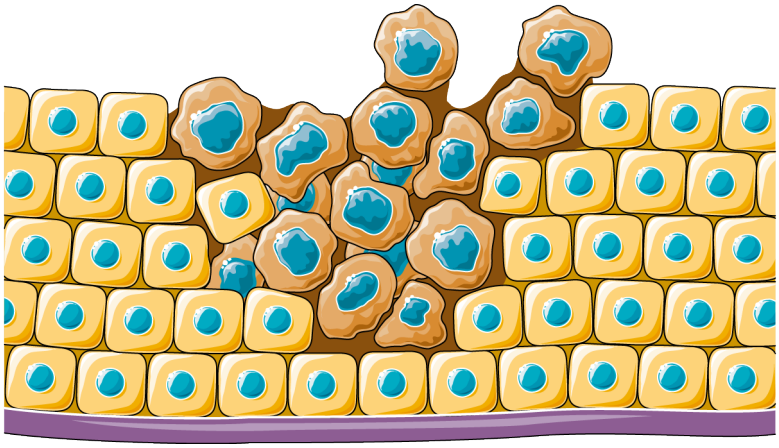


Figure 1.1: Cancerous cells multiply uncontrollably, forming tumors that invade healthy tissues, disrupting their structure and function. These cells consume resources like nutrients and oxygen, while also manipulating nearby normal cells to support tumor growth, further weakening surrounding tissues (from smart.servier.com, free medical images).

the lives of countless individuals worldwide. This discipline not only addresses the complex nature of cancer but also strives to unlock new therapeutic possibilities and improve patient outcomes. The high mortality rate associated with cancer is, in part, a consequence of extended life expectancy, which, while a testament to advancements in healthcare, also increases the population's susceptibility to developing cancer. As people live longer, the likelihood of encountering the myriad of factors that can lead to cancer rises.

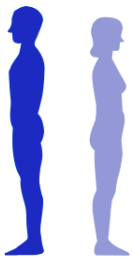
Effective solutions in oncology must address the constantly evolving landscape of cancer epidemiology and leverage technological innovations in medical imaging, surgical tools, and therapeutic approaches. For instance, the integration of advanced imaging technologies can significantly improve diagnostic accuracy and treatment planning.

Furthermore, these solutions must tackle the practical challenges faced by healthcare professionals, such as ensuring that new technologies and treatments are seamlessly integrated into existing clinical workflows, are cost-effective, and enhance overall patient care. Only by balancing innovation with practicality can the field of oncology continue to make significant strides against cancer.

Currently, deep-seated tumors, those located in areas that are challenging to access surgically, contribute significantly to cancer-related mortality. In 2024, it is projected that 22% of cancer deaths (Fig. 1.2) will be due to such tumors, including those in the liver, pancreas, brain, and prostate. Addressing these tumors is therefore a critical priority in oncology research.

Estimated New Cases

Male			Female		
Prostate	299,010	29%	Breast	310,720	32%
Lung & bronchus	116,310	11%	Lung & bronchus	118,270	12%
Colon & rectum	81,540	8%	Colon & rectum	71,270	7%
Urinary bladder	63,070	6%	Uterine corpus	67,880	7%
Melanoma of the skin	59,170	6%	Melanoma of the skin	41,470	4%
Kidney & renal pelvis	52,380	5%	Non-Hodgkin lymphoma	36,030	4%
Non-Hodgkin lymphoma	44,590	4%	Pancreas	31,910	3%
Oral cavity & pharynx	41,510	4%	Thyroid	31,520	3%
Leukemia	36,450	4%	Kidney & renal pelvis	29,230	3%
Pancreas	34,530	3%	Leukemia	26,320	3%
All sites	**1,029,080**		**All sites**	**972,060**	



Estimated Deaths

Male			Female		
Lung & bronchus	65,790	20%	Lung & bronchus	59,280	21%
Prostate	35,250	11%	Breast	42,250	15%
Colon & rectum	28,700	9%	Pancreas	24,480	8%
Pancreas	27,270	8%	Colon & rectum	24,310	8%
Liver & intrahepatic bile duct	19,120	6%	Uterine corpus	13,250	5%
Leukemia	13,640	4%	Ovary	12,740	4%
Esophagus	12,880	4%	Liver & intrahepatic bile duct	10,720	4%
Urinary bladder	12,290	4%	Leukemia	10,030	3%
Non-Hodgkin lymphoma	11,780	4%	Non-Hodgkin lymphoma	8,360	3%
Brain & other nervous system	10,690	3%	Brain & other nervous system	8,070	3%
All sites	**322,800**		**All sites**	**288,920**	

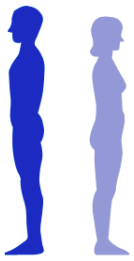


Figure 1.2: Projected Figures for 2024 regarding the incidence of new cancer cases and associated fatalities, categorized by types of cancer, within the United States population, encompassing both males and females [87], reproduced with permission.

To effectively target deep-seated tumors, the focus has shifted towards innovative non-surgical approaches. These methods are essential as they reduce the complexities and risks associated with traditional surgical interventions. By emphasizing non-invasive treatments, oncology aims to improve patient outcomes and reduce the burden of deep-seated tumors on global health.

To help patients, medical imaging is indispensable at every stage of cancer treatment: diagnosis, treatment planning, treatment guidance, and treatment evaluation. The integration of computer vision into medical imaging is anticipated to significantly enhance patient care from the moment of diagnosis.

Anomaly detection: Computer vision can automatically detect anomalies, such as tumors or masses, in medical images. This serves as a crucial first step towards diagnosis, enabling earlier and more accurate identification of cancerous regions.

Segmentation of anatomical structures: By accurately segmenting anatomical structures, computer vision helps determine disease progression and aids in precise treatment planning. This ensures that targeted therapies are delivered effectively, reducing damage to surrounding healthy tissues.

Image registration: Computer vision techniques can merge images captured at different times and/or from different modalities, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET) scans, into a single comprehensive image. This fusion of information provides a holistic view of the patient's condition, enhancing the accuracy of diagnosis and the effectiveness of treatment plans.

By incorporating these advanced computer vision techniques, medical imaging not only streamlines the workflow for healthcare providers but also ensures that patients receive the most accurate and effective care throughout their cancer treatment journey.

In this work, we introduce computer vision tools designed to facilitate the ablation process of deep-seated tumors through irreversible electroporation (IRE), with a particular focus on hepatocellular carcinoma (HCC) and pancreatic cancer. This intricate procedure leverages electric fields to effectively ablate the tumor, demanding accurate and real-time data analysis for successful implementation. Our paramount objective is to achieve precise estimations of the treated area during the procedure, thereby ensuring optimal clinical outcomes. This endeavor encompasses a range of image analysis challenges, with this thesis specifically delving into the complexities of image registration and object localisation. While the proposed techniques are specifically designed to enhance IRE applications, their versatile potential extends across various domains, encompassing both oncology and broader computer vision applications.

1.1.1 Deep seated tumors

We focus on two types of cancer caracterised by deep seated tumors, namely liver cancer and pancreas cancer.

Liver cancer: Liver cancer is the sixth leading cause of cancer death [87] and the first leading cause of death by cirrhosis [29]. The increasing number of diagnosed cases results from a change in epidemiology: previously, liver cancer was mainly caused by viral hepatitis, however, it is now increasingly linked to alcoholic and non-alcoholic fatty liver disease following obesity.

Early diagnosis based on surveillance leads to improved prognosis (50.8% 5 year survival rate) as opposed to symptom-based only (27.9%) [29]. In general, multiphase CT or MRI are used to identify the tumor. The former has a shorter duration and a lower cost but necessitates radiation exposure and, sometimes, iodinated contrast.

In particular, HCC is a type of cancer characterized by deep-seated tumors that primarily target the liver, a vital organ responsible for filtering blood and producing bile to aid in digestion. The liver is situated in the upper right quadrant of the abdomen, nestled beneath the diaphragm and above the stomach and gallbladder (in green in Fig. 1.3). Its proximity to major blood vessels, including the aorta (in red in Fig. 1.3) and the inferior vena cava (in blue in Fig. 1.3), makes surgical intervention particularly challenging in most cases.

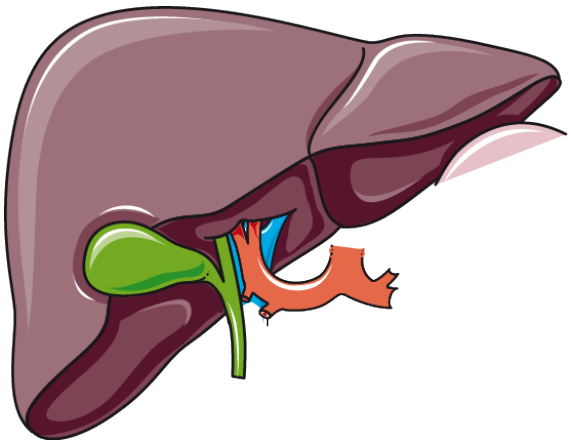


Figure 1.3: The liver and its surrounding anatomical structures, which include the gallbladder depicted in green, the aorta in red, and the inferior vena cava in blue (from smart.servier.com, free medical images).

For these reasons, there is an urgent need for intensive research in both prevention and treatment. Personalised screening methods, which can detect cancer at its earliest stages, and the development of advanced serological biomarkers are crucial. These biomarkers can predict the onset of disease and monitor the effectiveness of treatments, paving the way for more targeted and effective interventions.

Pancreas cancer: Pancreatic cancer ranks as the third leading cause of cancer-related deaths globally [87]. The prognosis remains grim, with a mere 4% of patients surviving beyond five years following diagnosis [94]. This malignancy is predominantly linked to a combination of genetic predispositions, increasing age, lifestyle factors such as obesity, and the harmful effects of smoking. In the majority of cases, patients suffer conventional ductal adenocarcinoma, the most common and aggressive form of this cancer [51].

The pancreas (Fig. 1.4) plays a crucial role as both an exocrine gland, secreting digestive enzymes into the small intestine via ducts, and an endocrine gland, releasing hormones such as insulin and glucagon into the bloodstream. Positioned

posterior to the stomach, its location near major blood vessels, such as the superior mesenteric artery and vein, ensures efficient nutrient and hormone transport throughout the body, while complicating surgical access to the organ.

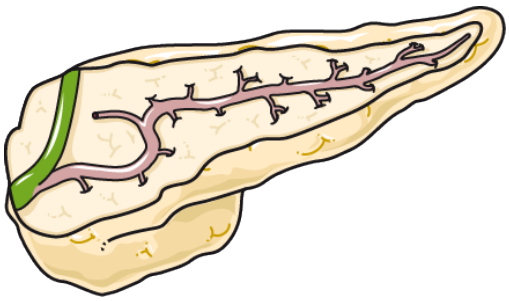


Figure 1.4: Cross-sectional view of the pancreas: the bile duct, highlighted in green, carries bile from the liver and gallbladder to the duodenum to aid in digestion, particularly the breakdown of fats. The pancreatic ducts, shown in pink, transport digestive enzymes produced by the pancreas to the duodenum, where they help in the digestion of proteins, carbohydrates, and fats (from smart.servier.com, free medical images).

Currently, routine screening for pancreatic cancer is not standard practice, largely due to the limited sensitivity of existing imaging modalities for early-stage detection. This limitation underscores the urgent need for more effective screening techniques, particularly for identifying pre-invasive lesions, as their early resection could significantly reduce the risk of developing invasive pancreatic cancer. Among the various potential screening tools, endoscopic ultrasound has emerged as a particularly promising option. This imaging modality is valued for its ability to provide high-resolution images and precise tissue sampling, making it an invaluable tool in the early detection and staging of pancreatic cancer [94].

Most pancreatic cancer are clinically silent, as symptoms start appearing when the disease has progressed to surrounding tissues, or distant organs. Thus, most patient are diagnosed with unresectable cancer. Treatment often focuses on prolonged survival or downstaging to eventually ablate the tumour.

Systemic treatments: Today, the landscape of treatment options for deep-seated tumors is incredibly diverse, reflecting the complex nature of cancer itself. No single protocol has emerged as universally effective, underscoring the need for a tailored approach in each individual case. Systemic therapies aim to combat cancer at a holistic level, targeting the body as a whole to inhibit tumor growth and spread.

Chemotherapy: Chemotherapy drugs are administered to combat the growth

of cancerous cells by inhibiting their cell division mechanisms. However, these potent drugs do not exclusively target the cancerous site, resulting in widespread systemic effects. This non-specificity means that healthy, rapidly dividing cells, such as those in the bone marrow, digestive tract, and hair follicles, are also affected. Consequently, patients often experience a range of side effects, including fatigue, hair loss, gastrointestinal issues, and increased susceptibility to infections due to suppressed immune function [75][21][89].

Anti-angiogenic therapy: Innovative cancer therapies now include drugs specifically designed to inhibit angiogenesis, the process by which tumors develop new blood vessels to sustain their growth. By targeting angiogenesis, these drugs prevent the formation of blood vessels within the tumor, thereby starving it of essential nutrients and oxygen needed for proliferation. This approach is often combined synergistically with chemotherapy, enhancing its effectiveness. Chemotherapy benefits from the improved access to the tumor due to reduced blood flow caused by angiogenesis inhibitors. Together, these therapies aim to shrink tumors and improve patient outcomes by depriving cancerous cells of their blood supply and delivering cytotoxic agents directly to the tumor site [107][4].

Immunotherapy: The patient's immune system is stimulated in order to target cancerous cells effectively. Immunotherapy stimulates specific components of the immune system, such as T-cells, to recognize and attack cancer cells. This stimulation often involves therapies that activate immune checkpoints or genetically engineer immune cells to enhance their targeting capabilities. By boosting immune responses against cancer, immunotherapy offers a promising strategy to achieve long-term remission and improve survival rates. Moreover, reprogramming immune cells to specifically target cancer cells has shown significant potential in overcoming the challenges posed by tumor heterogeneity and resistance [20]. This personalized approach not only minimises damage to healthy tissues but also enhances the body's ability to recognize and eliminate malignant cells, marking a pivotal advancement in oncology [35][24][92][84].

Loco-regional treatments: Other approaches focus on targeting tumors more locally, frequently incorporating surgical intervention as a critical component of treatment. These localised strategies aim to minimise collateral damage to surrounding healthy tissues while maximising the effectiveness of the intervention. Surgical techniques can range from complete tumor resection to minimally invasive procedures, each tailored to the specific characteristics of the tumor and the individual patient's needs.

Transplantation: In the demanding process of organ transplantation, the entire diseased organ is removed and substituted with a healthy equivalent from a donor.

This major surgical procedure is inherently complex and requires a prolonged recovery period. To maximise the chances of a successful transplant, it is crucial to ensure genetic compatibility between the donor and recipient. This compatibility helps to minimise the risk of organ rejection. Furthermore, patients must undergo continuous treatment with immunosuppressive drugs, which effectively dampen the immune system's response, thereby reducing the likelihood of the body rejecting the new organ. This balancing act between preventing rejection and maintaining immune function is a cornerstone of post-transplant care.

Tumor resection: During this operation, the diseased section of the organ is excised. This major surgery requires significant medical expertise and carries inherent risks. Patients may experience complications such as infection and bleeding post-operation. In the case of the liver, despite its remarkable ability to regenerate and restore its function over time, the recovery process can be extensive and demanding. Additionally, the long-term effects can vary, with some patients potentially facing persistent health challenges. Ensuring optimal recovery and managing any long-lasting effects require careful postoperative care and monitoring.

Electro-chemotherapy: Bleomycin, a chemotherapy agent known for its limited cellular penetration compared to more efficient drugs like cisplatin, is administered in conjunction with the application of an electric field. These electric pulses create temporary pores in the membrane of cancer cells, significantly enhancing the drug's ability to infiltrate malignant tissues. This innovative technique not only optimises the delivery of the drug, ensuring that it reaches its intended target more effectively, but also minimises systemic side effects by concentrating the therapeutic action precisely where it is needed most. By focusing the treatment directly on the tumor, this method represents a promising advancement in cancer therapy, maximising efficacy while safeguarding healthy surrounding tissues [92][88][33].

Transarterial radio-embolisation: Utilizing microspheres loaded with radioactive isotopes represents a sophisticated approach in targeted therapy. These microspheres, administered intravenously near the diseased tissue, emit radiation, damaging the DNA of nearby cells, particularly cancerous cells, and disrupt their ability to replicate. This technique, while effective, is associated with significant costs and complexities in administration and monitoring. Moreover, exposure to radiation can lead to various side effects, including damage to healthy tissues surrounding the treatment site. Despite these challenges, targeted radiation therapy remains a critical tool in oncology, offering a tailored approach to treating localised cancers with minimal systemic impact [82][2][15][97][91].

Photo-dynamic therapy: Photo-sensitisers, once injected into the bloodstream, are activated by light of a specific wavelength. This activation unleashes reactive oxygen species that overwhelm cancer cells' defenses, leading to their demise through

processes like apoptosis and necrosis. Photo-dynamic therapy stands out for its ability to precisely target malignant cells while minimising damage to healthy tissue, marking it as a promising therapeutic strategy in cancer treatment [108][54][11].

Radio-frequency ablation: High-frequency electric fields are administered via electrodes strategically positioned near the affected site. The thermal effects generated by these pulses induce cell death through necrosis, effectively eradicating the cancer cells [66][57][37].

Micro-wave ablation: High frequency microwave energy is delivered through antennas precisely inserted near the disease site. This electromagnetic energy is focused to raise the temperature of cancerous tissue, leading to thermal destruction of cancer cells. This technique offers a targeted and minimally invasive treatment approach in oncology, effectively removing tumors while minimising damage to surrounding healthy tissue [67].

Cryo-therapy: A hollow probe, filled with cooling gaz, is inserted near the disease site. This method uses extreme cold to destroy cancer cells through induced necrosis, ensuring precise targeting of tumors with minimal impact on healthy tissue [106][9][58].

IRE ablation: Inserted electrodes emit high-intensity electric pulses near the tumor, creating pores in cancer cell membranes. This process triggers apoptosis, a programmed cell death mechanism, ensuring targeted destruction of cancerous tissue [14][27].

It is worth noting that, prior to IRE ablation, surgical as well as non-invasive local treatments were nearly impossible in the case of pancreatic cancer due to the high risk of pancreatitis and the possible damages to adjacent vessels [53]. Also, responses to chemotherapy, both adjuvant and neo-adjuvant, are rather poor in that case [51].

1.1.2 Electroporation therapies

One promising method for the non-surgical ablation of deep-seated tumors is electroporation based ablation.

The fundamental principle underlying this technique is electroporation, utilized in various biological and medical applications. Electroporation involves the application of a pulsed electric field to cells, which induces the formation of transient nanopores within the cell membrane. These nanopores facilitate the permeabilisation of the membrane to ions and macromolecules.

Depending on the intensity and duration of the electric field applied, the resulting membrane damage can be either reversible or irreversible as seen in Figure 1.5. In the case of reversible electroporation, the cell membrane repairs itself after

the electric field is removed, restoring the cell to its normal state and functionality (Fig. 1.6). This reversible process is crucial for applications that require temporary access to the cell interior without causing permanent damage.

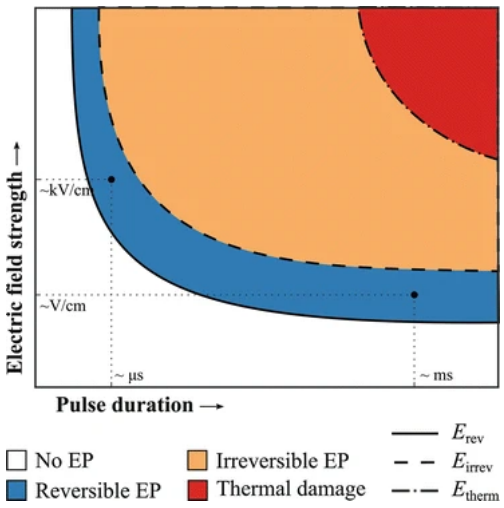


Figure 1.5: Effects of electroporation, reversible, irreversible, as well as thermal damage, in relation to variations in pulse duration and electric field intensity [59] reproduced with permission.

Conversely, irreversible electroporation occurs when the electric field causes extensive and permanent disruption of the cell membrane, leading to a critical loss of homeostasis. This irreversible damage triggers cell death through apoptosis, a programmed cell death mechanism [19]. This particular kind of cell death takes hours [27], making evaluation from directly observing the tissues impossible.

The applications to such an increase in cell membrane permeability are numerous in medicine.

Drug delivery: Electro-chemotherapy is an innovative therapeutic technique that synergistically combines chemotherapy with electroporation to significantly enhance the delivery of chemotherapeutic drugs, such as bleomycin, into cancerous cells. This method addresses a key limitation of certain chemotherapeutic agents—namely, their inability to efficiently penetrate cell membranes due to their hydrophilic nature. By utilizing electroporation, electro-chemotherapy facilitates the direct entry of otherwise non-permeant cytotoxic drugs into the tumor cells. In practice, the procedure begins with the targeted injection of chemotherapy drugs into the tumor site. The subsequent application of precisely timed electric pulses causes transient pores to form in the cell membranes, allowing the drugs to diffuse into the cells effectively [88][33]. Once inside, the drugs exert their cytotoxic effects, leading to

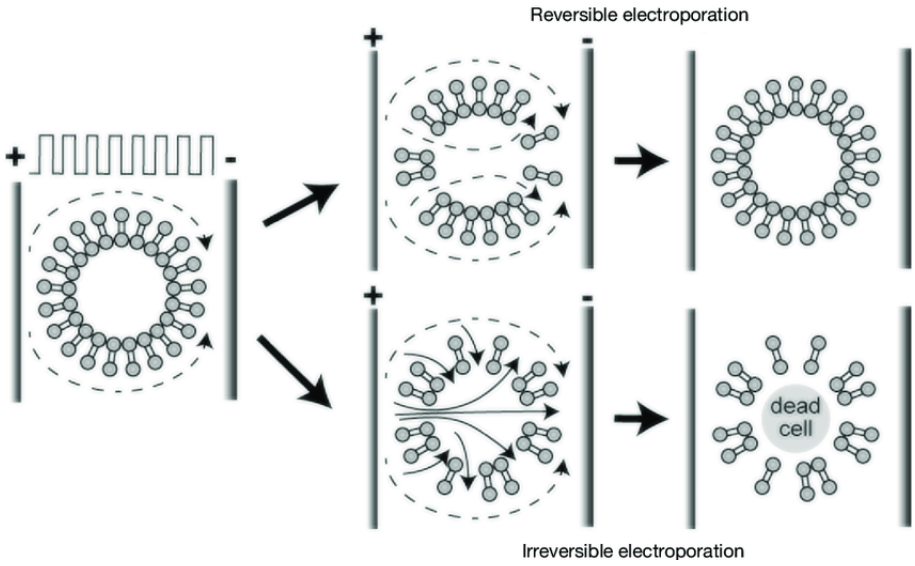


Figure 1.6: The application of electric pulses to cells induces the formation of pores in their plasma membrane, a process known as electroporation. When these pores are reversible, the membrane recovers, allowing the cell to restore its normal function. However, if the pulses cause irreversible damage, the disruption in cellular homeostasis becomes too severe, leading to cell death, typically through apoptosis [5], reproduced with permission.

increased cancer cell death. This strategic combination offers several advantages over traditional chemotherapy. Most notably, it allows for significantly lower drug dosages while still achieving high efficacy, thereby reducing the risk of systemic side effects and toxicity that often accompany higher-dose chemotherapy treatments.

Gene Electro-Transfer (GET): GET is an innovative technique utilized to introduce foreign DNA into cells by applying precise electric pulses. The induced pores allow DNA molecules to enter the cell's cytoplasm. Once inside, the foreign DNA can integrate into the cell's genome, enabling its expression to produce specific proteins or modulate cellular functions [79]. This method holds significant promise in therapeutic applications, such as DNA vaccines, where the introduced genetic material can stimulate an immune response by instructing cells to generate antigens.

Immunotherapy with electroporation: Both irreversible and reversible electroporation not only disrupt cellular membranes but also play a pivotal role in activating a systemic antitumor immune response, thereby enhancing the overall efficacy of immunotherapy treatments [32]. The creation of transient or permanent

pores in cell membranes leads to the release of intracellular material, including tumor-specific antigens. This leakage triggers a robust immune response that extends beyond the immediate treatment site, a phenomenon known as the abscopal effect. In this context, tumor antigens released from the treated site are recognized by the immune system, which subsequently mounts a broader attack on tumor cells throughout the body, even those not directly targeted by the electroporation procedure. Furthermore, the moderate tissue damage associated with the application of electric fields during electroporation initiates a localised inflammatory response. This tissue disruption attracts immune cells, such as macrophages and dendritic cells, to the site of electroporation. These cells, along with the release of cytokines and other signaling molecules, help orchestrate a more vigorous antitumor immune response. The combination of these processes—tumor antigen release and immune cell recruitment—primes the immune system for more effective recognition and destruction of tumor cells, thereby complementing and amplifying the effects of immunotherapy.

Tissue ablation: IRE ablates problematic tissues with precision and minimal collateral damage. This technology is particularly effective in targeting and eliminating tumors [27], where conventional methods might pose a higher risk to surrounding healthy tissue. Beyond its application in oncology, IRE is also utilized in the field of cardiology, where it ablates myocardial tissue to treat arrhythmias. By creating controlled disruptions in the cell membranes of abnormal heart tissue, IRE helps restore normal heart rhythm, offering a promising alternative to traditional ablation methods with potentially fewer complications and improved outcomes [43].

1.1.3 Focus on IRE ablation

IRE provides a minimally invasive, non-surgical method for the ablation of deep seated tumors: mainly liver [27], pancreas [14], prostate [10], and kidney [69]. The procedure requires a direct current generator and the corresponding electrodes. For example, the Nanoknife needles employed by our collaborating interventional radiologists are 15 cm in length, providing optimal reach and accurate placement during procedures. Additionally, they boast a diameter ranging between 16 and 19 gauge, facilitating precise application and minimising patient discomfort.

IRE ablation is distinguished by its focal-energy approach, which minimises thermal effects, thereby preserving the extracellular matrix of surrounding tissues and vital structures such as organs and blood vessels. This targeted energy delivery ensures that nearby tissues experience minimal damage during the procedure and that fluidic function in blood vessels is not compromised [74].

The recovery process following IRE is facilitated by the preservation of the tissue scaffold. Normal tissue regrowth occurs rapidly post-procedure, with limited

formation of scar tissue [73]. This stands in stark contrast to thermally based techniques, which often lead to more significant tissue damage and scarring [53].

Furthermore, this technique does not suffer from the heat sink effect which often compromises the completeness of ablation in traditional thermal techniques due to heat dissipation through blood flow. It is worth noting that blood vessels still impact IRE as the heterogeneity of tissues need to be taken into account in the estimation of the delivered electric field [74].

Also, localised temperature elevation can occur near the electrodes within the tumor's surrounding tissues, influenced by factors such as the applied voltage and electrode spacing. Proper electrode alignment is crucial to mitigate this risk: inadequate parallelism increases the likelihood of thermal damage in areas where electrodes are closest together [31].

In general, multiple electrodes are necessary for a single tumor, when most thermal techniques only require one. Also, unlike thermal methods that induce necrotic cell death, leading to immediate changes visible on medical imaging, IRE mostly triggers apoptosis, a process where cells undergo programmed death hours post-procedure. This delayed effect makes the precise delineation of the effective treatment area challenging immediately after the intervention. Such intricacies necessitate meticulous pre- and per-procedural planning to accurately assess and optimise the diverse parameters crucial for achieving successful outcomes.

Pulse parameters: The pulse parameters determine the electric field the cells are exposed to. They include the pulse strength (typically ranging between 1400 and 2000V), the pulse shape (square waves are most often used but there are also sine waves, etc), the pulse duration (between 50 and 100 µs), the number of pulses (80 to 100) generally delivered in trains of several tens [27], the pulse polarity, and finally the delay between the pulses [31].

Electrode positioning: The location of electrodes plays a critical role in defining both the spatial distribution and the intensity of the electric field applied during procedures. Ideally, electrodes should be positioned in parallel alignment to ensure uniform field distribution, at a similar depth within the tissue to optimise efficacy, and separated by at least 2mm from major blood vessels to minimise thermal damage [1]. The number of electrodes typically ranges between 2 and 6, depending on the size and location of the target area, with variations in electrode shapes such as plates, clamps, needles, or catheters offering tailored applicability [31].

Currently, establishing a universal protocol remains elusive due to the diverse scenarios encountered, particularly in the context of deep-seated tumors. However, needle electrodes emerge as the preferred choice for the ablation of deep seated

tumors due to their ease of handling and relatively low invasiveness. Despite these advantages, needle electrodes are susceptible to creating uneven electric field distributions, which can impact treatment efficacy. This variability necessitates careful consideration of electrode placement and configuration to ensure optimal therapeutic outcomes.

1.1.4 Clinical and numerical workflows

IRE-based ablation stands out as a highly intricate procedure, demanding meticulous planning and evaluation. The complexity lies in the precise application of electric pulses to achieve targeted tissue ablation while preserving surrounding structures. This work emphasizes the critical role of per-operative evaluation, proposing innovative solutions to enhance the accuracy of efficacy estimation during IRE procedures. By integrating advanced imaging techniques and real-time monitoring, clinicians can better assess treatment outcomes and optimise the procedure as it is performed.

In that respect, the interventional radiologist needs to check the effective area of treatment, corresponding to the region where the electric field intensity reaches the threshold for IRE. A numerical workflow was proposed in [27] to fully integrate within the current clinical workflow followed by the contributors to the database used in this work. It utilizes the medical images already captured along with the electric pulses information recorded by the generator to provide an estimation of the effective area of treatment. The symbiosis between clinical and numerical workflows is shown in Figure 1.7 for the per-operative and post-operative sessions.

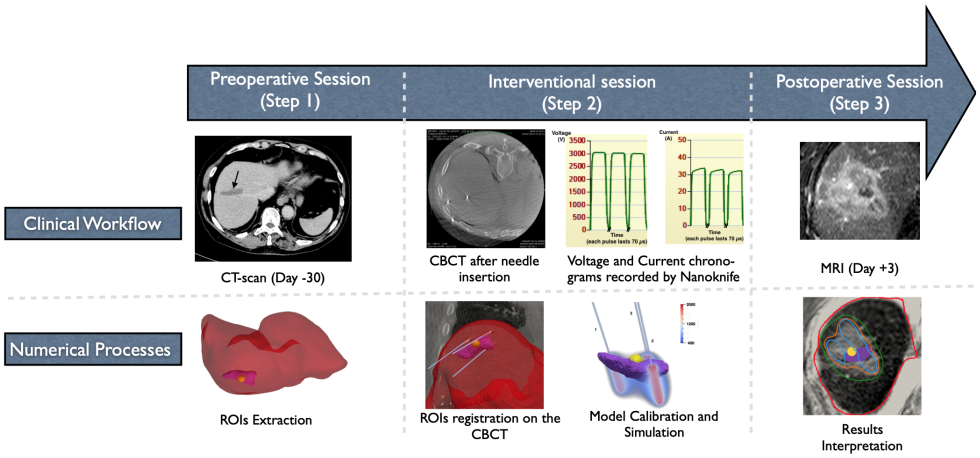


Figure 1.7: Combination of clinical and numerical workflows within the scope of deep-seated tumor ablation via IRE [27], reproduced with permission.

Pre-operative session: A CT-scan is conducted a few days before the operation to precisely localise the tumor. Advanced algorithms propose an ideal electrode configuration based on this data. However, these theoretical configurations often overlook critical anatomical constraints encountered during the treatment of deep-seated tumors, such as the presence of bones and nearby organs. Despite these limitations, these programs prove invaluable for treating superficial tumors in areas like the skin, subcutaneous tissue, prostate, and even certain bone tumors. Their ability to pre-plan electrode placements enhances procedural efficiency and accuracy, contributing significantly to patient outcomes [27].

Per-operative session: During the intervention, the patient is under general anesthesia and neuromuscular blocking agents. This cautious approach is crucial as the application of electric pulses during the procedure can potentially trigger involuntary muscle contractions, thereby complicating the accurate targeting of the tumor site.

The medical team then starts with a CBCT scan of the abdominal cavity. A CBCT is captured via the synchronous rotation of a radiogenic source and a scanning device around the patient [85]. This category of non-invasive medical imaging measures the tissue density. The sinograms of the projections are then used to reconstruct a 3D representation of the zone of interest, often through filtered retroprojection [23] as it is a well-understood and fast algorithm. The resulting 3D image is cylinder-shaped, as seen in Figure 1.8. This imaging modality offers advantages over traditional CT scans, requiring a reduced radiation dose and employing a lower-energy beam, which minimises exposure and enhances safety.

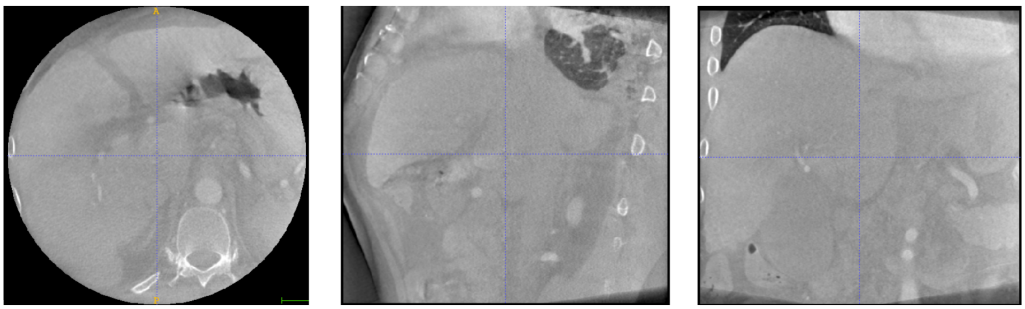


Figure 1.8: CBCT slices from our patient database, specifically for individuals undergoing treatment for liver cancer: (a) displays the axial plane, (b) the coronal plane, and (c) the sagital plane.

However, the benefits of this modality are tempered by significant challenges that impact its overall effectiveness. One of the primary issues is its pronounced

susceptibility to artifacts, particularly when high-density objects, such as electrodes, are present within the field of view. These artifacts can severely distort the images. Additionally, this imaging technique often suffers from poor contrast, making it difficult to distinguish between different tissue types or to identify subtle anatomical structures.

These limitations primarily arise from the use of a polychromatic X-ray beam, which contributes to the common phenomenon of beam hardening. Beam hardening occurs when lower-energy photons are absorbed more readily than higher-energy ones as the beam passes through dense materials, resulting in uneven attenuation and non-uniformity in the final images. This effect not only degrades image quality but also complicates the interpretation of the results.

Furthermore, CBCT is inherently limited by its partial field of view. Unlike traditional CT scans that can capture comprehensive images of larger anatomical regions, CBCT is designed to focus on specific areas of interest, often excluding surrounding tissues. This narrow targeting can lead to a lack of boundary information, which is critical for the accurate application of computer vision tools. Without a full context, these tools may struggle to accurately analyse the images, potentially compromising the precision of automated diagnostics or treatment planning.

The dimensions are 512 voxels in the anterior-posterior and the right-left direction, and between 195 and 512 voxels in the superior-inferior direction. The voxel size is isotropic and between 0.362 and 0.453mm.

This initial scan serves to map out the precise geometry and location of the region of interest, facilitating accurate electrode placement and ensuring optimal treatment outcomes. The tumor and the affected organ main structures are segmented either manually—which is still common but time consuming, laborious because of the variability in shape, location, number and unclear boundaries of tumors (in the case of liver cancer, it is even more so due to the low contrast between the organ and its surroundings)—or semi-automatically—using, for instance, a snake algorithm, as it is the case here [27].

The intervention proceeds with the percutaneous insertion of needles guided by an interventional radiologist. Employing a free-hand technique, each needle is carefully placed to closely mimic the configuration outlined in the pre-operative proposal. Throughout this delicate process, real-time ultrasound and fluoroscopic guidance are utilized. These advanced imaging modalities ensure not only the accurate placement of electrodes but also parallel alignment and uniform depth. Such precision is crucial for the success of IRE, as it enhances the efficacy of the treatment while minimising risks to surrounding tissues.

Following the initial placement of electrodes, a second CBCT scan is performed to confirm their precise positioning via a localisation algorithm. However, due to their high density, the electrodes generate significant artifacts that obscure clear

visualisation of the tumor. This limitation motivates the need for a second scan to accurately assess the electrode placement and the surrounding tissue, separately.

Ensuring precise needle localisation is paramount as deviations from the planned positions can introduce uncertainties in the electric field distribution. These uncertainties impact the effectiveness of the treatment area, influencing the therapeutic outcome evaluation [27].

Then, we join the geometrical information and the needle localisation into one frame of reference. To do so, the segmented regions of interest are registered onto the last CBCT scan in order to obtain the electrodes position with respect to the tumor [27]. Ideally, this procedure takes into account the rigidity of the structures (bones, needles, etc) to avoid unwanted distortions.

Concurrently, electric pulses are administered, synchronised with the patient's electrocardiogram (ECG) to coincide with the cardiac refractory phase, ensuring minimal disruption to cardiac function. Moreover, comprehensive records of both pulse intensities and their timing, known as chronograms, are documented. These recordings serve crucial purposes in subsequent evaluation of the electric field, facilitating further adjustments and evaluations.

An estimation of the delivered electric field is provided to the physicians for them to adjust their approach if necessary. They may decide to add an extra train of electric pulses to increase the area of effective treatment or do a pull-back—that is, adjusting the depth of the needles—to adapt the depth of the treatment. The information consists in the visualisation of the electric field as seen in Figure 1.9. Generally, the isolines corresponding to the areas of reversible and irreversible electroporation on the CBCT with the needles and the propagated tumor segmentation are shown for clarity. When a change is made, the procedure may be repeated to ensure a full treatment of the tumor.

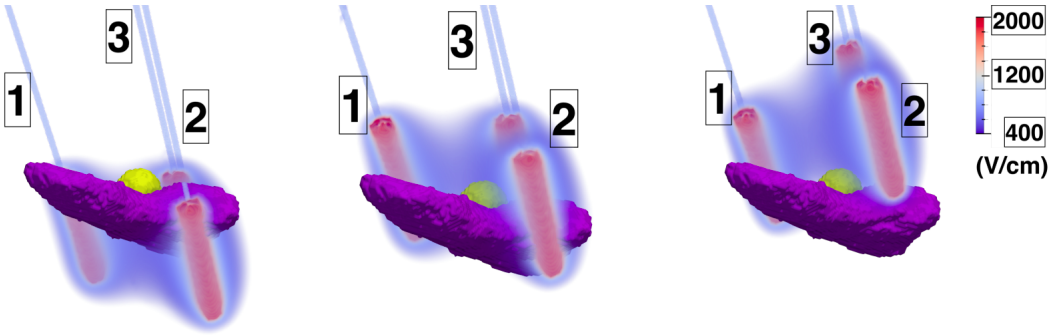


Figure 1.9: Electric field simulation during an IRE HCC ablation: on the left for the initial needle placement, in the center after the first pull-back and on the right after the second pull-back. The tumor is depicted in yellow and the liver in purple [27], reproduced with permission.

Post-operative session: Ultimately, several days post-operation, the utilization of MRI proves instrumental in evaluating the treatment's efficacy. An MRI is a sophisticated non-invasive imaging technique that offers a highly detailed 3D anatomical depiction without exposing the patient to harmful radiation. Utilizing powerful magnets, an MRI generates a magnetic field that aligns protons within the body's tissues. Subsequent application of a radiofrequency current disturbs these aligned protons, causing them to emit signals as they return to their original alignment state [3]. The MRI sensors capture these emitted signals, producing detailed images of the body's internal structures.

To enhance image clarity, contrast agents like gadolinium can be administered intravenously, accelerating the proton realignment process and resulting in brighter images. This imaging modality excels in visualising soft tissues rich in water, making it invaluable for diagnosing conditions affecting the brain, spinal cord, joints, and organs.

Despite its diagnostic advantages, MRI systems are costly to install and maintain, and examinations can be time-consuming due to the detailed nature of the scans. However, its unparalleled ability to provide non-invasive, detailed anatomical information makes it indispensable in modern medical diagnostics.

Each sequence within the MRI showcases distinct facets of the therapeutic response, providing a comprehensive view of the affected area's healing and recovery. By employing various imaging sequences, and contrast-enhanced imaging, clinicians can discern different characteristics of tissue healing, presence of residual tumors, and overall treatment impact:

- T1 weighted shows the largest zone of cell death, i.e. IRE,
- delayed phase of gadolinium enhanced T1 weighted shows the thermally impacted regions: near the needle location (and scar tissue from previous thermally-based treatments, if any),
- T2 shows the largest zone affected by electroporation, i.e. both reversible and irreverible electroporation.

1.1.5 Challenges for numerical methods

To provide relevant information regarding the effective treated area, we must face several challenges, inherited from the clinical settings.

First and foremost, CBCT stands out as the most informative imaging modality available in the operating room, particularly when compared with alternatives such as ultrasound and fluoroscopy. Despite its advantages, CBCT presents a challenge with its relatively low signal-to-noise ratio, which complicates the clarity of the images produced. As a result, the analysis of these images becomes a nuanced

endeavor, whether it involves tumor segmentation, precise localisation of electrodes, or image registration. This complexity underscores the necessity for advanced analytical techniques to enhance the utility of CBCT in surgical settings.

Second, the entire application operates on commodity hardware, a crucial consideration in clinical practices. In many hospitals, access to high-performance computing clusters equipped with GPUs for accelerated computation is often limited. This reality imposes significant constraints on the design of numerical tools, necessitating innovative solutions that can optimise performance within the confines of readily available technology. As such, it becomes imperative to create tools that not only meet the demands of clinical efficiency but also adapt seamlessly to the existing infrastructure in healthcare settings.

Additionally, the patient is on the operating table during the electric field estimation, which introduces significant time constraints to the process. The estimation must be completed in just a few minutes to remain practical, particularly given that multiple estimates may be necessary throughout a single procedure. This is essential to ensure that the adapted approach effectively ablates the entire tumor. Such urgency underscores the importance of developing rapid, accurate estimation methods that can seamlessly integrate into the fast-paced environment of the operating room, ultimately enhancing patient outcomes and procedural efficiency.

1.2 Contributions

The primary objective of this work is to equip interventional radiologists with crucial and highly relevant insights necessary for the successful execution of IRE ablation procedures. Central to this goal is the precise estimation of the electric field delivered during the procedure, a task that demands a comprehensive array of advanced image processing techniques. These techniques are employed to extract critical information essential for accurate evaluation, particularly the precise positioning of the needles relative to the tumor.

By harnessing state-of-the-art imaging methodologies, this thesis seeks to significantly enhance the precision and effectiveness of IRE ablation. This improvement is anticipated to lead to better therapeutic outcomes for patients, while simultaneously advancing the field of interventional radiology by providing radiologists with enhanced tools for decision-making during procedures. Moreover, the work presented here addresses key clinical challenges in collaboration with the AP-HP Avicenne Hospital, ensuring that the solutions proposed are both clinically relevant and aligned with the practical needs of healthcare professionals. Through this partnership, the research aspires to bridge the gap between cutting-edge technological innovation and real-world clinical applications, ultimately contributing to the

refinement of interventional radiology practices and patient care.

The accurate estimation of the delivered electric field is essential for determining the precise location of the effective treatment area, providing interventional radiologists with invaluable real-time guidance. This information empowers them to make critical adjustments during the procedure, ensuring that the ablation is optimally targeted. By doing so, radiologists can ensure that the entire tumor is effectively treated, significantly reducing the risk of incomplete ablation and minimising the likelihood of cancer recurrence.

In the numerical framework previously developed by the MONC research team [28], the linear static model is used:

$$\begin{cases} -\nabla \cdot (\sigma \nabla \phi) = 0 & \\ \phi = \phi_{src} & \text{at the active needles,} \\ \phi = \phi_b \text{ such that } \int_{active} \sigma \partial_n \phi = 0 & \text{at the passive needles,} \\ \sigma \nabla \phi \cdot \mathbf{n} + \alpha \phi = 0 & \text{at the simulation boundaries,} \end{cases} \tag{1.2.1}$$

where ϕ is the linear static potential, ϕ_b is the potential of the boundary, σ is the tissue conductivity, ϕ_{src} is the electric potential at the electrode, $\mathbf{n}$ is the normal to the boundary and α is a weight.

This model is widely utilized in medical applications, particularly because of its conservative nature, which tends to underestimate the electric field strength. This built-in safety margin is crucial, as it guarantees that areas identified by the model as electroporated have indeed undergone successful and complete electroporation, thereby reducing the risk of partial or incomplete treatment. Such conservative estimation is invaluable in clinical environments, where precision and reliability are of the utmost importance. Ultimately, this model serves as a critical tool in achieving optimal therapeutic outcomes, providing interventional radiologists with the confidence that their procedures will be both effective and safe, contributing to improved patient care and advancing medical practices.

To complete the problem formulation, we impose Dirichlet boundary conditions on the active needles, ensuring a fixed potential that drives the electric field necessary for the procedure. Simultaneously, floating potential conditions are applied to the passive needles, allowing their potential to adjust dynamically in response to the surrounding electric field, thereby influencing the overall distribution of the field within the treatment area. Additionally, Fourier-Robin boundary conditions are employed on the simulation boundaries, which serve to model the exchange of electric energy between the simulated domain and its external environment, ensuring a more realistic and accurate representation of the physical system.

This comprehensive combination of boundary conditions is critical to capturing the intricate interactions that occur during the electroporation process. By accurately modeling the behavior of active and passive needles within the tissue

and maintaining the physical realism at the simulation boundaries, we can ensure that the numerical model provides a highly reliable prediction of the electric field distribution.

The effective evaluation of the treatment area in IRE ablation relies on several critical steps. First, accurate segmentation of the tumor is essential, as it allows for the adaptation of the tissue conductivity, σ, which is known to vary between malignant and healthy tissues. This step not only aids in adjusting the parameters for an improved electric field estimate but also ensures later that the tumor is fully encompassed within the electroporated zone, thereby reducing the risk of incomplete ablation.

Equally important is the localisation of the electrodes, which serves as the foundation for estimating the electric field. Precise localisation ensures that the electric field is appropriately mapped, minimising the risk of incomplete ablation or damage to surrounding tissues.

Lastly, the alignment of CBCTs, is critical. This step involves registering the initial imaging, which outlines the tumor's location and boundaries, with the subsequent scan that captures the actual position of the needles. This alignment is paramount for ensuring that the targeted tumor area in the planning stage corresponds precisely to the area treated during the procedure. By achieving this seamless fusion of data from different stages of the process, the clinician can be confident that the intervention is both accurate and comprehensive.

Together, these tasks—detailed further in the subsequent sections—are not merely technical operations but are integral to advancing the precision, safety, and efficacy of modern interventional radiology procedures. They play a crucial role in improving patient outcomes and pushing the boundaries of what is possible in minimally invasive cancer treatments.

In this thesis, we focus on two pivotal tasks: the precise localisation of the electrodes and the registration of imaging scans, both of which are fundamental to optimising the accuracy and success of interventional procedures.

1.2.1 Automated electrode localisation

This part was published in [47].

Our first contribution focuses on improving the automatic coarse segmentation of needles, a vital step for precisely delineating the electrode tip and subsequently estimating the electric field. The manual localisation of these electrode tips within a 3D volume is not only labor-intensive but also consumes considerably more time than the proposed automated method. Furthermore, this manual process demands the undivided attention of a radiologist, which can be particularly challenging when staffing is limited in the operating room. By streamlining this aspect of

the procedure, we aim to enhance efficiency and allow medical professionals to concentrate on critical tasks, ultimately facilitating better patient care.

We achieved a significant improvement in the electrode localisation by integrating deep learning into the existing pipeline. Specifically, we employed a convolutional neural network (CNN) architecture, thoughtfully designed to address the unique challenges posed by our data.

First, we adapted the U-Net architecture to suit the task of segmenting extremely thin objects in 3D images. This adaptation involved modifying the network to handle the intricate details and dimensions of needle-like structures. To further refine the process, we implemented a patch-selective training strategy, which significantly improved class balance by focusing on relevant sections of the images.

Moreover, we optimised the inference threshold to mitigate the effects of the under-representation of our target objects within the dataset. This was crucial for enhancing segmentation accuracy.

While various network configurations were explored, the simplest model consistently delivered superior performance, both in terms of Dice coefficient and training efficiency. This streamlined approach not only improved segmentation precision but also reduced computational demands, making it a robust solution for our specific application.

1.2.2 Partial fields of view registration

This part was published in [46].

Our second contribution introduces an innovative framework for auto-adaptive boundary conditions in solving registration problems. When dealing with partial fields of view, it is paramount to address the impact of mis-registration at the borders, which results from an incomplete neighborhood of the pixels along the perimeter. These boundary errors can propagate inward, significantly affecting the regions of interest. Our novel approach adapts the boundary conditions on a voxel-by-voxel basis, ensuring optimal results by mitigating the mis-registration at the boundary. We propose to inject flux information computed online in the boundary conditions and to estimate hyperparameters automatically through energy minimisation.

While this thesis primarily focuses on the intra-operative registration of two CBCT scans, the versatile method we propose holds significant potential for a wide array of image registration tasks. This includes multi-modal registration, which integrates different imaging modalities, as well as registration to anatomical atlases for enhanced diagnostic accuracy. The innovative boundary conditions we have introduced are designed to excel in scenarios with partial fields of view, ensuring that our approach minimises errors at the image periphery and enhances

overall registration accuracy. Consequently, the improvements derived from our method are expected to be particularly impactful in such challenging cases, offering substantial advancements in the precision and reliability of medical imaging.

We conducted a thorough analysis of the impact of the automatically adaptable boundary conditions applied during registration on the electric field estimate for one patient from our database. We show an enhancement in the quality of electric field estimation, a critical factor in ensuring effective treatment outcomes. Since all spatial data fed into the electric field model is derived from CBCT images captured during the procedure, registration becomes crucial in achieving a precise estimation of the treated area. The refinement of the registration process ensures that the tumor's position relative to the electrodes is determined with greater accuracy. A proper spatial alignment is essential for optimising the delivery of the electric field to the tumor, maximising its impact on the targeted tissue. Together, these advancements image registration contribute to a more precise, controlled, and effective treatment process, ultimately improving patient outcomes.

1.3 Thesis outline

The current section sets the context for this thesis work: the ablation of deep-seated tumor via IRE. We have identified various computer vision needs to facilitate the procedure, including object segmentation — to outline the electrodes, the tumor, etc — and image registration — to easily compare images taken in different modalities and/or at different times during the diagnosis and the procedure.

The remaining is organised in two chapters, one for each contribution, succeeded by a general conclusion.

The first chapter of this thesis presents a comprehensive literature review on the field of image segmentation, encompassing both traditional methodologies and cutting-edge deep-learning techniques. It delves into a detailed exposition of our innovative method designed to enhance the coarse segmentation of needles, which is crucial for the subsequent precise localisation of the electrode tip. This chapter compares the outcomes achieved through the proposed approach with those generated by the previously employed thresholding algorithm, highlighting significant improvements. It concludes with an in-depth discussion and analysis of these results, underscoring the advancements and potential implications of our method in the broader context of medical imaging and computational accuracy.

The second chapter is structured similarly to the first, beginning with a thorough review of existing methods for image registration, with a specific focus on boundary condition usage. This foundational overview sets the stage for introducing our innovative framework, which automates the adaptation of boundary conditions within the context of variational methods. This chapter not only delineates the

theoretical underpinnings of our approach but also provides comprehensive results derived from extensive testing across multiple databases. These results are analysed and discussed, highlighting the efficacy and robustness of the proposed framework in enhancing image registration accuracy and reliability. We additionally investigate the effect of such boundary condition applied in registration on the evaluation of the electric field during an IRE ablation. We detail the method for the estimation of the electric field as delivered by the electrodes. We then proceed to a thorough analysis and discussion of the impact on the precision and accuracy of treatment area estimation. This critical evaluation highlights the improvements and potential applications of our approach in various medical and technological fields.

Finally, we synthesize the findings, providing a conclusive discussion that underscores the significance of these results and outlines future directions for research and practical applications.

In brief

This work seeks to enhance the treatment of deep-seated tumors, such as those in the liver and pancreas, which are difficult to access with conventional methods. IRE ablation offers a promising alternative by using high-intensity electric fields to permeabilise cancer cell membranes and induce cell death. However, its complexity—demanding precise electric field application through multiple electrodes—requires careful planning and evaluation.

The objective of this research is to assess the real-time execution of IRE ablation, utilizing advanced computer vision techniques like image segmentation for needle localisation [47] and image registration to align intra-operative scans [46]. The dataset comprises abdominal CBCT scans, with MRI validation, and the study proposes innovative solutions for needle localisation and image registration, analysing the latter's impact on electric field estimation. By improving electrode placement accuracy and field calculation, this work aims to enhance the efficacy of IRE ablation in treating these difficult tumors, leading to better clinical outcomes.

Chapter 2

Needle localisation

In this chapter, we tackle the critical challenge of needle localisation within CBCT. This task is essential for accurately estimating the electric field, as electric pulses are delivered to the tumor site through the tips of the needles. However, the process is fraught with difficulties; the low signal-to-noise ratio and pervasive artifacts in the imaging can significantly hinder precise localisation. Navigating these complexities is vital for ensuring the effectiveness of the treatment, making our exploration of advanced methodologies in this area all the more important.

We propose to integrate deep learning into the original coarse-to-fine algorithm, enhancing the initial coarse segmentation of the needles. We employ a modified U-Net architecture, designed to address the unique challenges presented by the imaging modality and the characteristics of the objects of interest. A significant hurdle we faced was the pronounced class imbalance within the dataset. To effectively tackle this issue, we adopted an asymmetric U-Net design, which allows for greater emphasis on the high-level information processed by the decoder. Additionally, we crafted training patch sets specifically to bolster the representation of needles, and we fine-tuned the inference threshold to better balance the network's tendency to predict background voxels with higher confidence. This multifaceted approach not only improves segmentation accuracy but also enhances the overall robustness of our localisation method.

The CNN is subsequently combined with the Hough transform, yielding an analytical representation of the needle that can be seamlessly integrated into the electric field estimation process, though the object of interest is not directly visible on the CBCT due to artefacts. In contrast to the previous thresholding method, our approach demonstrates enhanced accuracy and stability, significantly improving the reliability of needle localisation. This advancement not only optimises the precision of the electric field estimation but also paves the way for a more accurate delineation of the treated area, ultimately contributing to better outcomes in clinical applications.

Contents

The method detailed in this chapter was published and orally presented in the peer-reviewed international conference MELECON:

- Eloise Inacio, Luc Lafitte, Olivier Sutter, Olivier Seror, Baudouin Denis de Senneville, Clair Poignard. *Automated needle localisation for electric field computation during an electroporation ablation* IEEE - MELECON 2022 - 21st Mediterranean Electrotechnical Conference, Jun 2022, Palerme, Italy (10.1109/MELECON53508.2022.9842866),

and orally presented at the World Congress on Electroporation 2022.

2.1 Role of needle localisation in IRE evaluation

To adequately assess the efficacy of the IRE procedure, and adapt the treatment plan in real time accordingly, accurate computation of the electric field is imperative. This necessitates precise and swift localisation of the electrodes responsible for administering the electric pulses. These electrodes serve as the primary source generating the electric field essential for electroporating cancerous cells. The accuracy of electrode placement directly influences the distribution and strength of the electric field, thereby impacting the treatment's outcome and safety. By ensuring precise electrode localisation, one can optimise the therapeutic efficacy of IRE while minimising potential risks associated with inaccurate field application.

Since their actual position differs from the theoretical one proposed by a specialised algorithm, to achieve greater accuracy, this data is extracted from medical images, specifically through the use of CBCT within our operational workflow. CBCT proves highly practical in surgical settings, delivering minimal radiation doses that enable rapid, successive scans by radiologists. However, the modality is challenged by its inherent drawbacks, such as a low signal-to-noise ratio and the presence of artefacts. These complexities pose significant challenges for computer vision tasks aimed at extracting reliable information from CBCT images.

There are two primary types of segmentation tasks in computer vision. The first, semantic segmentation, involves assigning each pixel in an image to a specific class, effectively creating a detailed map where each pixel is labeled according to its corresponding object or region. This process results in a segmentation mask that identifies and distinguishes between different classes, allowing for precise analysis and understanding of the visual content. The second, instance segmentation, similarly attributes a class to each pixel but also differentiate between different objects belonging to a same class (Fig 2.1). Since we do not need to identify the needles individually, we focus on instance segmentation for the remaining of this work.

In this chapter, we present a novel coarse-to-fine localisation algorithm that seamlessly combines the advanced capabilities of deep learning with the time-tested robustness of the Hough transform technique. This innovative approach is specifically designed to process CBCT scans, enabling the accurate analytic identification of the electrode tip responsible for delivering the electric field.

By leveraging deep learning for the initial coarse segmentation, we capture the broader region of interest with efficiency and speed. Following this, the Hough transform is employed for precise fine-tuning, allowing for the exact localisation of the electrode tip. This strategic integration not only ensures accuracy in pinpointing the tip's position but also enhances the overall precision of the electric field estimation, making the method both reliable and adaptable for clinical ap-

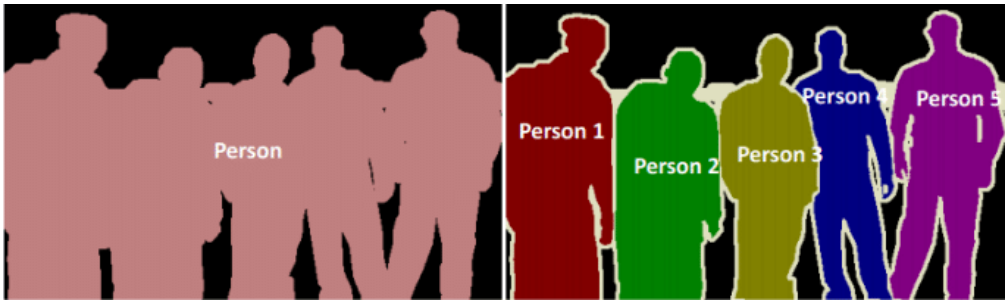


Figure 2.1: Semantic segmentation (left) differs from instance segmentation (right) in that it assigns a single label to each class of object throughout an image, whereas instance segmentation provides distinct labels for each individual instance of an object [93] ©2019.

plications. In doing so, our approach bridges the gap between modern machine learning techniques and established analytical methods, offering a powerful solution for high-precision medical imaging tasks.

The main challenges are the nature of the modality—CBCT are practical as they are fast and low-dose but the capture has a low signal-to-noise ratio—, the nature of the needles—electrodes are very dense, so they provoke extensive artifacts on the scan and appear thicker—and the task itself—the segmentation of very thin objects takes inherently place on highly skewed data. Indeed, across the dataset, there is one voxel belonging to needles for 10^4 voxels belonging to the background on average.

This chapter is structured as follows. Section 2 offers a comprehensive literature review on segmentation techniques, beginning with traditional approaches and culminating in advanced deep-learning-based methods. Section 3 deals with the clinical data, including the ground-truth generation for a deep learning application, the data preprocessing and the assesment of the method's efficacy. Section 4 delineates the contribution of this chapter: a CNN based coarse segmentation of the needles. Subsequently, section 5 delves into the fine localisation of the needles, using a Hough transform. Section 6 encapsulates the key contributions of this work and outlines potential future directions for research in this domain. Finally, section 7 concludes on the proposed method and its results.

2.2 Object segmentation in the literature

This section explores the various existing segmentation methods, with a focus on applications in medical imaging. We begin with traditional techniques, such as thresholding and contour optimisation, which have long been used for their sim-

plicity and effectiveness in specific scenarios. These foundational methods lay the groundwork for understanding the evolution of segmentation approaches. We then transition to advanced deep-learning-based techniques, which have revolutionized the field by offering unprecedented accuracy and adaptability. These cutting-edge methods leverage complex algorithms and datasets to achieve precise and robust segmentation, addressing limitations of traditional approaches and opening new frontiers in medical image analysis.

2.2.1 Traditional techniques

In this subsection, we delve into the two principal categories of traditional segmentation algorithms: level set-based methods and region growing methods. Level set-based methods are renowned for their ability to handle complex topological changes, allowing for the precise delineation of object boundaries even in challenging scenarios. On the other hand, region growing methods start from a seed point and iteratively expand the region by including neighboring pixels that share similar properties, making them particularly effective for segmenting homogenous regions. Both approaches have been pivotal in advancing the field of image segmentation, each offering unique advantages that cater to different types of segmentation challenges.

Level set based methods: Rooted in edge detection theory, level set-based methods possess an inherent sensitivity to noise. Despite this challenge, these techniques are known for their ability to track and delineate moving fronts and dynamically evolving shapes with precision. In the realm of level set methods, contours are represented as the zero level set of a higher-dimensional hypersurface. This conceptual framework allows the curve's evolution to be analogous to the transformation of a three-dimensional level set function. By leveraging the dynamic nature of the level set function, these methods adeptly handle the intricacies of varying shapes and structures, ensuring that segmentation remains both precise and adaptable. Thus, despite their sensitivity to noise, level set-based methods stand out for their robust capability to manage and interpret intricate, evolving geometries in advanced imaging applications.

The common movement formula, governing the behaviour of the contour, is:

$$\frac{\delta\varphi}{\delta t} + F|\nabla\varphi| = 0, \tag{2.2.1}$$

where t is the time, φ is the smoothing function representing a surface, and F is the speed function describing the evolution of the surface and depending on image characteristics.

The level set function is ideally maintained near a signed distance function, which provides a smooth representation of the evolving contour. However, it seldom attains a precise value of zero on the desired edge, as its accuracy fundamentally relies on the image gradients. These gradients guide the contour evolution, influencing the positioning and shape of the segmented boundaries.

A variational formulation is possible where the energy to optimise consists of an internal energy P, penalising deviation from a signed function, and an external energy $e_{\mathbb{G}}$, driving motion towards image features:

$$\mathbb{E}(\varphi) = \alpha P(\varphi) + e_{\mathbb{G}}(\phi) \tag{2.2.2}$$

where α is a weight and $\mathbb{G}$ is the edge indicator function.

Narrow band methods excel by reducing computational costs through selective level set updates. However, their effectiveness hinges on the dynamic management of narrow band points, which poses a significant challenge. Ensuring timely and accurate updates in response to evolving contours or features requires sophisticated algorithms and careful parameter tuning. This dynamic adaptation is crucial to maintain the method's efficiency and accuracy in various applications, from medical imaging to computer vision tasks.

More recently, homogeneity-based level set methods have emerged to achieve optimal image partitioning by leveraging regional information rather than edge detection. They exhibit robustness against noise and have the capability to detect objects characterised by smooth gradients or no sharp edges. However, their application is constrained by the formulation's computational demands, particularly in scenarios requiring segmentation into more than two regions. This limitation arises from the inherent complexity of maintaining homogeneity across the image, which necessitates intensive computational resources and careful parameter tuning for effective implementation. The new energy functional is:

$$\begin{aligned}\mathbb{E}(c_1, c_2, C) = \alpha_1 Length(C) + \alpha_2 Area(inside(C)) + \\ \alpha_3 \int_{inside(C)} |I(\mathbf{r}) - c_1|^2 \mathbf{dr} + \alpha_4 \int_{outside(C)} |I(\mathbf{r}) - c_2|^2 \mathbf{dr},\end{aligned} \tag{2.2.3}$$

where α_i are fixed weights, $I(\mathbf{r})$ is the intensity of the image at $\mathbf{r}$, c_1 and c_2 are the intensity best describing the two partitions of the image.

A worth-mentioning subcategory is contour optimisation, which represents a sophisticated blend of automation and human interaction in image processing. This method offers significant flexibility by allowing manual adjustments while leveraging automated algorithms to refine contours. Despite its semi-automated nature, contour optimisation retains a degree of subjectivity stemming from the reliance on manual inputs for final adjustments. This approach is particularly valuable in tasks requiring precision and nuanced adjustments that automated

methods alone may not achieve effectively.

Such a standard approach is snake algorithms where a parametrised curve or surface, in 2D and 3D respectively, provided manually, iteratively evolves according to an energy minimisation criteria [52]. This energy formulation integrates an internal and an external energy, which are further refined by internal bending forces within a state space framework. However, their primary limitation lies in their inability to effectively handle significant topological changes. When objects undergo substantial transformations or deformations that alter their shape drastically, snake algorithms struggle to adapt the contour appropriately, leading to inaccuracies in segmentation results.

In [81], they associate a genetic algorithm, with selection, mutation and recombination of candidates, with a snake algorithm. The overall order complexity is satisfyingly low: $O(n\lambda G)$ with n the number of control points, λ the number of individuals per generation and G the number of generations. The selection process employs a rank-based selector, evaluating candidates based on differences in their performance. This approach effectively applies selective pressure, ensuring optimal candidate selection even when energy disparities are minimal. The snake, characterised by discrete control points, navigates this intricate landscape with precision. Operating within a vast non-convex search space, the algorithm avoids local minima, thus enhancing its robustness in capturing global optima. This capability is crucial in tasks where complex geometries or highly variable environments require accurate and reliable solutions.

Two variations of the genetic active contour algorithm have been introduced. In the supervised version, comprehensive global parameters are initially computed, evaluated by the active contour algorithm to refine their effectiveness. Contrarily, the unsupervised approach employs a nuanced strategy where the genetic algorithm dynamically computes localised parameters specific to each control point. These refined parameters are then seamlessly integrated into the active contour algorithm's iterative process, ensuring precise adjustments in response to image features.

Compared to traditional generic snake algorithms, both methodologies demonstrate notable advancements, consistently yielding satisfactory outcomes in various segmentation tasks. These advancements underscore the algorithm's adaptive capabilities in optimising parameterization and contour refinement, enhancing its applicability across diverse image processing applications.

Region growing based methods: Region growing based algorithms leverage a collaborative strategy that integrates both region and line detection methodologies, effectively navigating the complexities of pixel interactions within an image. By dynamically examining the connectivity and similarity of neighboring pixels, these

algorithms construct coherent regions based on predefined criteria. This methodical process not only identifies distinct regions but also delineates their boundaries with precision.

At the foundation of region growing based image segmentation, lies an elementary algorithm that clusters pixels based on their immediate attributes, initiating the process of delineating distinct areas within an image. As segmentation techniques advance, more intricate methodologies emerge, evolving from basic pixel aggregation to the fusion of rudimentary regions. These advanced approaches necessitate a refined framework capable of accurately defining regions, precisely outlining boundaries, and seamlessly integrating segmented areas.

Local techniques: They delve into the details of individual pixels or their immediate surroundings, analysing properties such as color, intensity, and texture. Blob coloring exemplifies this approach, where contiguous groups of pixels with similar attributes are categorised and often distinguished from their surroundings.

Global techniques: They harness the collective attributes of a vast array of pixels across an entire image, aiming to delineate meaningful partitions effectively. These methods analyse extensive pixel data to establish optimal thresholds, such as those employed in thresholding algorithms.

Splitting and merging techniques: The manipulation of regions involves intricate processes facilitated by region adjacency graphs (RAGs). These graphs provide a structured framework where each node symbolises a distinct region identified within an image. The edges between nodes in the RAG denote adjacency relationships, capturing the spatial connectivity between regions. This approach allows for efficient merging and splitting operations based on predefined criteria, such as similarity in pixel attributes or geometric properties. By leveraging RAGs, algorithms can systematically analyse and transform regions, adapting them to the complexities of different images and tasks.

2.2.2 Deep-learning techniques

Machine learning, and more specifically deep learning, has been increasingly used for many image processing tasks such as object recognition, segmentation, classification, or target tracking. This recent surge is attributed to significant advancements in graphics processing units (GPUs), which have enabled more efficient computations. Consequently, this has facilitated the development of increasingly complex and deep neural network architectures capable of handling intricate visual data. Additionally, the expansion in the availability and diversity of datasets has played a pivotal role. Larger and more varied datasets provide ample training examples, allowing deep learning models to generalise better and achieve

higher accuracy in diverse applications across various domains.

Unlike most traditional techniques, deep-learning methods extract features automatically by noticing patterns in data (or tasks for reinforcement learning). These features are often more effective than hand-crafted features [80] and less laborious to set in place.

However, for methods requiring ground truth (supervised and semi-supervised learning), the segmentation quality strongly depends on the ground-truths precision [103]. Also, for 3D imaging, this kind of methods has high computational and memory costs, calling for a careful implementation.

Amongst machine learning methods for object recognition, and more specifically object segmentation, CNNs are dominant due to their better generalisation ability [44] [62] and their ability to retain spatial coherence. Convolutional networks consist in layers of operations, most of which are parametrisable.

Fully-connected layer: Each input x_i is linked to each output y by a connexion with a weight. The overall output is computed in considerations of those weights w_i, the inputs x_i and the bias b. A scheme of the connectivity is presented in Figure 2.2 and the exact computation of the output from each cell is as follows:

$$y = \sum_i w_i \cdot x_i + b. \tag{2.2.4}$$

where i is the input number.

Convolution: The input data is processed using a learnable kernel. The kernel size determines the receptive field of the convolution, influencing the feature extraction capabilities of the layer. Additionally, padding is employed to adjust the spatial dimensions of the input, ensuring that the output dimensions match the input when needed. This padding can be applied symmetrically to maintain spatial resolution or asymmetrically to adjust output dimensions as required by the network architecture. These fixed parameters play a crucial role in defining how information is extracted and propagated through the network, impacting the network's ability to learn hierarchical representations from the input data.

Normalisation: normalisation in neural networks plays a critical role in stabilizing training dynamics and enhancing performance. By mitigating the impact of differing feature scales and internal covariate shifts, normalisation methods ensure smoother gradient descent trajectories and faster convergence [1] [2]. This not only expedites training but also improves the model's ability to generalise [3]. It is a fundamental technique that contributes to the overall efficiency and effectiveness

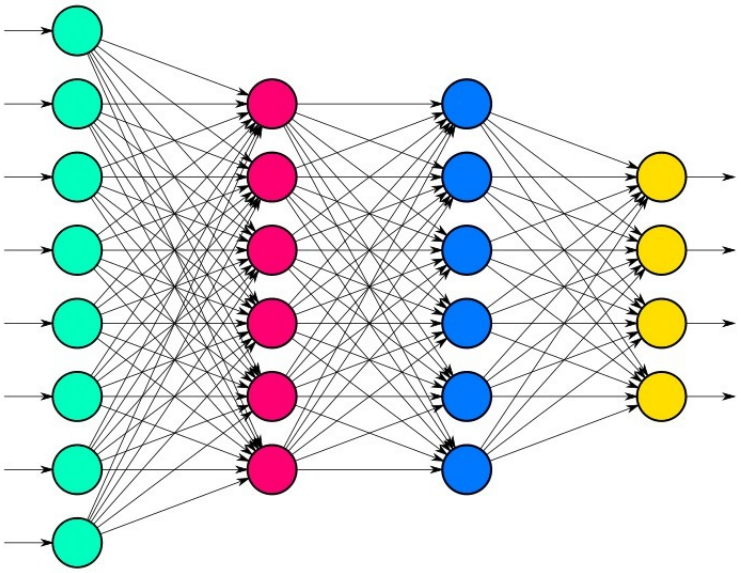


Figure 2.2: In a fully-connected layer, each neuron in a given layer (excluding the input layer, which is represented in green) is interconnected with every neuron from the preceding layer. When the input is an image, it undergoes a process of flattening, where each individual pixel of the image is mapped to a corresponding neuron in the green input layer, reproduced with permission.

of deep learning models. In the most common normalisation, namely batch normalisation, the inputs x_i, often sent in batch to the network, are normalised with learnable shift β and scale γ:

$$\hat{x}_i = \gamma \frac{x_i - \mu_i}{std_i} + \beta, \tag{2.2.5}$$

with μ_i the mean and std_i the standard deviation of the i-th layer.

Non-linearity: Non-linear activation functions are pivotal in neural networks because they introduce non-linearities, allowing the network to learn and approximate complex mappings between inputs and outputs. These functions enable neural networks to capture intricate patterns and relationships that linear functions cannot represent effectively. Rectified Linear Unit (ReLU), which sets negative values to zero and passes positive values unchanged, is favored for its simplicity and efficiency in training deep networks:

$$ReLU(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{otherwise.} \end{cases} \tag{2.2.6}$$

Pooling: Pooling layers are integral in CNNs for spatial down-sampling, which helps manage computational resources and control over-fitting. These layers operate by dividing the input into smaller regions and applying a function (i.e., maximum or average) to aggregate information within each region. On the one hand, max pooling retains the most significant feature within each neighborhood, enhancing robustness to variations in the input. On the other hand, average pooling provides a smoother representation by averaging values, useful in scenarios where fine-grained details are less critical.

Dropout: Dropout is a regularisation technique that combats over-fitting by randomly deactivating neurons or features during training. This stochastic process prevents the network from relying too heavily on specific neurons or features, ensuring that the model generalises well to unseen data.

In the following, we only consider supervised learning. This training method requires data and ground-truth to optimise the parameters. First an objective (or loss) function is defined to measure the distance between the outputs of the network and the ground-truth. A parametrisation giving a minimum (ideally the global minimum) of this loss function is then obtained by retro-propagating the loss through the layers of the network, starting from the end. Each step, the parameters are updated following, for instance, the gradient descent algorithm. The parameters are modified in the direction opposite the gradient, thus towards a minimum:

$$w_{n+1} = w_n - l_r \nabla L(w_n), \tag{2.2.7}$$

where w_n is the parameter at iteration n, l_r is the learning rate and L is the loss function to optimise.

The iterative process of the algorithm continues until a predefined stopping criterion is met, often determined by a specified number of iterations or the convergence of certain parameters. This ensures that the algorithm reaches an optimal state or a satisfactory approximation. Once in the inference phase, the parameters remain static, and the algorithm applies its learned model to new data without needing ground-truth labels for prediction.

Enhancing performance often hinges on integrating skip connections or leveraging context vectors derived from pooling operations across the entire image. On the one hand, skip connections facilitate gradient flow and enable the network to retain fine-grained details essential for accurate segmentation or classification tasks. On the other hand, context vectors synthesized from global pooling capture holistic image context, aiding in robust feature extraction and improving spatial awareness. Both mechanisms synergistically improve CNN capabilities by balancing local and

global information flow, thereby elevating overall model efficacy and adaptability.

A number of relevant architectures have been proposed to solve the segmentation problem. SegNet's unique up-sampling method aids in retaining spatial information crucial for accurate pixel-wise predictions. HRNet's strategic use of skip connections ensures that fine details are preserved throughout the hierarchical processing stages. VNet's adoption of the Dice coefficient for loss computation underscores its effectiveness in handling imbalanced data distributions in 3D contexts. Tensormask's novel approach with 4D tensors exemplifies how segmentation can benefit from predictive modeling and strategic data handling techniques [68].

However, the most used architecture in deep-learning method for medical image segmentation is undoubtedly the U-net architecture. It consists in an encoder-decoder structure as depicted in Figure 2.3. On the one hand, the contracting path takes into account context and reduces the input to a representation in a smaller latent space, behaving much like a classification CNN—accumulating semantic information while losing spatial infomation. On the other hand, the symmetric expanding path allows better localisation of objects of interest and recovers an output of similar dimension to the input from the smaller representation. Moreover, skip connections permit the combination of high resolution information from the encoder, with the up-sampled features of the decoder. A convolutional layer then learns to merge the activation features in a sensible way. In the original work [78], only the valid part of the convolution was used, inducing a need for the cropping of the encoder feature maps before concatenation with the decoder feature map, and an extrapolation by mirroring the input to obtain a segmentation mask of the same size as the image to segment. In the encoder, down-sampling is done via max pooling while in the decoder, up-sampling is done by up-convolution.

On the learning side, batch training is performed to minimise the overhead. The incorporation of a pixel-wise softmax in the loss function ensures precise probability assignment per pixel, crucial for detailed segmentation tasks. Weighted cross-entropy then adjusts these probabilities based on class frequencies in the ground truth, prioritising learning in under-represented regions. This methodology not only improves boundary delineation but also enhances overall model performance in semantic segmentation applications.

It was initially proposed for the segmentation of structures in microscopical images, specifically targeting neuronal structures and cells. The data augmentation pipeline is intricately tailored for this purpose, emphasising techniques like elastic deformation to simulate realistic variations in microscopic images. Additionally, the network incorporates implicit augmentation through drop-out regularisation at the end of its contracting path, enhancing generalisation capabilities.

This pioneering network, known for its efficacy in biomedical image segmentation, achieved significant recognition by winning the ISBI cell tracking challenge in 2015,

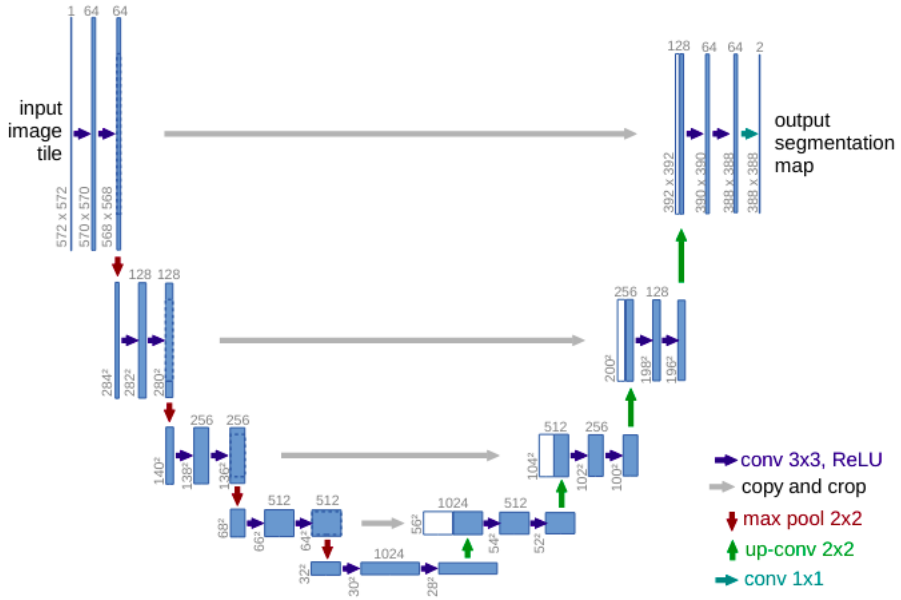


Figure 2.3: Original U-Net architecture proposed in [78], reproduced with permission. It features encoding and decoding paths linked by skip connections, which enhance the flow of information by directly passing outputs from earlier to later layers, preserving detailed spatial features.

underscoring its superiority in accurately delineating cell boundaries and neuronal structures

In [103], a state-of-the-art modified U-net architecture is applied for real-time needle localisation in ultrasound imaging. The inherent challenges of this modality, including a low signal-to-noise ratio, presence of artifacts, and speckle effects, necessitate advanced computational techniques for accurate localisation.

Ground truth data is derived through expert annotations of needle centers across selected ultrasound slices. These landmarks undergo linear interpolation to construct a detailed needle representation. Subsequently, this representation is expanded to accurately reflect the needle's true diameter, enhancing the model's robustness and precision in needle localisation tasks.

The improved architecture, shown in Figure 2.4, incorporates specialised attention gates to effectively mitigate false positives. This enhancement is crucial given the ultra-thin nature of needles, akin to those utilised in IRE ablation procedures. The attention gates are designed to refine the model's focus on intricate details, such as needle boundaries and subtle features amidst noisy ultrasound images. However, they introduce additional parameters for the model to learn, specifically the attention coefficients, making the network heavier.

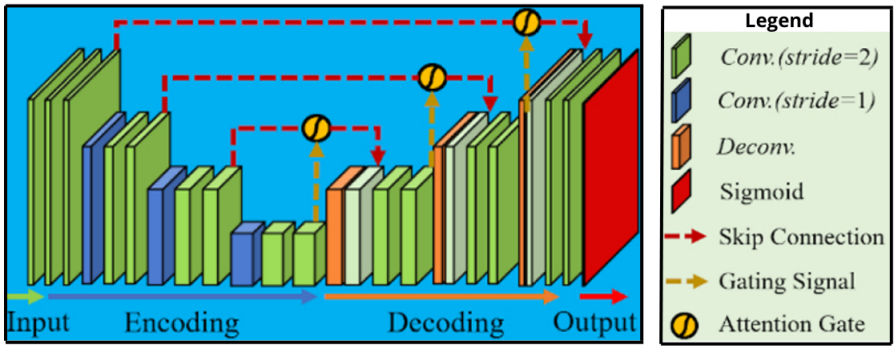


Figure 2.4: The U-Net architecture has been adapted to incorporate attention gates, as detailed in the study [103]. This modification enhances the network's ability to focus on relevant features by incorporating mechanisms that selectively emphasise important regions of the input data, reproduced with permission.

Also, the conventional composite loss function is substituted with a synergy of total variation, which enhances spatial coherence by selectively boosting the vertical axis—specifically aligned with the insertion direction of needles—and cross entropy. This novel approach not only refines boundary detection but also optimises segmentation fidelity in complex ultrasound images.

Moreover, the integration of deep supervision introduces companion objective functions tailored for each hidden layer. This strategic augmentation not only facilitates the extraction of more discerning features but also mitigates the challenge of vanishing gradients, thereby fortifying the network's capacity to achieve robust and precise needle localisation.

The proposed method led to a statistically significant improvement compared to the regular U-Net. 96% of the needles in their database were located within 0.442 ± 0.831 mm at the tip, where the radioactive source is located.

Oftentimes, good implementation and configuration are more important than the architecture itself [49]. This is why taking into consideration the importance of pre-processing, post-processing and adapted learning strategy is crucial. To adress this issue, [48] provides a full self-adapting pipeline creating a suitable CNN from a vanilla U-Net (2D, 3D or cascade, where a first network is fed a down-sampled version of the input, the resulting smaller segmentation map is then up-sampled and concatenated with the original input to segment, before being fed to another network for refinement, see Figure 2.5).

The proposed pipeline goes as far as adjusting the loss function (a multi-class dice loss with cross entropy), the optimiser settings (ADAM optimiser with adaptable learning rate), the data augmentation parameters, etc. The pre-processing consists in cropping out eventual padding and resampling with a third order spline

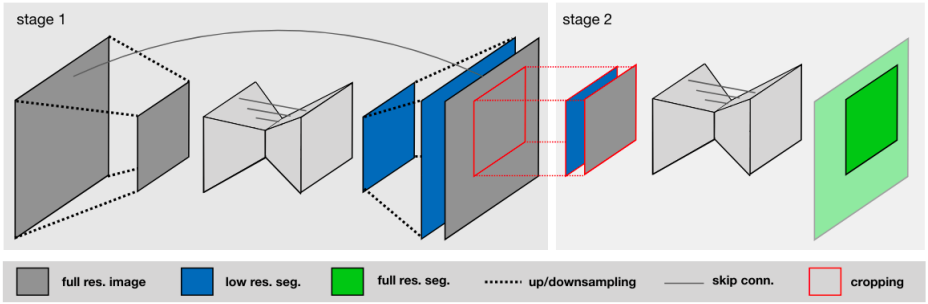


Figure 2.5: The vanilla cascade network, based on a U-Net architecture, employs a sequential arrangement of multiple neural networks. In this setup, each network's output serves as the input for the subsequent network, once concatenated with the original image. This process continues through the sequence, with each successive network operating at progressively higher resolutions [48], reproduced with permission.

interposation to ensure the same voxel spacing while the post processing keeps the largest connected components in the segmentation masks. 5-fold cross validation is always performed and all the resulting networks are used as an ensemble to increase robustness. This framework leads to state-of-the-art results on various segmentation tasks without the need for manual tuning of the various parameters that come with CNNs.

To address the substantial memory demands of processing 3D images, innovative 2.5D architectures were devised. These sophisticated networks allow for deeper and broader configurations, enabling comprehensive analysis of 3D datasets through multiple cross-sectional slices. Moreover, in the realm of anisotropic data, where voxel dimensions vary across different axes, 3D segmentation methodologies often exhibit inferior performance compared to their 2D counterparts [48], thus motivativating the resampling step in the nn-Unet.

A network architecture is proposed in [70], designed around a straightforward encoder-decoder framework featuring a U-Net-like decoder (see Fig. 2.6). This architecture employs strided convolution for efficient down-sampling. To enhance the network's performance, both bottleneck residual connections (illustrated in Fig. 2.7) and basic block residual connections are incorporated. These additions bolster the network's ability to preserve important features and gradients. Additionally, the inputs are upscaled using two-fold bilinear upscaling, addressing the challenges associated with segmenting small, critical objects such as brain metastases.

Furthermore, they employ mixup augmentation on a single dataset to subsequently validate their model across multinational datasets comprising hundreds of brain MRI scans. This approach aims to rigorously assess the model's resilience and

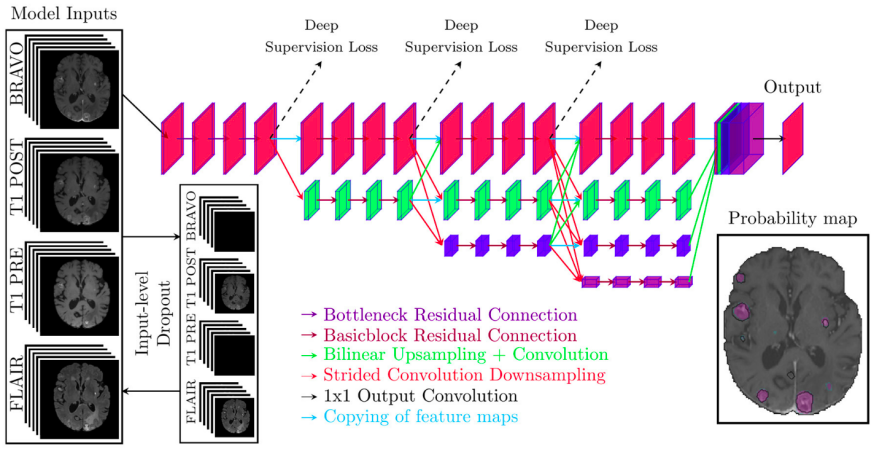


Figure 2.6: The modified U-Net architecture designed for 2.5D segmentation, as proposed in [70], incorporates several advanced features: deep supervision to enhance the learning process, bottleneck structures to optimise computational efficiency and feature extraction, and both basic and residual connections to improve gradient flow and overall network performance, reproduced with permission.

its ability to generalise beyond specific data sources. Prior to analysis, preprocessing involves brain extraction to isolate relevant structures and coregistration to ensure spatial alignment for accurate comparative evaluations.

The loss function is a compound loss with batch-wise focal Tversky loss and weighted binary cross entropy—where additional weight is attributed to metastases slices. This puts emphasis on true positives to handle class imbalance. To further reduce memory requirements, mixed precision training, where inputs are randomly cropped, is used. The 2.5 CNN was trained alongside its 3D counterpart for comparison. The authors end up recommending the 3D version as they observed a considerable reduction in false positives while maintaining a satisfying sensitivity.

Transformers, a prominent category of deep-learning models, have gained traction in medical image segmentation due to their capacity for capturing global dependencies. Despite their rising popularity, transformers typically exhibit lower performance compared to CNNs under identical data augmentation conditions [49]. Likewise, they may suffer from large memory and computational requirements, especially when processing 3D data.

This method comes from the field of natural language processing and is based on self-attention [62]. Early approaches integrated CNNs with transformers' self-attention mechanisms to enhance spatial context awareness. Over time, this hybrid model transitioned into vision transformers, where traditional convolutional layers were entirely substituted with self-attention layers. In this advanced paradigm, im-

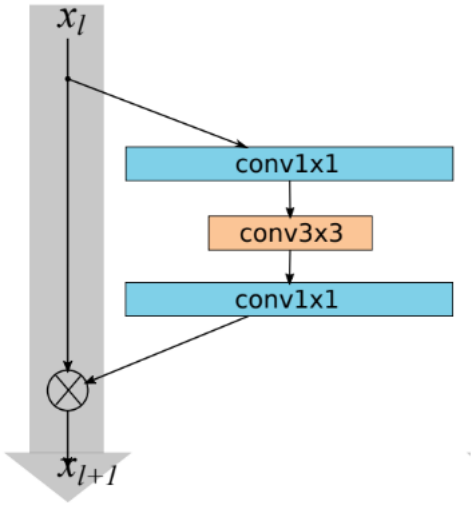


Figure 2.7: The bottleneck residual connection incorporates a residual block that includes a "bottleneck" structure, where a sequence of convolutional layers is compressed into a narrower dimensionality before being expanded back to its original size. This approach allows the network to more effectively learn and retain important features by facilitating smoother gradient flow and reducing computational complexity, while still capturing intricate patterns in the data [40], reproduced with permission.

ages are segmented by dividing them into manageable patches that are sequentially processed by the transformer. This patch-based approach allows the transformer to effectively capture global dependencies across the entire image, enabling robust and accurate segmentation

The vanilla transformer has an encoder-decoder structure made of transformer block, i.e. a multi-head self-attention to attend multiple parts of the input and a position-wise feed-forward network [72] (Fig. 2.8). The integration of residual connections and layer normalisation within both encoder and decoder components enhances computational efficiency and gradient flow. Notably, while both encoder and decoder share these optimisation techniques, the decoder introduces additional mechanisms essential for sequence generation tasks. Specifically, it employs attention masks to prevent the model from attending to future tokens during training. Moreover, positional encodings are incorporated to impart sequential context, crucial for tasks like language translation and image captioning.

The loss function in transformer-based models often employs a combination of various loss components tailored to specific tasks, known as compounded loss. This comprehensive approach ensures that the model optimises across multiple

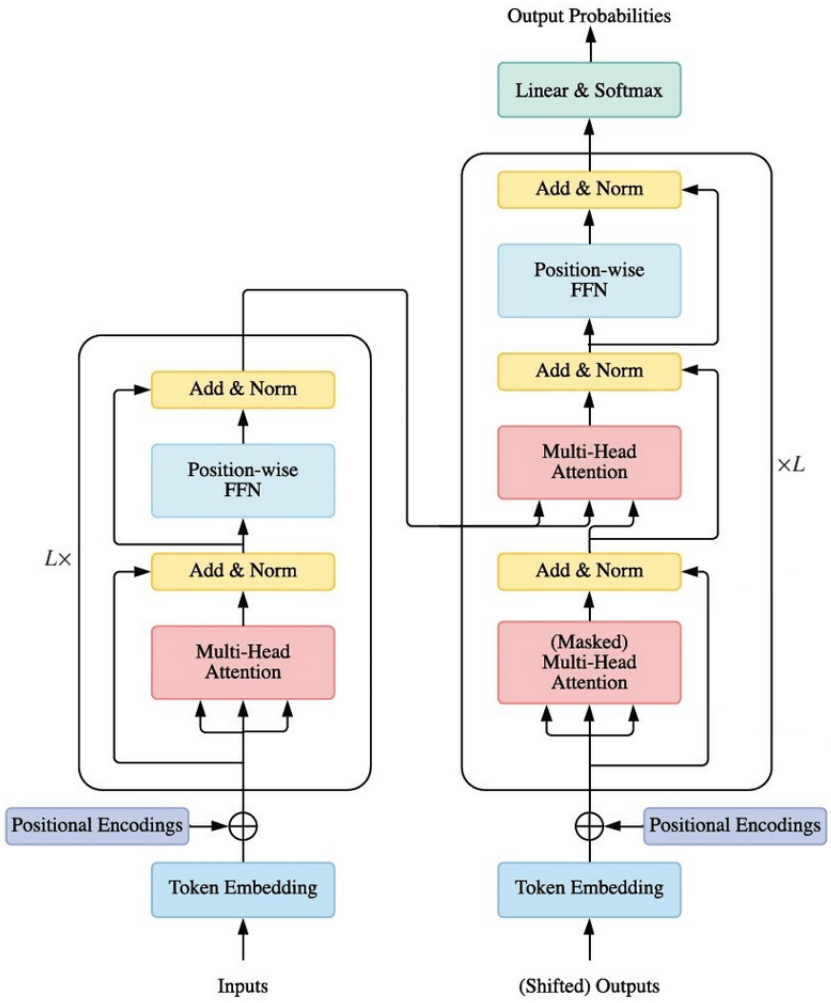


Figure 2.8: The vanilla transformer architecture comprises two primary components: the encoder, positioned on the left side of the diagram, and the decoder, on the right side. The encoder processes and transforms the input data into a contextual representation, which is then utilised by the decoder to generate the final output [72], reproduced with permission.

objectives, balancing accuracy and robustness in predictions.

Newest designs for semantic segmentation include the use of category queries with cross-attention in the decoder to better model global context [99], multiple class tokens with patch tokens to obtain multi-class attention map [98], quad-tree transforms for fine boundary/object [56] and dual frameworks combining transformer and CNN to look into global and local context respectively [101].

Integrating traditional graph models like conditional random fields (CRFs) and Markov random fields (MRFs) with deep-learning approaches enhances the contextual understanding and precision of segmentation tasks. These models leverage structured dependencies among labels, refining predictions by considering neighbouring pixel relationships.

In the DeepLabv2 architecture, the synergy between a CNN equipped with dilated convolutions and a probabilistic graphical model further enhances localisation accuracy. By integrating these techniques, DeepLabv2 not only captures fine details through dilated convolutions but also refines segmentations based on global context provided by the graphical model [68].

2.2.3 Conclusion on the literature review

Building on insights from previous research, it is evident that deep learning methods remain the most effective approach for fundamental computer vision tasks, such as coarse segmentation. Among these techniques, CNNs continue to deliver superior performance, even in the face of significant advancements in transformer-based models. The U-Net, in particular, has garnered substantial attention due to its powerful yet straightforward architecture, making it a favored choice in segmentation tasks.

While numerous U-Net variants—incorporating elements such as attention gates or deep supervision—have shown improved performance in specific cases, these enhancements are often the result of highly customised modifications tailored to particular datasets or challenges. Given this, we have chosen to begin with a basic U-Net architecture, adapting it to suit the unique requirements of our task: the precise segmentation of extremely thin structures. This pragmatic approach allows for a balance between leveraging the proven strengths of the U-Net and incorporating targeted adjustments necessary for our specific application.

2.3 Handling the clinical data

2.3.1 Ground-truth generation

Given that we are leveraging supervised learning to train our CNN for needle segmentation on CBCT images, it is imperative to have accurate ground-truth data for comparison during the optimisation process. However, the current data provided by radiologists is limited to the coordinates of the needle tip, which falls short of the comprehensive information needed for effective segmentation learning. Consequently, we must undertake the creation of our own detailed ground-truth annotations to ensure the network's ability to accurately segment the needles in

the images.

To do so, we utilise a semi-automatic segmentation technique, as implemented in ITK Snap [100], combining a traditional method introduced in section 2.2.1, i.e. the snake algorithm, with machine learning. We manually placed spheres along the needles that are later iteratively optimised to fit the shape of the needle as seen on the CBCT (Fig. 2.9). To do so, a random forest classifier is trained using features extracted from the coarse labels placed by the user. The active contour algorithm is then applied on the segmentation proposed by the classifier. The contour C evolves such that:

$$\frac{\partial C}{\partial t} = (s(C) + \alpha\kappa_C)\mathbf{n}, \tag{2.3.1}$$

with κ_C the mean curvature, imposing smoothness of the solution, s the speed function, $\mathbf{n}$ the unit outward normal vector and α a scalar parameter [100].

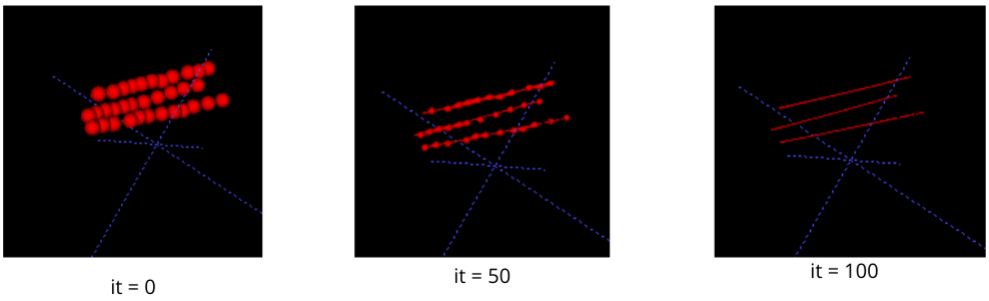


Figure 2.9: Ground-truth generation for a patient in our database is achieved using the snake algorithm implemented within ITK-Snap [100]. At iteration zero (it=0), the manually placed spheres are displayed as initial reference points. As the algorithm progresses through successive iterations, we observe the gradual refinement and convergence of the contour optimisation process, resulting in an increasingly accurate representation of the target structure: the electrode.

2.3.2 Data preprocessing

In machine learning, the partitioning of data into training and test sets is fundamental for robust model development and evaluation (Figure 2.10). The training set, comprising data from 8 randomly selected patients in our scenario, serves as the foundation for refining the algorithm through iterative adjustments and parameter tuning. This iterative process ensures that the model learns patterns and correlations within the data.

Conversely, the test set, consisting of the remaining 8 patients, acts as an impartial measure of the model's performance on unseen data. This critical

evaluation step simulates real-world scenarios, gauging how effectively the algorithm generalises to new, unseen patient data. By maintaining this separation, we mitigate the risk of over-fitting, where the model performs exceptionally well on training data but poorly on new data. Thus, the careful partitioning into training and test sets ensures the reliability and applicability of machine learning models in healthcare and beyond.

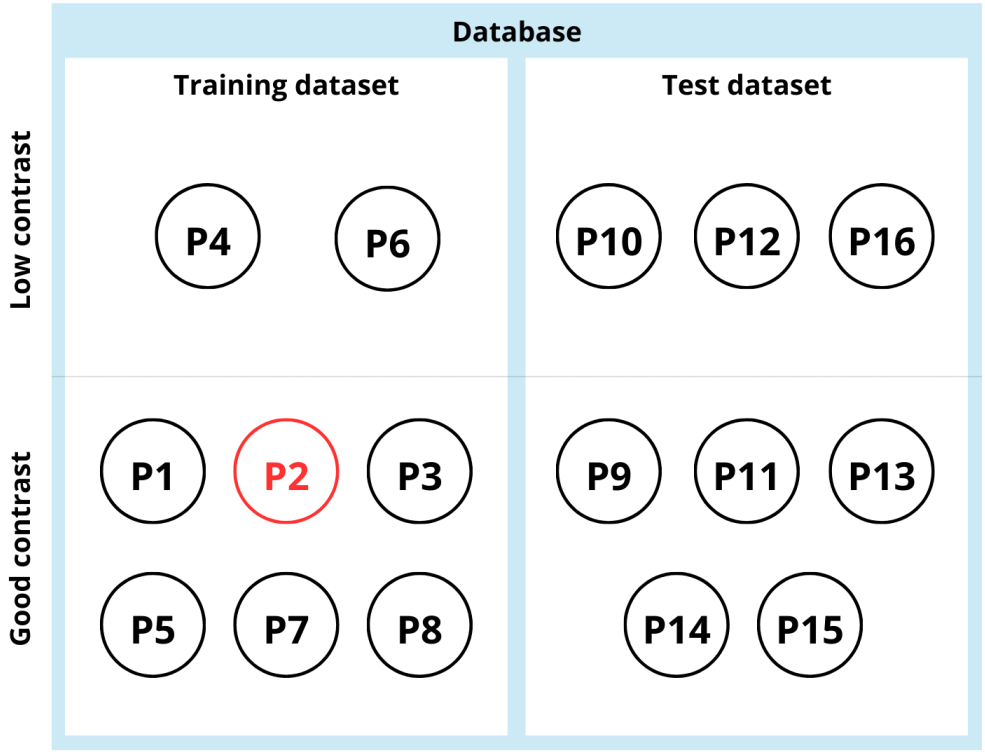


Figure 2.10: Partitioning of the dataset into a training dataset and a test dataset. The patient in red was used for validation. Additionally, information about the contrast is provided, where the contrast is considered low when the signal to noise ratio, defined as $SNR = \left(\frac{\mu}{std}\right)^2$ over a patch of the image, is lower than 5.

Furthermore, the dimensions of the scans are standardised, the dimension in the superior-inferior direction being otherwise between 195 and 512 voxels.

Finally, by standardising the scaling process across the entire dataset, consistency is maintained in the representation of intensity values. This approach ensures that features of interest, such as anatomical structures or pathological findings, are uniformly visualised and interpreted. Furthermore, applying uniform scaling improves comparability between different scans and facilitates accurate quantitative analysis. In our case, the values are thresholded between -999 (the padding

value for the cylindrical CBCT to be stored as a cubic matrix) and 1500 before normalisation.

2.3.3 Assessment

The Dice coefficient serves as a metric in evaluating the effectiveness of the proposed CNN for the needle segmentation task. By quantifying the overlap between the predicted segmentation (output $\hat{y}$) and the actual segmentation (ground-truth y), it provides a measure of how accurately the CNN identifies and delineates regions of interest within an image:

$$Dice = \frac{2\sum_{\mathbf{r}\in\Omega} y_{\mathbf{r}} \cdot \hat{y}_{\mathbf{r}}}{\sum_{\mathbf{r}\in\Omega} y_{\mathbf{r}} + \sum_{\mathbf{r}\in\Omega} \hat{y}_{\mathbf{r}}}. \tag{2.3.2}$$

where $\mathbf{r}$ is the coordinates, y is the ground truth, $\hat{y}$ is the estimate and Ω is the image domain.

2.4 Coarse segmentation with a U-Net

In the following sections, we present the overall method later used in clinical settings to precisely localise the needles delivering electric pulses. The approach is coarse-to-fine: starting by a rough segmentation using deep-learning with a suited loss function and a patch-selection optimisation strategy, and finishing with a Hough transform improved via a voting procedure. A good first approximation of the needle is necessary for the proper evaluation of the needle tip localisation, though the Hough transform is able to adress the issue of potentially missing data in the coarse segmentation, by providing an analytical representation of the object of interest.

The main challenges arise from the nature of the imaging modality—low signal-to-noise ratio—and the nature of the needles—their density causes extensive artefacts on the CBCT. Moreover, the dataset utilised in this study is relatively small, a common limitation in oncology, consisting of only 16 patient samples. Compounding this issue is the significant class imbalance, which presents a formidable challenge for training an effective segmentation model. As highlighted in Table 2.1, the proportion of voxels containing needle information is remarkably low, never exceeding 0.1% of the total volume in the 3D scans. This is due to the extreme thinness of the needles, which occupy a minimal number of voxels compared to the overall image dimensions.

The scarcity of needle-related voxels severely constrains the algorithm's ability to accurately learn the distinct features necessary to segment these objects. The limited representation of needle information within the training data poses a risk

Patient	Needle proportion	Dimension	Spacing (mm)
1	0.007%	512x512x510	0.453
2	0.028%	512x512x474	0.453
3	0.022%	512x512x380	0.362
4	0.010%	512x512x270	0.452
5	0.010%	512x512x504	0.452
6	0.013%	512x512x512	0.452
7	0.015%	512x512x382	0.453
8	0.012%	512x512x510	0.453
9	0.009%	512x512x510	0.453
10	0.032%	512x512x384	0.452
11	0.047%	512x512x194	0.453
12	0.011%	512x512x394	0.453
13	0.013%	512x512x406	0.452
14	0.006%	512x512x509	0.453
15	0.015%	512x512x464	0.452
16	0.003%	512x512x472	0.453

Table 2.1: General characteristics of the data collected on a set of patients whose liver cancer was treated by IRE. The needle proportion is computed as the percentage of voxels including needle information in the whole image. The values are extremely low, indicating a challenging class imbalance within the dataset.

of the model failing to generalise, potentially leading to poor performance and reduced accuracy. If not carefully addressed through techniques such as data augmentation, loss function adjustment, or specialised architectures, this imbalance could significantly hinder the effectiveness of the deep learning models. Thus, mitigating this challenge is crucial to achieving accurate and reliable segmentation outcomes.

2.4.1 Adapted U-Net

As highlighted in section 2.2, the U-Net architecture emerges as one of the most efficient and adaptable deep-learning techniques for segmentation tasks. Recognising its potential, we have decided to customise the U-Net framework to address the unique challenges posed by our specific task, namely, the segmentation of thin objects. Additionally, we will tailor it to accommodate the characteristics of our dataset, which is both small in size and imbalanced. This approach will allow us to leverage the strengths of the U-Net, optimising it for our particular requirements

and improving segmentation accuracy.

We retain some of the architectural and pipeline choices from the original article [78].

Encoder-decoder architecture: The architecture includes a contracting path, reducing the spatial dimensions while preserving essential features, to effectively extract and capture contextual information from the input data. This is complemented by an expanding path designed to achieve precise localisation of critical features, by recovering the spatial information. This dual-path approach enhances the model's ability to understand the broader context while maintaining fine-grained accuracy in feature detection.

Skip connections: At each level of the U-Net architecture, we incorporate connections between the encoder and decoder paths. The feature maps from the encoder are concatenated with the feature maps from the decoder before being processed as a single input. These connections facilitate a deeper network structure while preserving the seamless flow of information throughout the model. The shorter connections between layers close to the input and those near the output enable the integration of various levels of perception (low level edges and high level shapes for instance) [45], similarly to the human visual system, thus enhancing the network's ability to capture both high-level context and fine-grained details simultaneously.

Max pooling layer: Within the encoder, the down-sampling process employs max pooling, a technique that condenses the input by selecting and preserving the maximum values in each division. This operation not only reduces the spatial dimensions of the data but also retains the most prominent features, ensuring that critical information is efficiently carried forward to subsequent layers for more advanced analysis and processing.

Large number of feature channels: Leveraging multiple channels enables the extraction of a richer and more diverse set of information from a single input. This multifaceted approach ensures that various aspects of the input data are captured, enhancing the depth and breadth of the analysis and yielding more comprehensive insights.

Patch-based approach: The inputs are divided into patches to artificially augment the volume of training data. This strategy not only bolsters the dataset but also ensures the preservation of high resolution within the constraints of time and memory. Indeed, larger inputs require more network layers to effectively

down-sample, increasing computational costs. It thus answers the clinical constraints in the case of IRE ablation. When inputs are excessively large, additional network layers are necessary to sufficiently down-sample the data during encoding. Conversely, if the input size is too small, it encompasses insufficient contextual information for thorough analysis, thereby impeding the model's performance. By using 64x64x64 patches, we balance data volume, memory efficiency, and resolution quality, ensuring better model accuracy while addressing clinical challenges.

To address the unique challenges inherent to our dataset and the specialised clinical environments for which our solution is tailored, we put forth the following comprehensive modifications. These adjustments are crafted to enhance the efficacy and precision of our model, ensuring it meets the stringent demands of medical applications.

3D convolutions: Given that both the data and thus the extracted patches are three-dimensional, the application of 3D kernels becomes essential for convolution operations. This approach enables the CNN to effectively capture spatial relationships in three dimensions, leveraging volumetric information from the data and patches. By using 3D kernels, the network can analyse not only the traditional 2D features across height and width but also depth, which is crucial for tasks involving volumetric data such as medical imaging. This approach enhances the model's ability to discern intricate patterns and structures within the 3D data, leading to more accurate and robust predictions.

Asymmetric encoder-decoder: In enhancing our network architecture, we augment the number of feature channels within the encoder phase, thereby introducing an asymmetry. This adjustment enables our model to effectively capture intricate local details, crucial for segmenting objects of exceptionally fine structures, such as the electrodes. By doubling the feature channels, our network enhances its capacity to discern subtle variations and nuances in the data, particularly beneficial for tasks where precise localization is paramount. This approach aligns with recent advancements in deep learning architectures tailored for medical image segmentation, emphasising the importance of comprehensive feature extraction to achieve accurate results [69].

Padded convolutions: In our approach, we integrate zero-padding into the convolution process to uphold consistent feature map dimensions across our customized U-Net architecture. By padding the feature maps with zeros, we maintain the spatial dimensions throughout the network, making it easier to concatenate the feature maps from the encoder with those from the decoder. This approach preserves spatial fidelity and facilitates smoother integration of features, allowing

for more efficient and accurate segmentation without the distortion that might arise from cropping.

Up-sampling: In our segmentation model's decoder phase, we opt for interpolative resizing over traditional deconvolution, as used in the original paper, to enhance computational efficiency without compromising performance. By employing interpolation techniques, we efficiently scale up the feature maps to match the original input dimensions. Interpolative resizing uses simpler operations like bilinear interpolation to scale the feature maps, which reduces the computational cost and memory usage compared to deconvolution. Deconvolution, or transposed convolution, can introduce higher computational demands due to its more complex calculations and the potential for artifacts, such as checkerboard patterns. By opting for interpolation, we can upsample the feature maps more efficiently, accelerating the segmentation process. This method retains the accuracy and detailed localisation required for tasks like medical image analysis, without the computational burden associated with deconvolution.

2-level decoder: The size of our patches being considerably smaller than those used in the original U-Net ($64 \times 64 \times 64$ vs 572×572), only 2 levels are used to avoid over-fitting. Over-fitting occurs when a model becomes overly complex, learning not just the underlying patterns in the training data, but also noise and specific details that do not generalise well to unseen data. By limiting the number of levels to just two, we simplify the model. This helps the network focus on more general features rather than memorising the training data, improving its ability to generalise to new data. Additionally, reducing the complexity of the model can lead to more efficient training and faster convergence

No batch training: In cases where the dataset is small, such as having only 16 patients, the model may face issues during training when using standard batch normalisation, which relies on calculating statistics (mean and variance) across a batch of data. With such a small batch size, these statistics might be unreliable, which can negatively affect the performance and generalisation of the model. To overcome this challeng, layer normalisation is used instead. Unlike batch normalisation, which normalises across the batch dimension, layer normalisation normalises across the features of each individual sample. This helps stabilise training and improve the model's ability to learn, especially when there is a limited amount of data. By considering the CBCTs individually, the network applies normalisation at the level of each scan rather than the entire batch, improving training efficiency and model performance with small datasets like the one described.

Loss function: The choice of the appropriate loss function plays a pivotal role in guiding a model towards effectively learning the task at hand, especially in the realm of image segmentation. Within segmentation tasks, there exist four categories of loss functions, each tailored to address specific challenges and nuances [50]:

- distribution based: measures the difference between two probability distributions. Specifically, it quantifies the dissimilarity between the predicted probability distribution generated by a model and the true distribution derived from the actual data labels. For instance, the binary cross-entropy is computed as:

$$L_{BCE} = -\frac{1}{|\Omega|}\sum_{\mathbf{r}\in\Omega}(y_{\mathbf{r}}log(\hat{y}_{\mathbf{r}}) + (1-y_{\mathbf{r}})log(1-\hat{y}_{\mathbf{r}})), \tag{2.4.1}$$

  where Ω is the image domain, $\mathbf{r}$ is the voxel coordinate, y is the ground-truth segmentation and $\hat{y}$ is the estimated segmentation.

- region-based: exemplified by the Dice coefficient (Eq. 2.4.2), delves into the intersections among distinct regions within a segmentation task. This metric, often referred to as the overlapping index, evaluates the overlap between true positive, true negative, false positive, and false negative regions. It can be adapted as a loss function to minimise:

$$L_{Dice} = 1 - Dice(y,\hat{y}) = 1 - \frac{2\sum_{\mathbf{r}\in\Omega} y_{\mathbf{r}}\cdot\hat{y}_{\mathbf{r}} + \epsilon}{\sum_{\mathbf{r}\in\Omega} y_{\mathbf{r}} + \sum_{\mathbf{r}\in\Omega}\hat{y}_{\mathbf{r}} + \epsilon}. \tag{2.4.2}$$

  where $\epsilon = 10^{-3}$, in our implementation, is a smoothing factor.

  Despite its widespread adoption due to its effectiveness in handling class imbalances and boundary misalignments, the Dice coefficient is non-convex. This characteristic can occasionally hinder its ability to achieve globally optimal results, as it may converge to local minima rather than the optimal solution,

- boundary based: emphasises the fidelity of predicted boundaries to their true counterparts in segmentation tasks. This kind of metric quantifies the maximum distance between points of the estimated and ground-truth boundaries. By prioritising the accurate localisation of boundaries, the Hausdorff distance, for instance, enhances the precision of segmentation models, ensuring robust performance in scenarios demanding high spatial accuracy:

$$D = max_{\hat{y}\in\hat{Y}} min_{y\in Y} ||\hat{y} - y||_2, \tag{2.4.3}$$

- compounded loss: integrates diverse loss components to enhance model performance by leveraging complementary strengths. For example, the combo loss (Eq. 2.4.4) intertwines the Dice loss, emphasising spatial overlap accuracy, with cross-entropy, which penalises probabilistic deviations between predicted and actual distributions. This fusion addresses nuances across segmentation tasks, balancing pixel-wise precision with holistic structural integrity. By synergising distinct loss metrics, compounded approaches navigate complex data distributions more adeptly, fostering robust learning frameworks crucial in fields like medical imaging and autonomous systems.

There is no universal loss function. The choice needs to adapt to the data set (class imbalance, skewness, etc) and the task at hand (impact of false positives/negatives on the solution quality, etc). The loss function we propose to use is composed of the dice loss function and a binary cross-entropy instead of a pixel-wise softmax with weighted cross-entropy:

$$L = L_{BCE} + L_{Dice} \tag{2.4.4}$$

The class imbalance is further dealt with in the learning and inference framework themselves.

A variant, thereafter refered to as 3Loss-U-Net, proposes to add a term corresponding to the Tversky loss to further adapt to the class imbalance thanks to an extra parameter β:

$$L_{Tversky} = 1 - \frac{1 + \sum_{\mathbf{r}\in\Omega} y_{\mathbf{r}} \cdot \hat{y}_{\mathbf{r}}}{1 + \sum_{\mathbf{r}\in\Omega}[y_{\mathbf{r}} \cdot \hat{y}_{\mathbf{r}} + \beta(1 - y_{\mathbf{r}})\hat{y}_{\mathbf{r}} + (1 - \beta)y_{\mathbf{r}}(1 - \hat{y}_{\mathbf{r}})]} \tag{2.4.5}$$

where $(1 - y_{\mathbf{r}})\hat{y}_{\mathbf{r}}$ are the false positives and $y_{\mathbf{r}}(1 - \hat{y}_{\mathbf{r}})$ are the false negatives.

ADAptive Moment (ADAM) optimiser: We employ the ADAM optimiser, a powerful algorithm in machine learning that effectively navigates complex optimisation landscapes, accelerating convergence towards optimal solutions while mitigating the risk of getting stuck in local minima. It is a combination of gradient descent with momentum and Root Mean Square Propagation (RMSP). Parameters are updated using multiple iterations of gradients, as follows:

$$w_{t+1} = w_t - m_t^*(\frac{l_r}{\sqrt{v_t^*} + \epsilon}), \tag{2.4.6}$$

with the bias corrected momentum:

$$m_t^* = \frac{m_t}{1 - \beta_1}, \tag{2.4.7}$$

the bias corrected velocity:

$$v_t^* = \frac{v_t}{1-\beta_2}, \tag{2.4.8}$$

the moment:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\frac{\delta L}{\delta w_t}, \tag{2.4.9}$$

the velocity:

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\left(\frac{\delta L}{\delta w_t}\right)^2, \tag{2.4.10}$$

where l_r is the learning rate, t the iteration number, ϵ a small positive constant, β_1 and β_2 the decay rates of the moment and velocity respectively.

The architecture resulting from the afore-mentioned modifications is represented in Figure 2.11.

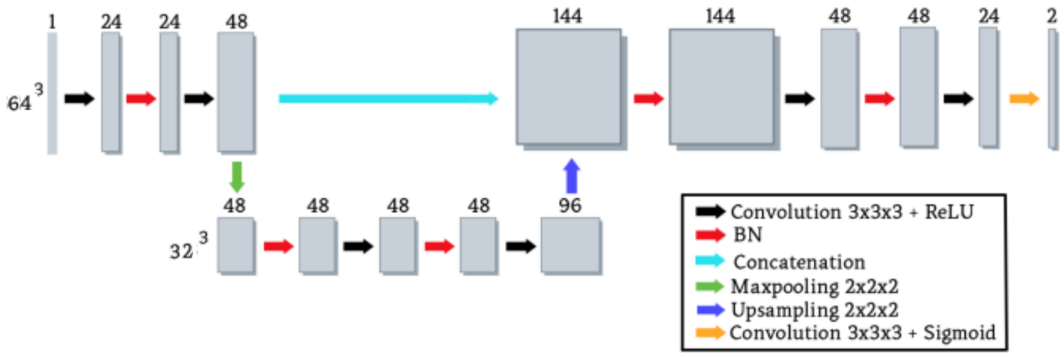


Figure 2.11: The U-Net architecture has been customised to suit the specific requirements of our dataset. This modified version incorporates two distinct levels of resolution, up-sampling techniques to increase feature resolution, and twice as many feature maps in the decoder compared to those in the encoder. These adjustments allow the network to better capture and represent intricate details pertinent to our data.

2.4.2 Post-processing optimisation

The over-representation of background compared to needles in the dataset facilitates the classification of background voxels. Therefore, the network recognises background voxels with more confidence than it does needle voxels. This translates

to higher probabilities for the background channel of the output. Thus, when computing the binary segmentation masks, it is useful to adjust the inference threshold θ in order to take into account the increased confidence. Indeed, reducing it allows to artificially give more weight to the needle prediction despite the skewness in the dataset since the conventional threshold $\theta = 0.5$ proves suboptimal given the dataset characteristics.

The inference is therefore finely tuned to align perfectly with the training dataset, establishing a robust foundation for the analysis. This calibrated value will remain steadfast for the remainder of the thesis, ensuring consistency and reliability in our findings.

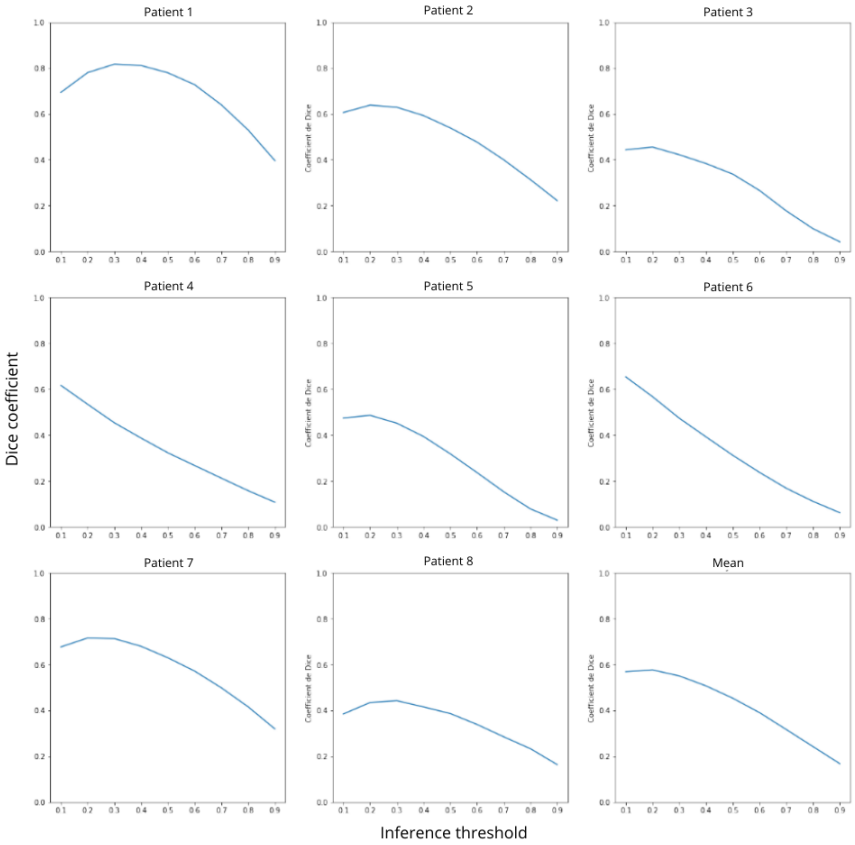


Figure 2.12: Dice coefficients against inference thresholds for all 8 patients from the training dataset, and their average. In average, a lower inference threshold of 0.2 allows to counter balance the under-representation of needles.

Low quality image: In CBCT presenting with weak signal and low contrast (patient 4 and 6), the network struggles to detect the needles. The relationship

between inference threshold and Dice coefficient is linear (Fig. 2.12) , that is the more we allow positives, whether true or false, the more the network predicts needles. The linear relationship highlights how cautious adjustments in thresholding can influence the network's propensity to detect needles amidst low contrast scenarios.

Satisfying quality image: In CBCT with higher signal-to-noise ratio and contrast, the neural network demonstrates enhanced capability in detecting needles. This results in a nuanced relationship between the inference threshold and Dice coefficient, characterised by a parabolic curve. As the threshold adjusts, the network's confidence in predicting needle voxels fluctuates, reflecting its ability to discern subtle features amidst clearer imaging conditions.

Overall, the performance of the network exhibits notable enhancement when utilising a lower threshold setting. This adjustment allows the model to capture finer details and nuances in the data, thereby improving its accuracy and sensitivity in identifying relevant features. In the subsequent sections of this chapter, we will adopt an inference threshold of 0.2 to optimise the network's performance and ensure robust results, as it has proven to lead to higher segmentation quality on the training dataset (Fig. 2.12).

2.4.3 Patch overlap

This section evaluates how the use of overlapping patches influences the network's effectiveness in image segmentation tasks. Overlapping patches can enhance segmentation quality by providing redundant coverage of image regions, improving the network's ability to capture intricate details and boundaries. Moreover, they contribute to robustness against variations and noise in input data, leading to more reliable segmentation outcomes.

However, this approach also impacts computational efficiency. The time required for processing increases as the overlap between patches grows, affecting real-time application feasibility. Balancing the trade-off between segmentation accuracy and computational cost is crucial for optimising network performance in practical scenarios.

We test 3 degrees of overlap: no overlap, overlap of half the patch size, overlap of 3 quarters of the patch size.

As can be seen in Table 2.2, overlapping patches barely improves compared to no overlapping when the overlap is half the patch size. In fact, performing a paired t-test shows that this increase in segmentation quality is not statistically significant ($p = 0.49$).

For the stronger overlapping tested, the performance deteriorates. The high

Patient	No overlap	Overlap = 1/2	Overlap = 3/4
9	0.7401	0.5176	0.4957
10	0.5198	0.5364	0.5073
11	0.3661	0.3657	0.3530
12	0.3094	0.5184	0.4945
13	0.3052	0.3779	0.3432
14	0.3721	0.3782	0.4801
15	0.6203	0.5679	0.5510
16	0.3928	0.3713	0.3002
Mean ± std	0.4532 ± 0.1581	0.4542 ± 0.0879	0.4406 ± 0.0934

Table 2.2: Dice coefficients for different degrees of overlapping patches. The increased redundancy does not improve the segmentation quality significantly.

Overlap	Training time (s)
0	3103
1/2	10446
3/4	48002

Table 2.3: Training time averaged over 5 training for different degrees of overlapping patches. The inference time, and thus the training time, notably increases.

redundancy hinders the CNN's ability to generalise to the test data. Furthermore, increasing the overlap results in increasing the computation time (Tab. 2.3). Therefore, we choose to discard the overlapping patch strategy proposed in [78] for the remainder of this thesis.

2.4.4 Investigating a new loss

Our exploration into a novel loss function integrating binary cross-entropy (Eq. 2.4.1), Dice loss (Eq. 2.4.2), and Tversky loss (Eq. 2.4.5) signifies a strategic approach to enhance model performance in complex tasks like image segmentation. By leveraging the strengths of each component, we aim to achieve robustness against class imbalance inherent to the segmentation task at hand. The binary cross-entropy component provides a fundamental measure of dissimilarity between predicted and ground-truth distributions, while the Dice loss emphasises overlap between predicted and ground-truth masks, particularly effective for unbalanced data. Additionally, the Tversky loss offers flexibility in adjusting model behavior

by fine-tuning the balance between false positives and false negatives.

β	Mean Dice coefficient ± std
0	0.3161 ± 0.0080
0.1	0 ± 0
0.2	0.3082 ± 0.0080
0.3	0.3678 ± 0.0659
0.4	0 ± 0
0.5	0.3342 ± 0.0310
0.6	0 ± 0
0.7	0.3067 ± 0.0081
0.8	0.3082 ± 0.0080
0.9	0 ± 0

Table 2.4: Mean Dice coefficients for different values of the 3 term loss parameter β. The addition of a Tversky term in the loss does not lead to improved segmentation quality.

The instability of the 3-term loss function in relation to parameter β is evident in our findings, where significant fluctuations in mean Dice coefficients were noted across different values of β sampled during experimentation (see Tab. 2.4). This sensitivity highlights the challenge of maintaining consistency and predictability in model performance when employing such loss functions. The variability observed underscores the need for more robust formulations or adaptive strategies to mitigate the impact of parameter changes on model outcomes, especially in tasks like image segmentation.

Furthermore, the optimal value $\beta = 0.3$ does not lead to better results than the combo loss (Dice= 0.4532). The decision to prioritise the combo loss henceforward is grounded in its demonstrated capability to consistently outperform $\beta = 0.3$ across our experiments. This strategic shift ensures that our focus remains on optimising model performance and achieving robust outcomes in U-Net training.

2.4.5 Patch-selective learning strategy

Addressing the representation imbalance between the needles and the background poses a significant challenge in our current task. To solve this problem, we propose a patch selection strategy based on the information included within the considered region of the image.

To comprehensively investigate the impact of needle presence on network performance, we curated distinct training subsets from our dataset. Each subset was

crafted to analyse how varying proportions of needle-inclusive patches influence network efficacy. Needle patches were defined as regions where the ground-truth segmentation masks confirmed the presence of at least one voxel belonging to a needle. This strategic distribution not only facilitated precise training but also enabled nuanced evaluation of the network's responsiveness to different concentrations of needle-related information. By systematically varying these subsets, we gained valuable insights into optimising the network's ability to discern and accurately process critical needle data amidst complex imaging scenarios. The proportions tested are 100%, 67%, 50% and 7% (corresponding to all available patches) of needle patches. To complete the patch sets with background patches, a CBCT scan from our training dataset was selected at random. Subsequently, patches were extracted from this chosen CBCT scan. This systematic extraction process was pivotal in ensuring comprehensive coverage of all abdominal structures within our training data, including vital organs such as the liver and intricate skeletal elements like the ribs. By adopting this approach, we aimed to optimise the neural network's ability to discern and accurately interpret various anatomical features present in complex abdominal imaging scenarios.

We anticipate achieving an equilibrium between the representation of needle information and background details within one of these datasets. Despite our methodological efforts, the network may still exhibit a stronger propensity for confidently predicting background elements, due to the predominant representation of background information compared to needles, which motivates the need for the other proposed adaptations, such as the inference threshold for instance.

Initially, a specialised subset of patches was curated, focusing exclusively on those containing crucial needle information. 38 distinct patches were identified within the comprehensive training dataset.

We then added 23 supplementary background patches into the initial patch set, deliberately sourced from a single patient, chosen through a random selection process. By including diverse background samples from the same patient, the dataset becomes more robust for training, facilitating the model's ability to generalise across different anatomical features

The third patch set integrates the entirety of the second patch set, with the last 35 patches sourced from the same patient, selected at random.

Finally, the last patch set gathers all the patches extracted from the 8 patients in our training database.

We seem to obtain an equilibrium for class representation around 50% of needle patches as seen in Table 2.5. This equilibrium not only optimises our model's performance timewise but also enhances the Dice coefficient significantly, all while maintaining a practical training duration. Subsequently, this curated dataset earns the designation of the "balanced set".

Training set	Mean Dice coefficient	Time/epoch (s)
Patch set 1 (100-0%)	0.1310	847
Patch set 2 (67-33%)	0.5179	1701
Patch set 3 (50-50%)	0.5754	2698
Patch set 4 (7-93%)	0.4532	6749

Table 2.5: Mean Dice coefficients following the training of the proposed architecture on patch sets including different proportions of needle information. As a balance is found between needle and background information, the segmentation quality increases.

2.4.6 Comparison with the nn-U-net

We detail a performance comparison between our adapted U-net model and the state-of-the-art nn-U-net algorithm [48] presented in the literature review (Section 2.2.2).

As a brief remainder, nn-U-net is an algorithm encompassing a pre-processing pipeline, a segmentation U-net whose hyper-parameters are optimised with regards to the dataset, and a post-processing pipeline. Known for its robust and self-adapting framework, it automatically configures itself to various datasets, providing high adaptability and strong baseline performance across numerous segmentation tasks. The authors propose 3 networks: 2D, 3D and cascade, the closest to our network being 2D and 3D.

By juxtaposing these models, our results highlight the advancements in segmentation technology brought by our enhancements to the U-net architecture, providing a compelling alternative to the already impressive nn-U-net.

Our proposed method consistently outperforms the other networks on average due to its precise alignment with the specific requirements of the task (Tab. 2.6). This tailoring ensures not only superior performance but also enhanced accuracy. The method exhibits a lower standard deviation, indicating a more reliable and consistent outcome. This reduced variability signifies that our approach provides stable and dependable results, making it an optimal choice for the complex segmentation of thin electrode segmentation.

The 3D version of the nnU-Net produces results comparable to those achieved by our U-Net before implementing the patch selection strategy, with a p-value of 0.48 indicating no statistically significant difference. This observation underscores a critical insight: the strategic selection of training data significantly enhances the efficiency of the learning process ($p = 0.04$). By meticulously curating the training dataset, our U-Net can more effectively focus on relevant features, thereby optimis-

Patient	Balanced Patch Set	2D nn-U-net	3D nn-U-net
9	0.7569	0.6396	0.6775
10	0.6487	0.7993	0.7507
11	0.4308	0.1948	0.3483
12	0.5419	0.4281	0.4381
13	0.4109	0.1086	0.4012
14	0.6529	0.1422	0.3936
15	0.7019	0.5780	0.5953
16	0.4590	0.0003	0.0003
Mean ± std	0.5754 ± 0.1326	0.2905 ± 0.2904	0.4506 ± 0.2335

Table 2.6: Dice coefficients for the proposed method against two versions of the nn-U-net [48], namely the 2D and the 3D versions. In average, the proposed approach outperforms the established nn-U-net.

ing performance and accuracy. This approach not only validates the importance of data selection in training neural networks but also highlights the potential for achieving superior results through targeted data curation.

Conversely, the 2D version demonstrates suboptimal performance. The implicit data augmentation achieved by providing slices rather than the entire 3D image fails to compensate for the deficiency of 3D contextual information. This limitation is particularly pronounced given the object's thin and longitudinal shape, which requires comprehensive 3D data for accurate segmentation. Consequently, the 2D approach struggles to capture the intricate spatial relationships and fine details necessary for precise segmentation, underscoring the importance of utilising full 3D imagery for such tasks.

2.5 Fine localisation with a Hough transform

The segmentation approach using deep learning presents inherent challenges in achieving the precision necessary for accurate electric field estimation. Firstly, due to their density, the electrodes often appear thicker in CBCT scans, complicating their precise localisation. Secondly, the segmentation of individual needles lacks continuity guarantees, further impacting the reliability of the segmentation results. These factors underscore the complexities involved in leveraging deep learning for achieving high-precision segmentation required in medical imaging and electric field simulations.

Therefore, the segmentation masks undergo additional processing to precisely

pinpoint the localisation of the needle tip, addressing any gaps or inaccuracies in the data. This refinement is crucial as it focuses solely on identifying the tip of the needle, which is pivotal for delivering electric pulses in IRE procedures. By isolating and accurately determining the needle tip, the method ensures optimal placement and efficacy of therapeutic treatments, enhancing patient outcomes and procedural success.

2.5.1 Hough transform improved with a voting procedure

To precisely localise the needles, we make use of the Hough transform to obtain the most represented needles, parametrised as lines, in the segmentation masks. The Hough transform is an algorithm for shape recognition that provides an analytic representation of the object of interest. Given a shape parametrisation and a point cloud, it consists in exhaustively searching the parameter space for any shape present in the image. Doing so, it fills an accumulator array, whose dimensions represent each parameter of the shape to detect. The values in the array correspond to the number of points belonging to a given shape. Local maxima then indicate the most represented shapes in the image.

Even though some bending may occur as the electrode is inserted, a line is still a good approximation for the needle tip. We choose the Roberts' optimal line representation in the form of a 4-tuple (b_x, b_y, x', y') where $\mathbf{b} = (b_x, b_y, b_z)$ is the unit vector normal to the plane containing the origin and the line to parametrise, and x', y' are the coordinates of the intersection between the line to parametrise and the plane previously defined, in a local 2D Cartesian frame, as represented in Figure 2.13. 4 is the minimum number of parameters required to uniquely define a line in 3D. Moreover, the representation has the advantages of presenting no singularities or special case and proposes an exact one to one correspondence with respect to the point and orientation representation [77]:

$$\mathbf{b} = (b_x, b_y, \sqrt{1 - b_x^2 - b_y^2}) \tag{2.5.1}$$

and:

$$\mathbf{p} = x'(1 - \frac{b_x^2}{1 + b_z}, -\frac{b_x b_y}{1 + b_z}, -b_z) + y'(-\frac{b_x b_y}{1 + b_z}, 1 - \frac{b_y^2}{1 + b_z}, -b_z), \tag{2.5.2}$$

with $x', y' \in [-\infty, \infty]$ and $0 \leq b_x^2 + b_y^2 \leq 1$.

To improve the time efficiency of the algorithm, the orientation space, i.e. for the parameters b_x and b_y, is discretised based on a tessellation of a platonic solid. The faces of the icosahedron are divided into four triangles as many times as necessary to successfully cover the orientation space. Additionally, to further limit the parameter space, the centre of the point cloud is defined as the origin [18].

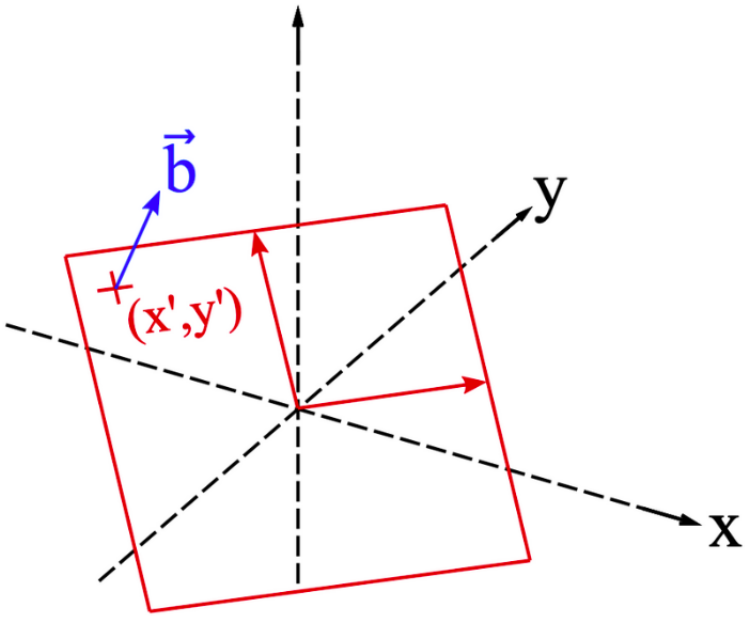


Figure 2.13: The diagram illustrates the relationship between the 4-tuple from Roberts' optimal line representation [17] and the corresponding line, depicted in blue. Additionally, it shows the constructed plane, marked in red, within a designated coordinate system, reproduced with permission.

Then, the binary segmentation mask undergoes a transformation into a point cloud representation, a crucial step in preparing the data for the Hough transform algorithm. This process involves converting the 3D binary mask, which outlines the segmented object's boundary, into a 3D point cloud format. By translating the binary mask into a rich, spatially detailed point cloud, the Hough transform can effectively analyse and interpret the geometrical characteristics and spatial relationships within the segmented object.

While filling the accumulator array, we allow some imprecision in the localisation of the points in the point cloud and use the radius of the needles as the maximum distance between the point and the detected line, that is 8 to 9.5 gauges.

The search for the $k \times N$ most represented lines, where k is a given coefficient and N is the number of electrodes, additionaly takes into consideration a minimum score for the line to be considered as detected.

Once the line with the most points is selected, a reverse Hough transform is performed in which every point belonging to the selected line is removed from the accumulator array. This allows to sequentially detect the required number of lines by simply considering the global maximum in the accumulator array. Moreover, the reverse Hough transform makes the process more efficient as each point in the

point cloud is only used twice by the algorithm.

Following this procedure, as there are too many detected needles, we further select the right representation of the electrodes with an elaborate voting procedure. Criteria are:

- the inverse number of points belonging to a line: to favor most represented needles,
- the inverse distance between the extremities of the segment: to favor long objects that are more likely to be indeed needles as opposed to ribs for instance,
- the mean distance from the points to their projection on the detected line: to favor lines where actual points are closed to the representation,
- the distance between the extremity and the center of the image: to favor lines that are close to the tumor and are thus more likely to be the needle tip.

Those scores are normalised across the set of $k \times N$ lines previously selected and their sum is considered as the final score. The N lines with the lowest scores are then used as the final representation of the needles tip.

2.5.2 Overall needle localisation

We are now undertaking a comprehensive evaluation of our integrated framework, which combines coarse segmentation using deep learning with precise localisation via the Hough transform. This dual approach leverages the powerful abstraction capabilities of deep learning for initial segmentation, followed by the classical robustness of the Hough transform to refine and accurately localise features. By synergising these techniques, we aim to achieve a precise analytic representation of the needle tips, capitalising on the strengths of both modern and traditional methods to handle complex image data effectively.

To evaluate our segmentation algorithm, we benchmark it against a previously utilised thresholding-based method. Initially, a high threshold of 500 is applied to the CBCT scans to achieve a coarse approximation of the structures. This step is crucial for isolating the primary features of interest. Subsequently, a lower threshold of 100 is employed to refine the delineation, ensuring that even the finer details of the objects are accurately detected. This dual-threshold approach allows us to balance between capturing the broader structural elements and the intricate nuances of the scanned objects, thereby improving the overall accuracy and reliability of the segmentation process.

To assess the effectiveness of integrating the CNN and Hough transform in localising needles, we quantify the quality of the needle localisation by computing the L2 norm, a standard measure that captures the spatial accuracy between the predicted and expert-marked coordinates of the needle tip:

$$L_2 = \sqrt{(x - \hat{x})^2 + (y - \hat{y})^2 + (z - \hat{z})^2} \tag{2.5.3}$$

where (x, y, z) are the true coordinates and $(\hat{x}, \hat{y}, \hat{z})$ are the algorithm estimates.

Patient	Average distance (mm) with thresholding	Average distance (mm) with proposed U-Net
9	76.98	1.88
10	0.97	1.44
11	1.25	0.72
12	1.11	1.17
13	7.69	5.80
14	4.99	6.07
15	3.66	2.18
16	18.13	1.56
Mean ± std	14.35 ± 25.94	2.60 ± 2.10

Table 2.7: Euclidean distance between needle tip coordinates estimated (either through the thresholding method, or the proposed deep-learning method) and the coordinates as determined by radiologists. In average, the proposed method outperforms the previously used thresholding technique.

The method incorporating deep learning successfully outperforms the previously used thresholding method, as seen in Table 2.7. Indeed, we observe a much lower average distance between the ground-truth and the coordinate estimates. This improvement in the needle localisation will in turn lead to a more accurate estimation of the electric field delivered.

2.6 Feasibility in clinical settings

The proposed algorithm is suitable for clinical application time wise. visualisation of the electric field onto the CBCT takes approximately 2 minutes, which is highly acceptable within a clinical workflow. Moreover, the precise localisation of the needles is achieved in just 10 seconds using standard hardware (an Intel 2.5 GHz i7 workstation (8 cores) with 32 GB of RAM), showcasing its remarkable

speed and practicality. The implementation of GPU acceleration will further enhance segmentation speed, making the process even more efficient. Despite these potential improvements, the current performances are already suitable for clinical use, ensuring timely and accurate needle localisation in real-time scenarios.

In our search of an optimal learning strategy, we experimented with several innovative variants. First, we investigated the impact of varying degrees of overlap in the patches fed to the CNN. This involved assessing how different levels of patch overlap influenced the network's ability to learn and generalise from the training data. Additionally, we evaluated the effectiveness of two distinct loss functions: the combo loss, which combines the Dice coefficient and the binary cross-entropy to improve training stability and performance, and the 3-term loss, designed to take into account the class imbalance. These experiments aimed to fine-tune the model's accuracy and efficiency, ultimately contributing to more precise and reliable outcomes.

The overlapping patch strategy presents mixed results. While it extends the inference duration, it does not significantly enhance the network's performance. This inefficiency arises because only a small number of patches contain relevant information about the object of interest. Consequently, the inherent data augmentation achieved through overlapping patches introduces unnecessary redundancy. This surplus of patches burdens the training process, increasing computational costs without delivering proportional benefits in accuracy or effectiveness.

A similar line of reasoning elucidates why a selective patch strategy yields superior performance compared to indiscriminately feeding all available patches to the network. It is crucial to strike a balance between discerning what constitutes the background and what represents the needle. Given the pronounced class imbalance in the dataset, where background information is more prevalent, the network naturally gravitates towards learning the background more effectively. However, by implementing a patch selection strategy that mitigates this skewness, we enable the network to better focus on the object of interest—the needle—while still maintaining its confidence in identifying the background. This targeted approach ensures that the network's learning process is more robust and nuanced, ultimately enhancing its predictive accuracy and overall performance.

The advancement in needle tip localisation is noteworthy, demonstrating significant improvements over previous methods in both precision and robustness. The new approach provides more accurate and reliable needle tip detection, surpassing the older thresholding-based technique. When applied to new CBCT scans, it consistently succeeds in identifying all needles, whereas the previous method occasionally failed to detect certain electrodes. This progress ensures a higher degree of accuracy and reliability in clinical settings, making the new technique a substantial advancement in needle localisation technology.

2.7 Conclusion

The performance of the proposed U-Net model is significantly influenced by several key factors. Foremost among these is the choice of loss function, which directly affects how well the model learns to distinguish between different features in the data. Additionally, the composition of the training data plays a crucial role; diverse and well-represented data ensures that the model generalises effectively across various scenarios. Lastly, the inference threshold, which dictates the decision boundary for classifying voxels, impacts the precision of the segmentation results. Each of these elements—loss function, training data composition, and inference threshold—interacts to determine the overall efficacy and accuracy of the U-Net model, making them critical considerations in the design and evaluation process.

As highlighted in [49], a more complex framework does not necessarily guarantee improved precision in results. For instance, the use of an overlapping strategy or a complex 3-term loss function—comprising three distinct terms—did not outperform the simpler approach employing a no-overlap scheme with a two-term loss function. This observation underscores that increased complexity in model design does not always translate into enhanced performance. Additionally, while it might seem intuitive that more training data would always lead to better results, this is not the case. Excessive data redundancy can actually impede the training process, making it harder for the model to discern meaningful patterns and leading to diminished performance. Therefore, both the complexity of the model and the quality of data need to be carefully balanced to achieve optimal results.

Incorporating this enhancement into the numerical workflow is anticipated to significantly improve the accuracy of electric field estimates. Our findings demonstrate that this modification refines the precision of needle tip localisation and should therefore enhance the overall quality of the electric field estimation.

In brief

Deep learning, particularly CNNs, has gained significant traction in the field of image analysis, including segmentation tasks. In this work, we address the problem of needle localisation using a coarse-to-fine approach. The initial coarse segmentation is enhanced through deep-learning techniques, which significantly improves the accuracy and robustness of the process.

However, several challenges complicate this task, stemming from the specific nature of the electrodes, the imaging modality, and the size of the dataset. To overcome these hurdles, we adapted a U-Net architecture and designed a tailored learning strategy. This strategy includes the selection of an appropriate loss function, as well as an effective patch selection process for training. After experimenting with several variants, the most successful model is a 3D U-Net combined with a combo loss function, trained on non-overlapping patches.

Once segmentation is complete, we further refine the results by adjusting the inference threshold. This step helps mitigate the imbalance between the needle class and the background, which is a common issue in medical image segmentation. Ultimately, this approach, combined with the fine localisation by the Hough transform, results in a more accurate and stable needle localisation compared to the previously employed thresholding algorithms, offering significant improvements in precision and consistency. The analytic representation can then be directly incorporated in the electric field simulation.

Chapter 3

Boundary conditions and image registration

This chapter delves into the intricate challenge of boundary conditions in image registration, a topic often overlooked despite its profound impact on the accuracy of registered images. Mis-registration stemming from boundary conditions can lead to significant distortions, yet this critical aspect remains under-explored in the existing literature. Typically, algorithms resort to either homogeneous Dirichlet boundary conditions for wide fields of view or homogeneous Neumann boundary conditions. However, neither approach sufficiently addresses the complexities of most scenarios, particularly in partial fields of view commonly encountered in clinical practice, where the region of interest is often close to the borders. It is, for instance, the case in IRE procedure, where multiple partial field of view CBCT are taken to successfully locate the tumor and the electrodes, which are often near the boundary. By examining these limitations, this chapter seeks to illuminate the pressing need for more nuanced boundary condition strategies that enhance the fidelity of image registration in medical applications.

In this work, we introduce a novel approach involving local boundary conditions that dynamically adjust to the unique characteristics of the images being registered. Specifically, we employ a Robin-type boundary condition, seamlessly integrating flow field information derived from an initial transformation estimate. To streamline the optimisation process, we reduce the set of hyper-parameters to just two critical hyper-parameters, which are fine-tuned through a comprehensive grid-search aimed at minimising the registration energy. This innovative methodology has demonstrated remarkable efficacy, significantly enhancing registration quality when benchmarked against ground truth in both mono-modal and multi-modal tasks. Our findings underscore the potential of this adaptive strategy to transform the landscape of image registration, delivering more accurate and reliable results in diverse applications.

The proposed boundary conditions consistently surpass the performance of widely used alternatives and validate their efficacy in specific contexts. For instance, in the realm of abdominal cavity imaging, where minor movements occur across all boundaries, our approach validates that the homogeneous Neumann boundary conditions effectively model these subtle shifts, illustrating that while our method excels in general, it also affirms the relevance of traditional approaches when applied to particular situations. This nuanced understanding highlights the versatility and adaptability of our boundary condition framework, paving the way for enhanced accuracy in diverse imaging applications.

Finally, we show that this enhancement in registration quality can significantly impact the accuracy of electric field estimates in the context of IRE ablation. This is particularly pertinent given that partial fields of view are utilised in the operating room to minimise radiation exposure to patients. By ensuring a precise transformation that aligns the tumor and the electrodes within a unified frame of reference, our approach directly influences the accuracy of electric field estimations and the overall assessment of the ablation procedure. This synergy between improved registration and electric field modeling will not only optimise clinical outcomes but also contribute to safer, more effective treatment protocols.

Contents

The method detailed in this chapter was published in the peer-reviewed journal Physics in Medicine & Biology:

- Eloise Inacio, Luc Lafitte, Laurent Facq, Clair Poignard and Baudouin Denis de Senneville. *Adaptive local boundary conditions to improve Deformable Image Registration*, Physics in Medicine & Biology, Volume 69, Number 16, August 2024 (10.1088/1361-6560/ad6952),

and orally presented at the conference Journées Maths Bio Santé 2023.

3.1 Role of the registration task

During the IRE ablation procedure, two CBCT scans are captured to ensure comprehensive imaging and precise needle placement. The first CBCT scan is performed prior to needle insertion, providing a clear and detailed view of the tumor and adjacent anatomical structures. This preliminary scan is crucial for assessing the spatial relationship between the tumor and surrounding tissues, facilitating accurate targeting. The second CBCT scan is conducted after the needle insertion, where the localisation algorithm, as discussed in the preceding chapter, is employed to pinpoint the needles with high precision. This post-insertion scan enables visualisation of the needle placement, ensuring optimal alignment and enhancing the accuracy of the ablation process. However, it often suffers from artifacts caused by the needle's density. These artifacts can significantly degrade image contrast, rendering the tumor less visible or even obscured. A CBCT after needle insertion is shown in Figure 3.1, demonstrating the extent of the artifacts. Thus, to overcome this challenge, two scans are captured, later registered to align the relevant structures.

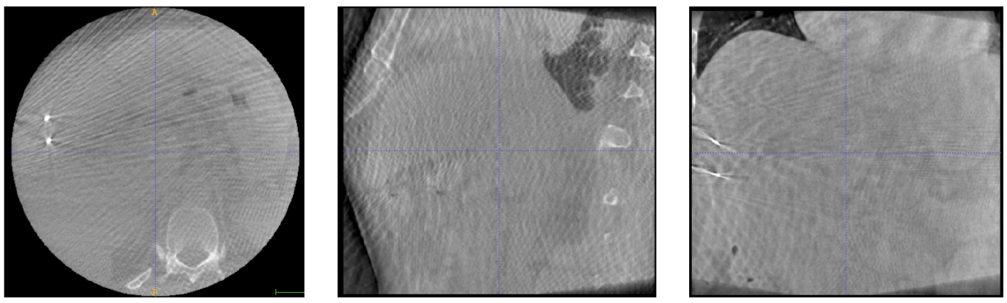


Figure 3.1: CBCT slices from our patient database, after needle insertion: (a) displays the axial plane, (b) the coronal plane, and (c) the sagittal plane. Artifacts include the dilation of the needles (in fact, they should not be more than 2 to 3 voxels wide) and streaking artifacts. Additionally, needles can be located close to the borders, leading to tissue compression that the registration must take into account.

Our primary objective is to provide interventional radiologists with critical insights into whether the tumor, along with the necessary treatment margin, falls within the effective treatment area. Achieving this necessitates a seamless integration of tumor location data from the initial CBCT scan with the estimated treatment area computed on the subsequent CBCT scan. To facilitate this, we implement a registration process that aligns the first CBCT with the second. Specif-

ically, this thesis introduces innovative boundary conditions designed to enhance the performance of various registration algorithms in a straightforward manner, adapting intelligently to the unique characteristics of the images involved. Furthermore, we demonstrate that optimising these boundary conditions via registration energy correlates directly with improved registration quality, as evidenced by comparisons with available ground truth data. This comprehensive approach not only bolsters the accuracy of tumor localisation with respect to the electrodes but also supports more effective treatment evaluation in interventional radiology.

3.1.1 Registration in the medical field

Image registration is a pivotal computer vision task that involves determining the optimal transformation to align a moving image with a reference image. Initially developed to address challenges in remote sensing and medical imaging, registration has evolved significantly, becoming a cornerstone in numerous disciplines. More specifically, the registration task is crucial in the medical field as it enables precise comparisons of anatomical images captured across various modalities and time points. Aligning medical images allows healthcare professionals to accurately track the progression of a patient's condition. This is essential for assessing the effectiveness of treatments by comparing pre- and post-intervention images. By superimposing images from different sources, such as CT scans, MRIs, or CBCT, practitioners can correct for potential discrepancies caused by variations in patient positioning, movement, or changes in imaging equipment. This ensures that any observed changes are attributable to the condition or treatment, not technical factors. The ability to visualise these aligned images in one unified view helps improve decision-making, enhances diagnosis, and facilitates personalised treatment planning, leading to more accurate and reliable patient care outcomes. Additionally, image registration can be a component of more complex workflows in medical imaging, where it facilitates the integration of information from multiple sources. For instance, in image data fusion, different images are aligned and combined to provide a comprehensive view of the patient's anatomy. In atlas-based segmentation, an image is registered to a predefined atlas that contains the region of interest's segmentation, allowing for accurate delineation of structures. The segmentation mask from the atlas is then transformed using the motion estimate to match the target image, enabling precise identification of areas requiring attention. Volumetric image reconstruction uses data from multiple 2D slices to create a 3D volume, helping visualise complex anatomical structures. Additionally, motion tracking allows the tracking of organ or tissue movement over time, which is crucial for accurate treatment planning in radiology or surgery. Finally, tissue elasticity estimation is facilitated by registering elastography images, where the deformation of tissue under mechanical stress provides insights into its stiffness, aiding in the

diagnosis of conditions like liver fibrosis or tumors.

In treatments requiring dose computation, such as calculating the electric field in IRE procedures, accurate assessment is crucial for ensuring effective and safe therapy. Image registration plays a pivotal role in this process by aligning and fusing information from various medical images taken at different times or modalities. For instance, this technique allows clinicians to more accurately evaluate the electric field distribution and tissue response, leading to improved precision in targeting abnormal cells while preserving healthy tissue. By integrating data from multiple sources, image registration enhances the evaluation and monitoring of the treatment's efficacy and safety. IRE represents just one example among various medical contexts where accurate dose calculations—such as the distribution of electric fields—are crucial. Other methods like radiotherapy [66] [57] [15] [97] or thermo-therapy [106] [9] [58] also require this kind of information for an optimal outcome.

In the realm of image-guided therapy, registration is essential for maintaining precision. Biological movements, such as breathing or gastrointestinal peristalsis, can cause shifts in the position of the targeted area, which may impact the accuracy of treatment delivery. Image registration compensates for these dynamic motions by aligning images taken over time. This ensures not only the effectiveness of the treatment but also minimises harm to surrounding healthy tissues, making it a critical tool in therapies that rely on high accuracy.

While our work focuses on a specific application, it is crucial to recognise that image registration is a foundational technique in medical imaging, playing a vital role across numerous domains.

Rigid registration: Rigid registration involves transformations that preserve the distances between any two points in an image, such as rotations and translations. This method is the most straightforward form of image registration, focusing on aligning images through basic geometric adjustments without altering the shape of the objects within them. Despite its appealing simplicity, rigid registration is often inadequate for capturing the complex movements observed in clinical contexts, where biological tissues might undergo significant deformations due to factors like patient motion or internal changes.

Nevertheless, rigid registration serves as a valuable preliminary step in many imaging workflows. By first applying rigid transformations to align images roughly, it can simplify and enhance the subsequent application of more sophisticated, deformable registration techniques. This preparatory step helps in achieving a more accurate and detailed alignment by reducing the complexity of the problem for advanced algorithms that handle non-rigid deformations.

Deformable Image Registration (DIR): DIR offers a greater range of flexibility compared to rigid registration, accommodating various transformations such as shearing, compression, and scaling. Unlike rigid registration, DIR can model complex anatomical changes and patient movements more accurately. This method is particularly beneficial for handling dynamic or non-rigid deformations, such as those occurring during needle insertion, where tissues may be compressed, or due to diaphragm motion during the breathing cycle, which can cause significant extension or compression of anatomical structures.

While DIR tends to be more computationally demanding due to the complexity of the transformations involved, it generally results in more precise and clinically relevant alignments. Given the nature of the movements encountered in our specific application —such as tissue compression and expansion— the enhanced accuracy of deformable registration makes it the preferred choice for achieving optimal results. Therefore, this thesis will focus exclusively on DIR techniques to address these complex motion scenarios effectively.

Image registration can be performed between images of the same modality, such as aligning two MRI scans, which generally consists in a simpler task due to consistent imaging characteristics. However, the process becomes significantly more complex when dealing with images from different modalities, such as aligning a CT scan with an MRI to combine the spacial resolution and dense tissue contrast of the former, with the soft tissue contrast of the latter [42]. This is because the images may vary in scale, contrast, and resolution, making it challenging to establish correspondences.

Multi-modal registration algorithms are designed to handle these complexities and can align images from different modalities by finding common features across varying image types. Interestingly, these sophisticated multi-modal algorithms can also be applied to mono-modal datasets, where they may still provide benefits. Conversely, using a mono-modal registration algorithm for multi-modal images is not feasible because it lacks the capability to address the differences in imaging characteristics and modalities. This distinction underscores the versatility of multi-modal techniques while highlighting the challenges involved in solving a multi-modal DIR problem.

On the one hand, mono-modal algorithms typically assume each object is present in both images and maintain its gray levels to some extent [86], which is not the case when considering different modalities. On the other hand, multi-modal registration algorithms address the challenge of aligning images from different modalities by relaxing the strict assumption of intensity preservation. This flexibility, however, comes with increased computational complexity. To manage this, some innovative approaches aim to simplify multi-modal registration by transforming the problem

into a mono-modal context. These methods involve converting images from one modality to another or employing modality-independent representations, such as local phase or Magnitude and phase-based normalised Mutual INformation Descriptor (MIND) [42]. These techniques facilitate registration by aligning images on features that are invariant to modality differences, thus bridging the gap between disparate imaging systems while potentially reducing the complexity of the task. However, the former is highly dependent on the quality of the conversion and the latter may not be discriminative enough to produce relevent transformation estimates.

In this chapter, we delve into the intricate issue of boundary conditions in image registration. These conditions play a crucial role in determining the behavior of the solution, particularly in regions where the neighborhood of pixels or voxels is incomplete. While it is well-established that constraints at the borders significantly influence the overall quality of the registration outcome, the subject has been relatively underexplored. This is primarily due to the complexity involved in defining and implementing appropriate boundary conditions that effectively address the inherent challenges. As a result, this area of study remains a formidable task, demanding both innovative approaches and a deep understanding of the underlying mathematical and computational principles.

Given the absence of ground-truth data in the available database dedicated to IRE procedures, we begin by evaluating the proposed framework for adaptable boundary conditions on publicly available medical datasets. This initial testing phase enables us to rigorously assess the framework's performance, allowing for a comprehensive evaluation of its effectiveness. By leveraging these public datasets, we can observe the improvements introduced by our method both quantitatively—through measurable metrics—and qualitatively—by visual inspection of the results. This approach not only validates the adaptability of the boundary conditions across various medical imaging tasks but also highlights the tangible enhancements offered by our framework, setting the stage for its application to IRE procedures.

We begin by examining existing solutions to the registration process itself in the literature, categorising them into three primary types: landmark-based methods, physical-model-based approaches, and data-based techniques. Additionaly, we summarise more specifically the research tackling boundary conditions in medical image registration, to motivate the proposed method in showing the influence of boundary conditions on the solution. Following this review, we introduce our novel enhancement to DIR algorithms through the use of automatically adaptable boundary conditions. This innovation aims to improve registration accuracy and robustness. Our proposed method is evaluated on two distinct datasets: a mono-modal lung CT dataset, which serves as a control for intra-modality registration,

and a multi-modal abdominal CT to MRI dataset, which tests the effectiveness of our approach across different imaging modalities. We discuss our findings and their implications for future research and clinical applications. The improvement due to the proposed boundary conditions in the context of IRE are finally further studied on a case extracted from our database.

3.2 Review of image registration techniques for medical images

In this section, we explore various approaches to medical image registration that have been developed so far. Registration algorithms can be classified into three main categories. The first category includes landmark-based methods, which require the identification of specific landmarks to align the images accurately. The second group is physical-based methods, where the emphasis lies on modeling the physical characteristics behind the transformation, allowing for more realistic estimation of deformations. Finally, data-based methods focus primarily on the image data itself, utilising the inherent features of the images to achieve the registration process.

3.2.1 Landmark-based approaches

Historically, landmark-based registration represents one of the earliest approaches to image registration algorithms. This method is intuitive in its search of a transformation between corresponding anatomical landmarks across images to align them. These landmarks can be either manually annotated by experts or detected automatically by algorithms. Despite its straightforward concept, landmark-based registration faces significant challenges, particularly in the context of multi-modal registration. In such cases, the features used as landmarks may not be visible or consistent across different imaging modalities, making the task of finding and matching these points notably complex. This limitation underscores the difficulties in achieving accurate alignment when the images do not share common visual features [25].

More recently in [41], landmarks are extracted through the Förstner keypoint operator. In this approach, local information around key landmarks is utilised within a block matching framework that leverages self-similarity descriptors to enhance accuracy. The motion estimation process involves optimising an L_1 similarity measure across a finely sampled range of potential displacements associated with the detected sparse but distinctive keypoints. To ensure inverse consistency, the backward motion from the fixed image to the moving image is also calculated. By averaging the resulting energy values, the method improves the reliability of the

matching process. Following the block matching phase, a dense motion estimate is derived through interpolation techniques such as thin plate splines, which provide a smooth transformation field. This estimate undergoes iterative refinement by re-executing the motion search with a narrower displacement range and/or additional landmarks, thus progressively enhancing the precision of the motion estimate.

In [3], Lobachevsky splines are used for their simple expression and compact support, to parametrise the transformation in a landmark based registration algorithm. Due to the inherent sparsity of the interpolation matrix and the fact that they lead to smaller linear system, they are proven more efficient than Gaussian or thin plates parametrisation. Here, landmarks are readily available which is rarely the case, especially in clinical settings.

3.2.2 Physical-based approaches

Physics-based models concentrate on the fundamental principles governing motion by incorporating physical laws into their design. These models can be broadly categorised into three types.

Regularised models: These models emphasise the incorporation of regularisation techniques to enhance the plausibility and stability of the solutions. By imposing constraints or penalties based on physical realism, they ensure that the generated motions adhere more closely to expected physical behaviors. This approach helps in producing more natural and feasible results while mitigating artifacts that might arise from purely data-driven methods. Additionally, most data-driven approaches without regularisation are ill-posed.

Diffusion regularisation: Also called Tikhonov regularisation, this technique focuses on minimising the first-order spatial derivatives of the deformation field. By smoothing out the gradients in the deformation field, diffusion regularisation helps to reduce abrupt changes and discontinuities. This results in a more gradual and physically plausible deformation, avoiding sharp transitions that are unlikely in anatomical structures.

Curvature regularisation: This approach targets the minimisation of the second-order spatial derivatives of the displacement field. By penalising higher-order variations, curvature regularisation effectively reduces oscillations and irregularities in the deformation field. This is particularly advantageous as anatomical structures typically exhibit smooth, continuous deformations rather than abrupt or oscillatory changes. Consequently, curvature regularisation contributes to producing anatomically realistic and smooth deformation fields [25].

A generalised formulation was proposed in [60], making use of fractional order derivatives:

$$S = \frac{1}{2}||\Delta^{\alpha/2}T||^2, \tag{3.2.1}$$

where $\alpha \in [1, 2]$, $\Delta = \nabla^2$ the Laplace operator and T is the transformation field.

It allows to regularise in between diffusion and curvature to obtain potentially better registration, both in terms of image similarity and transformation smoothness.

There are more specialised, physics-driven methods that leverage fundamental physical principles such as elastic and fluid regularisation [60].

Elastic regularisation: It uses the concept of elastic energy to guide the deformation process. The regularisation term is derived from the energy associated with elastic deformations, ensuring that the deformations remain physically plausible by penalising excessive changes. It assumes that the stress within a material is linearly related to the strain. As a result, this approach excels in maintaining smooth and realistic transformations but may struggle with large deformations, as it inherently limits the extent of allowable movement.

To paliate this issue in [39], they introduce a linear elastic continuum mechanical model designed specifically for large deformation estimation. This model leverages the principles of linear elasticity to provide a robust framework for understanding and predicting deformations. The model treats displacement estimates $\mathbf{T}$ as a sequence, such that:

$$I_m = \mathbf{T}_n(\mathbf{T}_{n-1}(...(I_f))) \tag{3.2.2}$$

and

$$I_f = \mathbf{T}_{2n}(...(\mathbf{T}_{n+1}(I_m))), \tag{3.2.3}$$

where I_f is the fix image, I_m is the moving image, and n is the number of iteration necessary for a transformation estimate. The sequential approach allows a dynamic representation of deformation over time or across various iterations, and provides a detailed view of how displacement evolves, enhancing the accuracy of the deformation predictions.

Furthermore, the motion is parameterised using 4D Fourier transforms, to estimate low-frequency components before high-frequency ones, allowing the algorithm to capture and register global features of the image first. By focusing on these broad, global features initially, the system ensures a robust alignment of large-scale structures before refining the details with higher-frequency adjustments.

The model incorporates a sophisticated second-order regularisation on the displacement field, applied across both spatial and temporal dimensions, utilising

a linear elastic operator. This advanced regularisation technique is specifically designed to capture the inherent mechanical properties of elastic anatomical structures.

To minimise inverse consistency error, the iterative process estimates both the forward ($T_{moving \to fix}$ from moving image to fix image) and the backward ($T_{fix \to moving}$ from fix image to moving image) and define a symmetric similarity.

Fluid regularisation: Alternatively, the fluid regularisation is based on the Navier-Stokes equations, which describe the behavior of fluid flow. By applying these equations to regularise deformations, this method ensures that the transformations adhere to the principles of fluid dynamics. However, solving the Navier-Stokes equations is computationally intensive and time-consuming, often requiring significant resources to achieve accurate and stable solutions.

Total Variation (TV) regularisation: An additional type of regularisation can be introduced to manage the shearing effects observed during processes like breathing, where the liver moves against the pleural wall. This motion causes friction and tension, requiring a sophisticated approach to handle the resulting deformation.

The method described in [95] employs isotropic TV regularisation, which is particularly effective in preserving sharp edges while minimising noise:

$$S_{TV}(T) = \sum_{\mathbf{r} \in \Omega} \sqrt{\sum_{i,j \in \{x,y,z\}} (\nabla_i T_j(\mathbf{r}))^2}. \tag{3.2.4}$$

This technique facilitates the modeling of non-smooth displacements, which are essential for accurately capturing complex anatomical interactions. The isotropic TV regularisation ensures that these displacements are represented while preserving spatial coherence, effectively balancing the need for capturing sharp changes with maintaining overall smoothness in the displacement field.

A significant limitation of this regularisation term is its inherent non-differentiability, which complicates the use of gradient-based optimisation methods. To overcome this challenge, the Alternating Direction Method of Multipliers (ADMM) is employed for cost function optimisation. ADMM is particularly well-suited for handling non-differentiable terms by decomposing the problem into more manageable sub-problems that can be solved iteratively. Additionally, the motion is parameterised using first-order B-splines, with control point displacements strictly constrained. This approach ensures precise control over the displacement parameters and prevents the issue of overshooting, thereby maintaining stability and accuracy in the optimisation process.

Purely physical models: These models are firmly rooted in fundamental physical principles, meticulously considering geometric configurations, material prop-

erties, and other critical physical attributes. By simulating the inherent physical forces and interactions, they strive to faithfully reproduce the real-world dynamics of objects and materials. This approach offers a highly detailed and accurate depiction of motion, capturing the true physical behavior of the system under study. However, such models are relatively rare and can be computationally intensive. They necessitate a detailed discretisation of the image to align with the physical properties of the tissue, which can be demanding in terms of computational resources and complexity.

For instance in [102], a finite element contact impact analysis provides a robust framework for simulating breathing motion by integrating detailed material properties and geometric configurations into a finite element model. Initially, it is crucial to accurately extract and define these properties to set up the model. The simulation begins with a state of zero motion, where the model starts with no initial displacement. As the simulation progresses, a gradual change in intrathoracic pressure is applied, mimicking the physiological process of breathing. This approach captures the dynamic interaction at the interface between the lungs and surrounding structures, allowing for precise modeling of lung deformation during respiration. Given the complexity of these interactions, the method relies significantly on a deep understanding of the underlying motion dynamics to ensure accurate and realistic results. This comprehensive approach is essential for reflecting the true behavior of lung mechanics in response to varying pressure conditions.

Parameterised methods: Lastly, parameterising the solution to the registration problem offers significant advantages by embedding physical constraints directly within the parameterisation itself. This approach can effectively eliminate the need for explicit physical constraints, as the chosen parameterisation inherently incorporates the necessary regularisation. By doing so, it provides enhanced control over motion properties and accelerates convergence, primarily because it reduces the dimensionality of the search space.

For example, by selecting an appropriate parameterisation, one can easily enforce constraints such as inverse consistency or sliding motion consistency. This is achieved without requiring additional computational overhead for these constraints. The optimisation process is typically conducted using nonlinear least squares methods, guided by a predefined similarity metric that measures how well the registration aligns with the target criteria. However, it is important to note that the cost function in this context can be non-convex, which introduces sensitivity to initial conditions and can affect the stability and reliability of the solution [13].

In [13], a parametric approach making use of cubic spline is associated with block matching. The parameter search is performed by solving a constrained non-linear least square problem on a subset of locations. The minimisers are the

spline coefficients solving an L_1 optimisation problem, with sparsest residual error vector for the block match estimate. There is no guarantee on the accuracy of the matching as the optimal point match might not lead to the optimal spatial accuracy but a lower error in the non-linear least square problem suggest higher spatial accuracy. The method is generalisable to other parametrisation and diverse similarity metrics may be used for the block matching.

3.2.3 Data based approaches

Data-based techniques leverage directly observable information from the data, focusing on pixel or voxel intensity values. This approach prioritizes data fidelity, ensuring that the registration process accurately reflects the observed intensities within the images. Such techniques are particularly valuable in applications where maintaining the integrity of the original data is crucial. Among the various subgroups within image registration methods, data-based techniques have garnered significant scientific interest due to their robustness. It is worth noting that regularisation is still mandatory as the minimisation of a data fidelity term alone is an ill-posed problem [86].

Emphasis on data fidelity A well known intensity-based registration algorithm is demons [36]. It is designed to be both swift and fully automated, streamlining the alignment process for efficiency. This approach is predicated on the assumption of small displacements, necessitating a multi-scale framework. This framework is crucial for mitigating voxel-wise displacements and minimising computational time, as it progressively refines the alignment at multiple levels of detail. In this method, voxel intensities in the fixed image generate local forces that guide the displacement of voxels in the moving image. such that:

$$\mathbf{T}^{t+1} = \mathbf{T}^t + \frac{(I_m^t - I_f)\nabla I_f}{(I_m^t - I_f)^2 + |\nabla I_f|^2}, \tag{3.2.5}$$

where $\mathbf{T}$ is the deformation field, t is the iteration number and I_m and I_f are the intensities of the moving image and the fix image respectively.

This mechanism ensures that the alignment is precise and adheres closely to the intensity patterns of the fixed image, enhancing the accuracy of the registration process while maintaining speed and automation.

However, this formulation presents an inherent challenge: it is underdetermined, meaning there is no single unique solution. This lack of a definitive solution arises from the formulation's inability to fully constrain the problem. Consequently, the stopping criteria employed are typically based on the convergence of successive solutions, rather than on direct image similarity metrics. This approach involves

evaluating the differences between iterations to determine when to halt the process, which helps ensure that the solution stabilises even in the absence of a clear, unique result. Such criteria are crucial for managing the complexity and ensuring the robustness of the solution in the face of the inherent underdetermination.

Some variants utilise a Gaussian filter to maintain geometric continuity while effectively reducing noise, ensuring that important structural details are preserved without distortion. Other approaches employ active forces or modified passive forces equipped with normalisation factors, which enable dynamic adjustment of strength to better match and align images. Additionally, the incorporation of inverse consistency mechanisms further refines accuracy by ensuring that transformations are consistent when applied in reverse, thus improving overall registration quality [36]. These enhancements collectively contribute to more robust and precise image registration outcomes.

Commonly used metrics are grounded in the principle of brightness constancy, which assumes that the intensity of corresponding pixels should remain consistent before and after movement. This assumption leads to the use of various quantitative measures, such as squared differences, the Pearson correlation coefficient:

$$\rho(I_f, I_m) = \frac{cov(I_f, I_m)}{std_{I_f} \cdot std_{I_m}}, \tag{3.2.6}$$

or mutual information:

$$MI(X, Y) = \sum_{\mathbf{r}} \sum_{\mathbf{r'}} P_{(I_f, I_m)}(\mathbf{r}) log(\frac{P_{(I_m, I_f)}(\mathbf{r})}{P_{I_f}(\mathbf{r}) P_{I_m}(\mathbf{r'})}, \tag{3.2.7}$$

which is less restrictive but global: the data fidelity term is the same everywhere on the image [25]. Thus, the latter performs poorly on non-rigid tasks.

On the one hand, brightness-based terms, which rely on the assumption of brightness constancy, often perform well in mono-modal scenarios where both images share the same intensity distribution. On the other hand, correlation-based terms are versatile and can handle both mono-modal and multi-modal image registration tasks effectively. These metrics evaluate the degree of linear relationship between image intensities, making them robust to variations in intensity distributions across different modalities and capable of accommodating non-rigid transformations.

An instance of information-based data fidelity term is presented in [90], where they formulate a symmetric energy functional using either mutual information or normalised mutual information. Symmetry is imposed through graph-based volume forms to adress asymetries due to discretisation of the solution:

$$D_S(T) = 1 + det(Jacobian(T)) \cdot T^*, \tag{3.2.8}$$

where T^* is the standard volume form of T on Euclidean space.

This formulation ensures that positive semi-definiteness is maintained, which guarantees that the objective function is both well-defined and has a meaningful minimum. Positive semi-definiteness ensures that the Hessian matrix of the objective function is non-negative, implying that the function is convex or at least not negatively curved. This property is crucial for ensuring stable optimisation and reliable solutions.

Additionally, the formulation induces invariance under volume-preserving transformations. This means that the function's value remains unchanged when transformations that preserve volume are applied, such as certain linear transformations or rotations. This invariance is essential for ensuring that the algorithm does not introduce bias based on specific orientations or scales of the input data. As a result, the formulation does not favor any particular transformation when constant images are processed, thus preventing skewed results due to preferential treatment of certain transformations.

Feature-based approaches: Relevant information can be extracted from the intensities to facilitate registration. These are generally less sensitive to noise and reduce redundancies and resolution requirements for optimal solutions [7]. In particular, the use of modality-independent features significantly enhances this process by allowing for a unified approach to both mono-modal and multi-modal registration. These features are designed to be invariant to the specific imaging modality, meaning they can generalise across different types of images (e.g., CT, MRI, etc) without needing modality-specific adjustments. This capability streamlines the registration process, enabling the alignment of images from diverse sources into a single cohesive framework, thereby simplifying and improving the accuracy of the registration task.

One such feature is extracted by Scale Invariant Feature Transform (SIFT) [63]. Its main advantage is its invariance to scale and rotation, which ensures that keypoints can be reliably detected regardless of changes in image size or orientation. Additionally, SIFT exhibits robust performance against affine distortions, occlusion, clutter, and noise, making it highly effective in challenging visual environments.

Originally designed for object recognition, SIFT has also proven valuable in image registration tasks. The algorithm employs a cascade filtering approach to feature extraction, focusing computational resources on specific areas of interest rather than the entire image, which helps manage its inherent computational demands. It operates by analysing scale space extrema, which allows for the extraction of stable features across varying scales. Each of these extrema is assigned a consistent orientation based on the local image gradient, leading to the generation of distinctive descriptors. These descriptors correspond to the keypoints

identified at different scales, facilitating accurate matching and alignment in both object recognition and registration contexts.

Another example is Speeded Up Robust Features (SURF) [7]. This particular feature is scale and rotation invariant. Furthermore, it offers better robustness and distinctiveness compared to SIFT. It is composed of a feature detector, using the Gaussian second derivative mask to detect large intensity gradient variations in multiple directions:

$$\frac{G_{xx}(x,y,\sigma)\cdot G_{yy}(x,y,\sigma) - G_{xy}(x,y,\sigma)^2}{\sigma^2}, \tag{3.2.9}$$

with

$$G_{ij}(x,y,\sigma) = \frac{\delta \mathbb{N}(0,\sigma)^2}{\delta i \delta j} * I(x,y), \tag{3.2.10}$$

and a feature descriptor, determining the orientation of each detected features with the local Haar wavelet response. This definition can be extended to higher dimensions.

In [64], the nearest correspondence between SURF descriptors is found using KD tree based nearest neighbor search. The transformation is then determined with RANdom SAmple Concensus (RANSAC) algorithm, using a perspective transformation model. RANSAC estimates the perspective transformation model parameters in a hypothesize and test framework. During the hypothesize phase, the parameters are determined from a random minimal set of relevant features. Then, the model is tested on the whole dataset to verify how much data is actually consistent with the parameters estimate. Those two steps are repeated until a stopping criteria is verified.

MIND is another such feature that can be used in multi-modal cases. It assumes that intensity distributions are locally reliable across modalities. It is computed based on self-similarity between patches such that:

$$MIND = \frac{1}{N} exp(-\frac{L_p(I,\mathbf{r},\mathbf{r}+\zeta)}{V(I,\mathbf{r})}), \tag{3.2.11}$$

where L_p is the voxel wise square distance between patches of size p, ζ defines the search region, N is a normalising constant, and V estimates the local variance to account for noise perturbation.

This feature framework looks for local structures preserved across modalities. MIND is robust to non-functional intensity relations, noise and non-uniform bias fields, a common artefact in MRI. Moreover, it is highly directional and multi-dimensional [42], thus carrying a good amount of information. However, it is sensitive to low contrast, patch size, and texture changes from one modality to the other [86].

Image gradients can also be effectively leveraged across modalities, as exemplified by the EVolution algorithm's data fidelity term [86]. This approach is grounded in the premise that structural contours remain discernible across various imaging modalities, despite potential discrepancies in intensity and texture representation. By harnessing these gradients, the algorithm capitalises on the underlying geometric similarities between images, allowing for robust alignment and enhancing the fidelity of multi-modal registration. This innovative utilisation of image gradients underscores the versatility and potential of gradient-based techniques in bridging the gaps between different imaging modalities, ultimately contributing to more accurate and meaningful analyses. The data fidelity term is defined as:

$$D(\mathbf{T}) = exp\left(\frac{\int_\Gamma |\nabla_{I_f}(\mathbf{T}(\mathbf{r})) \cdot \nabla_{I_m}(\mathbf{r})|\mathbf{dr}}{\int_\Gamma ||\nabla_{I_f}(\mathbf{T}(\mathbf{r}))||_2 ||\nabla_{I_m}(\mathbf{r})||_2 \mathbf{dr}}\right), \tag{3.2.12}$$

with Γ a patch centered at $\mathbf{r}$, I_f the fix image, I_m the moving image, T the transformation field.

By maximising edge alignment, EVolution effectively captures and preserves critical structural details, making it a powerful tool for high-precision multi-modal image analysis. This method necessitates a pronounced contrast between physiological tissues, ensuring precise differentiation. However, it demonstrates remarkable invariance to contrast reversals, enhancing its resilience. Additionally, it significantly boosts robustness to dynamic anatomical structures, both incoming and outgoing, offering a substantial improvement over MIND-based algorithms [86].

Finally, supervoxels may be used, considering regional intensities instead of local intensities. They have the advantage to preserve pixel relationships and image structure. For instance in [96], supervoxels, computed by iterative dichotomy are used in conjunction with information entropy to eliminate redundant feature points:

$$H(I) = -\sum_{i=1}^{L} P_i(\mathbf{r}) log(P_i(\mathbf{r})), \tag{3.2.13}$$

where $i \in [1, L]$,

$$P_i(\mathbf{r}) = \frac{h_i(\mathbf{r})}{\sum_{i=1}^{L} h_i(\mathbf{r})}, \tag{3.2.14}$$

and $h_i(x)$ is the cumulative number when the state is i.

The selected supervoxels are then registered with the RANSAC algorithm, thus improving registration speed. Though promising, this technique has only been tested on rigid transformations and on a mono-modal case at the moment.

We may also cite features such as local phase, local entropy, or gradient orientation [86].

Machine-learning based approaches: In recent advancements, deep learning-based methods have emerged as a powerful tool for image registration tasks. These methods have demonstrated remarkable improvements in performance, akin to their success in other areas of image processing. The application of deep learning has led to enhanced accuracy and efficiency in image registration, often surpassing traditional approaches.

Nevertheless, a significant challenge remains in fully understanding the underlying mechanisms that drive these deep learning models. Specifically, the generation of solutions and the metrics employed for evaluation in these methods can be opaque. This lack of transparency makes it difficult to justify their use compared to classical algorithms, which have well-understood processes and metrics. As deep learning models are often seen as black boxes, their adoption in registration tasks requires careful consideration of their effectiveness and the interpretability of their outputs. Furthermore, some annotation is crucial for the learning [8], but it is not easily obtainable and synthetic database may not represent natural movements.

Motion estimation: Deep learning offers a powerful approach for generating motion estimates directly from data, utilising sophisticated neural network architectures. These models can learn complex motion patterns by analysing large datasets, potentially yielding highly accurate predictions. However, designing an effective loss function for these tasks can be quite challenging. It often requires incorporating multiple components to capture various aspects of the motion, such as spatial consistency, temporal dynamics, and smoothness constraints. Balancing these terms to achieve an optimal performance can be intricate, as each term needs to be weighted appropriately to guide the learning process effectively.

A CNN doing so is SymReg-GAN [105]. A Generative Adversarial Network (GAN) comprises two distinct yet interconnected CNNs: the generator and the discriminator. The generator creates synthetic data samples with the goal of mimicking the characteristics of real data from a database. Conversely, the discriminator evaluates these samples, distinguishing between those generated by the generator and genuine samples from the database.

The interaction between these two networks is driven by an adversarial process (Fig. 3.2). The generator's objective is to produce increasingly convincing samples to deceive the discriminator, while the discriminator's goal is to improve its accuracy in distinguishing between real and fake data. This dynamic forms a game-like scenario where the generator and discriminator continuously challenge each other.

Their loss functions are intricately designed to reflect this adversarial nature. The generator's loss function is aimed at maximising the discriminator's error, while the discriminator's loss function is designed to minimise its error in classifying real and fake samples. The convergence of this adversarial game results in the generator producing high-quality samples that are indistinguishable from real data.

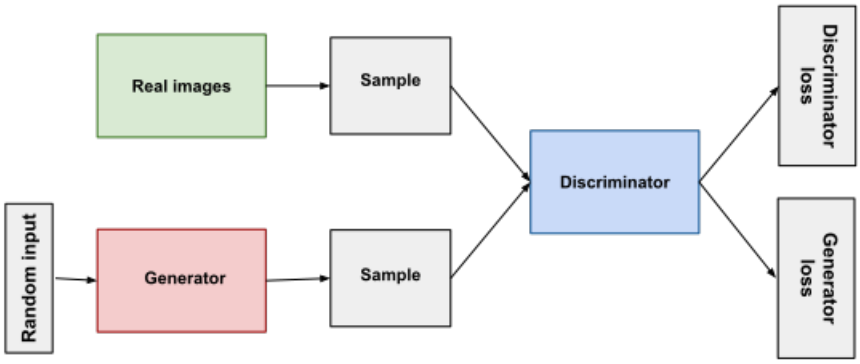


Figure 3.2: The overall framework for GANs encompasses two primary components: the generator and the discriminator. The generator is responsible for creating new data samples, aiming to produce outputs that closely resemble those from the actual dataset. Meanwhile, the discriminator's role is to distinguish between the samples generated by the generator and those from the real dataset. This adversarial process, in which the generator and discriminator are trained simultaneously, drives the continuous improvement of the generator's ability to produce realistic data, from developers.google.com.

The researchers utilised a sophisticated architecture involving two GANs, each conditioned by the images being registered to foster symmetry between modalities. This dual-GAN setup enhances the registration process by promoting alignment and coherence in the transformed images.

Each GAN comprises several integral components: a modality translator designed to mitigate modality-specific biases, a non-linear transformation regressor to handle complex, non-linear mappings between the images and a spatial transformer to provide an additional layer of flexibility by allowing spatial manipulations within the network. The loss includes a data fidelity term L_{fid} depending on both the forward and the backward motion estimates, a symmetry term L_{sym}, a regularisation term L_{smt} and an adversarial term L_{GAN}:

$$L_{fid}(I_f, I_m, \mathbf{T}, \mathbf{T}_{inv}) = \sqrt{((I_f - \mathbf{T}(H(I_m)))^2 + \epsilon^2} + \sqrt{((\mathbf{T}_{inv}(H_{inv}(I_f)) + I_m)^2 + \epsilon^2}, \tag{3.2.15a}$$

$$L_{sym}(I_f, I_m, \mathbf{T}, \mathbf{T}_{inv}) = ||_f I - H_{inv}(\mathbf{T}_{inv}(\mathbf{T}(H(I_f))))||_2^2 + ||I_m - H(\mathbf{T}(H_{inv}(I_m))))||_2^2, \tag{3.2.15b}$$

$$Lsmt(\mathbf{T}, \mathbf{T}_{inv}) = ||\nabla^2 \mathbf{T}||_2^2 + ||\nabla^2 \mathbf{T}_{inv}||_2^2, \tag{3.2.15c}$$

$$L_{GAN}(F, G, I_f, I_m, I_f^*, I_m^*, \mathbf{T}^*) = \mathbb{E}_{I_f^*, I_m^*, \mathbf{T}^*} log(F(I_f^*, I_m^*, \mathbf{T}^*)) \tag{3.2.15d}$$
$$+ \mathbb{E}_{I_f, I_m} log(1 - F(I_f, I_m, G(I_f, I_m))),$$

where I_f is the fix image, I_f^* is a fix image from the annotated database, I_m is the moving image, I_m^* is a moving image from the annotated database, $\mathbf{T}$ it the forward motion estimate, $\mathbf{T}_{inv}$ is the backward motion estimate, $\mathbf{T}^*$ is a transformation field from the annotated database, F is the discriminator, G is the generator, H is the modality translator from moving modality to fix modality and H_{inv} is the modality translator from the fix modality to the moving modality.

During the training phase, the network is designed to handle both labeled and unlabeled data effectively. When using labeled data, the network incorporates a supervised loss term, denoted as L_{SPV}, which helps guide the learning process by penalising incorrect predictions based on the true labels. This enables the network to learn from explicit examples and refine its performance accordingly.

CNN can also be used conjointly with parameter-based methods to estimate the parameters [26]. However, these methods still require expensive to obtain groundtruth without removing the iterative nature of the process.

Learned metric: CNNs can be effectively employed to develop modality-independent metrics, which are crucial for assessing the alignment of registered images against a reference image. In this context, a CNN learns to evaluate whether an image has been accurately aligned or registered with a fixed reference image, irrespective of the specific modalities involved.

However, a significant challenge arises in this process due to the nature of the metrics being learned. These metrics are frequently characterised by non-smooth and non-convex properties [26]. Non-smoothness implies that the metric can exhibit abrupt changes or discontinuities, while non-convexity indicates the presence of multiple local minima. These characteristics complicate the optimisation process, making it more difficult for conventional optimisation algorithms to converge to an optimal solution. Consequently, advanced optimisation techniques or modifications to the learning algorithms may be required to address these difficulties and achieve robust performance in modality-independent image registration. When accomodated properly, those metrics often lead to better registration accuracy [8]. The

resulting algorithm is still relatively computationally demanding as it requires an iterative registration process.

Learned regularisation: Another approach is to incorporate a high fidelity motion model in a Plug-and-Play algorithm [83]. The method in question blends the adaptability of traditional model-based techniques with the robust performance of modern machine learning algorithms. This approach alternates between utilising a model and applying a regulariser that is finely tuned based on data-driven insights. The regularisation can be informed by various advanced techniques, including Markov Random Field (MRF)—learnable undirected graph—, CNN image denoiser, CNN artifact remover, etc. By integrating these techniques through proximal and splitting algorithms, the method ensures a sophisticated interplay between model-based flexibility and the data-driven precision of machine learning. This results in a powerful tool for handling complex image processing tasks with both structural and statistical efficacy. [55]. In the case of image denoisers, they should be non-blind and handle wide range of noises since their generalisability could otherwise be very limited [104]. Instances of proximal and splitting methods include:

- ADMM: decomposes a large problem into smaller, more manageable subproblems, which are easier to solve. The ADMM algorithm iteratively updates the solutions to these subproblems, leveraging the method of multipliers to ensure convergence to an optimal solution. ADMM is particularly useful for problems involving large-scale data and complex constraints, where traditional optimisation methods may be inefficient,

- primal-dual splitting: solves large-scale convex minimisation problems by addressing both the primal and dual formulations simultaneously. This method is particularly useful for problems involving complex constraints and non-smooth terms. It works by decomposing the original problem into simpler subproblems, which can be solved iteratively. The primal-dual splitting approach applies gradient and linear operators directly without inversion, making it efficient for high-dimensional problems. By alternating between updates to the primal and dual variables, the method ensures convergence to an optimal solution,

- Fast Iterative Shrinkage Thresholding Algorithm (FISTA): works by combining proximal gradient steps with an accelerated gradient descent method, allowing it to reduce the number of iterations required to reach a solution. This makes FISTA more efficient and effective for large-scale problems where quick convergence is crucial,

- Half Quadratic Splitting (HQS): aims to minimise a cost function by alternating between two simpler optimisation problems. It works by splitting

the original problem into subproblems that are easier to handle, making the overall problem more tractable. HQS introduces auxiliary variables to reformulate the original non-convex problem into a sequence of convex problems, which can be solved iteratively. This approach leverages the quadratic form, allowing the problem to be decomposed and solved efficiently. The method is particularly useful for problems that are ill-posed or NP-hard.

These methods excel in improving performance by efficiently solving non-convex optimisation problems, especially in high-dimensional settings. Despite their effectiveness, these methods face limitations in generalisability because their learning phase is tightly coupled with the specific training data. Consequently, they may struggle with novel or unseen data. To address the challenge of large-scale images, such as those encountered in 3D medical imaging, approximations of gradients and proximal operators are often employed to manage computational complexity and ensure practical applicability.

CNNs learning a regularisation can also be integrated in a framework with another CNN estimating the transformation. Such a method was proposed in [83] (Fig 3.3). The loss function of an unsupervised registration CNN is composed, similarly to most variational technique, of a similarity term and a regularisation term. Here, the latter is a feasibility prior of motion learned by a U-Net architecture on the Jacobian of deformation fields derived from images of higher resolution than the one eventually fed to the registration network. It is used to regularise the Jacobian of the produced deformation field. The Jacobian matrix represents how the local deformation changes in space, and the feasibility prior helps to regularise these deformations by constraining them to plausible physical movements. This regularisation helps in reducing noise and artifacts, leading to enhanced robustness and accuracy of the registration. However, this approach is highly tailored to specific types of motion that the U-Net has been trained to recognize, which may limit its generalisability. Additionally, the requirement for high-resolution images for training the U-Net can be a limitation, as such data may not always be readily available in practical scenarios.

Bridge between modalities: To advance multi-modality deep learning, two primary approaches are utilised.

CNNs can be employed to translate data from one modality to another. This involves training networks to map information between different types of data, such as translating MRI images to CT scans or vice versa. This translation process helps integrate diverse data sources, allowing models to work cohesively across various modalities. For instance, a modality translator is used in the aforementioned SymReg-GAN [105].

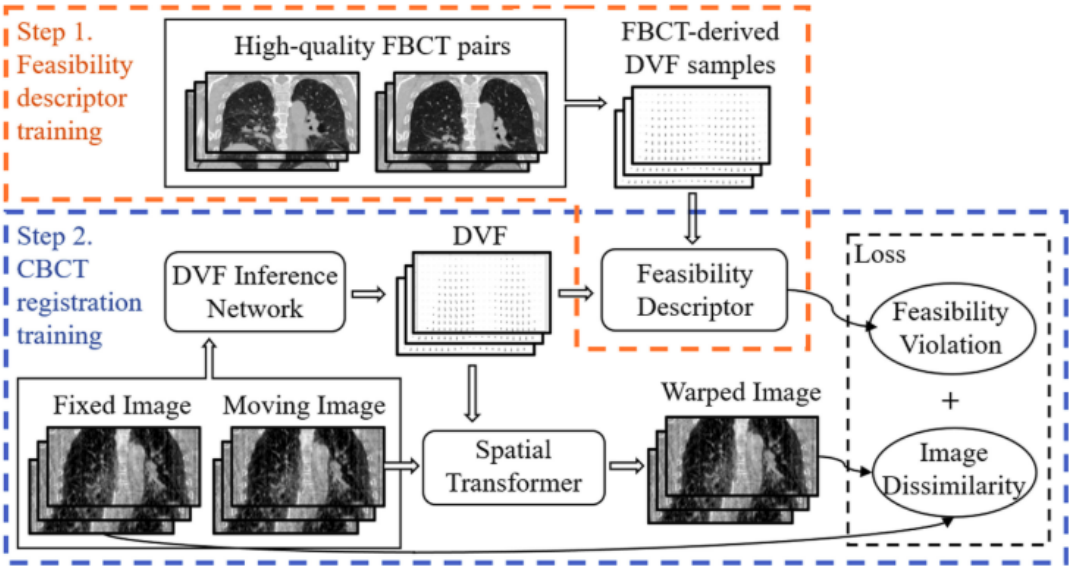


Figure 3.3: The methodology outlined in [83] involves an integrated approach where Fan Beam Computed Tomography (FBCT) serves as the high-resolution reference image. The deformation between the FBCT and CBCT images is quantified and described using Deformation Vector Fields (DVF), which provide a detailed representation of the spatial transformations required to achieve alignment between the two imaging modalities, reproduced with permission.

Alternatively, networks can be designed to learn modality-independent descriptors. These descriptors capture features that are invariant to the specific modality of the data. Once these features are learned, they can be used for image registration, aligning images from different modalities by comparing their shared, modality-agnostic characteristics

Both methods enhance the ability of deep learning models to effectively process and integrate multi-modal data.

3.2.4 The challenge of boundary conditions in image registration

Boundary conditions are constraints necessary for solving differential equations that describe physical systems. These conditions specify the behavior of a system at its boundaries, ensuring unique and meaningful solutions. Without proper boundary conditions, the equations may yield infinite or non-physical solutions, rendering them useless for practical purposes. Thus they play a crucial role in

determining the final quality of solutions for registration problems.

When neighboring information is lacking at the borders, the registration task becomes significantly more challenging. This difficulty arises because the lack of contextual data at these edges can lead to inaccuracies in alignment and motion estimation.

Research highlights that accurately estimating motion at the boundaries is currently one of the most challenging aspects of optical flow methods [6]. This difficulty is not isolated but rather extends to many variational methods used in image registration. The absence of adequate boundary information often results in suboptimal performance and reduced accuracy in these methods.

Mis-registration at the boundaries especially arises in scenarios where images capture only partial fields of view, a situation commonly encountered in medical imaging. Also, the issue may be caused by an absence of well-defined, sharp boundaries between the anatomical structures of interest and the surrounding background [12]. In such cases, the overlap between regions of interest and less distinct areas can lead to inaccuracies in aligning and integrating images. This problem is particularly prevalent in medical sciences due to the inherent complexity and variability in anatomical structures, which often results in blurred or ambiguous borders that challenge precise image registration and fusion. This error easily propagates towards the Region Of Interest (ROI), even when said region is at the center of the image. This is thus especially problematic in the case of partial fields of view where the ROI is inherently closer to the boundaries. These practical limitations come from the imaging sensor or the data acquisition process itself, when multiple scans need to be conducted within a reduced timeframe for instance.

At all times, it is crucial that boundary conditions complement the regularisation operator used in the solution process. If the boundary conditions are not aligned with the regularisation strategy, the problem can become ill-posed, leading to unstable or inaccurate solutions [16]. This alignment ensures that the regularisation process effectively addresses the inherent instability and ambiguity of the problem. The regularisation operator stabilises the solution by introducing additional constraints or smoothing, while the boundary conditions define the constraints at the edges of the domain.

Despite this problem being known, it is rarely adressed in literature. To our knowledge, the only study on the impact of boundary condition on image registration is [12]. A synthetic database was used to test generic boundary conditions, namely periodic:

$$T[0] = T[L], \tag{3.2.16}$$

where L is the size of the 1D vector T, homogeneous Dirichlet boundary, imposing a null motion field:

$$T[x] = 0, \tag{3.2.17}$$

where $x \in \delta\Omega$, the boundary, and homogeneous Neumann, restricting the derivative (or the shear tensor along the normal direction for higher dimensions):

$$\frac{dT}{dx} = 0, \tag{3.2.18}$$

where $x \in \delta\Omega$.

The study found that periodic boundary conditions deliver reasonably good results primarily when the ROI is sufficiently distant from the boundaries, minimising the effects of any mis-registration or boundary distortions. In contrast, homogeneous Dirichlet boundary conditions are advantageous only under particular circumstances where boundary movement is negligible. This requires prior knowledge on the motion to ensure accuracy. Overall, homogeneous Neumann boundary conditions tend to provide better performance across a wide range of scenarios.

However those simple boundary conditions are not suitable for very generic movements such as a rotation, as seen in Figure 3.4.

In the image registration literature, boundary conditions frequently receive minimal attention or are assumed to be homogeneous. When boundary conditions are specified, homogeneous Dirichlet conditions are typically employed for scenarios where the field of view is sufficiently expansive, ensuring that the ROI remains unaffected by edge effects or boundary movements. This choice is suitable when the domain's boundaries do not influence the critical areas being analysed. Conversely, in cases where the ROI is more susceptible to boundary effects—such as in smaller or partially observed fields—homogeneous Neumann boundary conditions are often utilised. These conditions help manage the influence of boundary flow field and are better suited for mitigating errors that could arise from boundary misalignments or constraints.

Addressing the issue of boundary conditions in image registration is rare and typically tailored to specific applications. These specialised approaches frequently shift the focus from conventional image boundaries to the boundaries of the ROI. This adjustment allows for handling complex transformations, including shearing and other non-smooth motions, which are not well-managed by standard methods. Such methods are highly sensitive to the quality of the segmentation process; if the segmentation is inaccurate, it directly impacts the registration outcome. Additionally, the transformation estimates are only valid within the segmented mask. Outside this mask, the estimated transformation field may lack meaningful or reliable information, which can limit the utility of the registration in broader contexts [95].

In [76], the approach leverages segmentation masks to reinforce the distinction between background and foreground pixels. By integrating these masks into

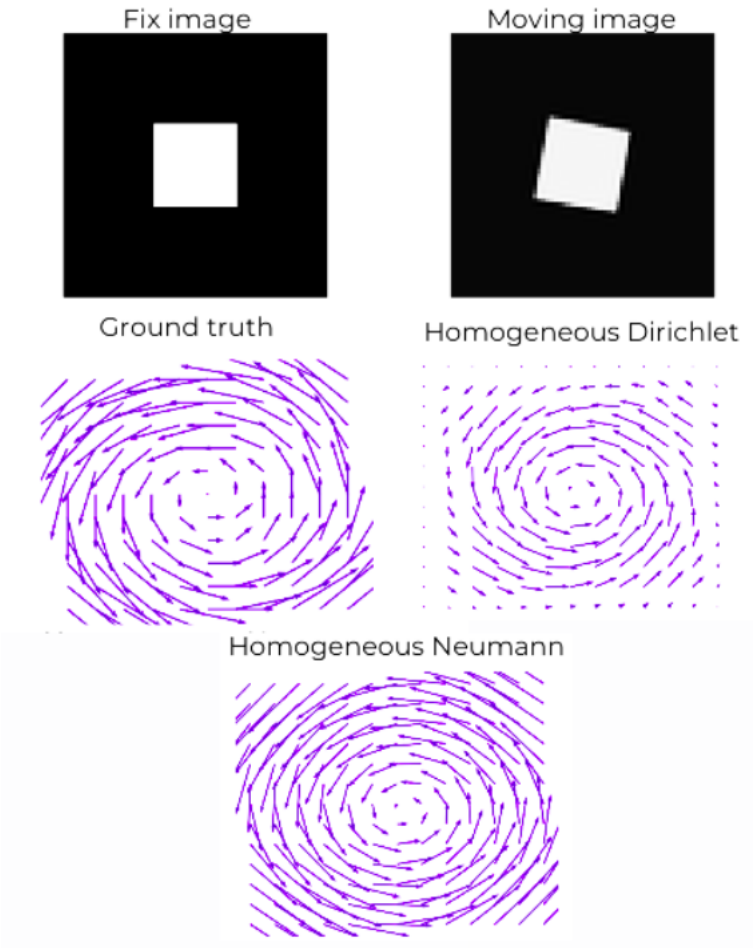


Figure 3.4: Motion field estimated by the EVolution registration algorithm [86] for different generic boundary conditions, namely homogeneous Dirichlet (imposing a null displacement) and homogeneous Neumann (imposing a null shear tensor), on a registration task consisting of a rotated square. Neither of the generic global boundary conditions is adequate for the estimation of a rotational movement.

the boundary condition framework, the method encourages background pixels to consistently remain registered in the background, and reciprocally for foreground pixels. This is achieved through a sophisticated implementation of free boundary conditions, where an additional term is introduced into the energy minimisation process. This term represents the conditional probability that a pixel does not belong to the background, thereby improving robustness against poor initialisations of the ROI. This marks a significant advancement over previous methods that utilised non-homogeneous Dirichlet boundary conditions on preliminary contours,

which were less adaptable. While segmentation masks are effectively employed in this technique, alternative strategies such as simpler thresholding functions or segmentations provided by CNNs could also serve as effective conditional probability distributions, potentially simplifying the implementation while maintaining performance.

Similarly in [102], the estimation of movement relies exclusively on the boundaries defined within the ROI. This approach is deeply influenced by the physical dynamics of breathing. Initially, the boundary conditions are set using a homogeneous Dirichlet condition, which provides a uniform constraint across the boundary. As the simulation progresses, these initial conditions are gradually refined to better reflect the pressure variations within the lungs. This adaptive approach allows for a more accurate simulation of the respiratory mechanics, capturing the complex interplay between lung expansion and contraction. The gradual adaptation ensures that the boundary conditions evolve in response to the physiological changes, thereby enhancing the fidelity of the movement estimation and improving the overall model accuracy.

In machine learning approaches, incorporating boundary conditions significantly enhances the physical realism of motion fields. By embedding these constraints, models can generate more accurate and plausible predictions that align with real-world physics. However, integrating such constraints can introduce complexity, making it more challenging to precisely constrain and solve the models. The trade-off involves balancing the benefits of improved physical accuracy with the increased difficulty in managing and optimising the constraints effectively.

In [22], a U-Net architecture is employed to estimate transformation fields using a one-shot learning strategy. This entails training the network to perform image registration on the pair of image under consideration until convergence is achieved. To ensure the quality and physical plausibility of the estimated transformation fields, boundary conditions are seamlessly integrated into the loss function. Specifically, the loss function incorporates a regularisation term that promotes boundary smoothness at patch borders, reducing artifacts and enhancing continuity. Additionally, a cyclic constraint is applied, which enforces that the sum of deformation vectors along the trajectory of a boundary voxel approximates zero. This cyclic term helps maintain consistency and prevent drift in the deformation fields, contributing to more accurate and reliable registration outcomes.

A thorough examination of previous research reveals a notable gap: the absence of a comprehensive, unified framework for boundary conditions that can be applied broadly across diverse medical image registration tasks. Most of the existing methods reviewed in the context of boundary conditions are highly specialised, designed to address the nuances of specific applications—whether it be the registration of lung scans during the respiratory cycle or mammographic images taken at various

stages.

This fragmentation highlights a critical need for a more generalised approach to boundary conditions, one capable of adapting to the intricacies of different types of motion estimation with minimal or no supervision. Such a framework would bridge the current divide, offering a versatile solution that extends beyond highly tailored scenarios. This motivates our pursuit of a robust generalisation, designed to flexibly accommodate the specific motion characteristics inherent to each medical imaging task, while maintaining broad applicability across a variety of contexts.

3.3 Framework for auto-adaptive boundary conditions

We address the longstanding challenge of boundary conditions in image registration through the introduction of an auto-adaptive local boundary condition methodology. This approach dynamically adjusts to the distinct characteristics of each registration task, thereby ensuring superior accuracy and robustness in registration tasks. By customizing the boundary conditions to meet the specific demands of each registration scenario, we can significantly enhance the alignment and integration of diverse image datasets. This advancement paves the way for more precise and dependable results across a wide array of applications.

The proposed method is applicable to most registration algorithm though we focus here on variational solutions, a deterministic class of methods that allows to precisely determine the impact of the proposed boundary conditions. This class of registration algorithm is chosen as it is fast and requires relatively few hyper-parameters. In this framework, the registration task is formulated as an optimisation problem involving an energy functional. This functional typically consists of two components: a data fidelity term D, which ensures that the registered image aligns accurately with the fixed image, and a regularisation term S, which promotes the physical plausibility and smoothness of the solution:

$$E(\mathbf{T}) = \int_{\Omega} D(I_f, \mathbf{T}(I_m)) + S(\mathbf{T}) d\mathbf{x}. \tag{3.3.1}$$

By balancing these terms, the method effectively aligns images while maintaining realistic and coherent transformations, thus enhancing the robustness and reliability of the registration process.

Variational methods have gained widespread use, highlighted in the previous literature review, due to their modular nature. This modularity empowers these methods to address diverse regularisation challenges, balancing the imposition of physical principles with the fidelity to observational data. Such flexibility is instrumental in fields where precise modeling of complex systems is paramount,

ensuring robust and accurate solutions tailored to the unique demands of each problem.

A boundary condition serves a crucial role in the formulation of the variational problem. It can either confine the search space, as seen with the Dirichlet boundary condition, which specifies the values a solution must take on the boundary, or it can add a term to the energy functional, as demonstrated by the Neumann boundary condition, which imposes conditions on the shear tensor of the transformation along the boundary. By doing so, Neumann conditions add complexity to the energy functional, guiding the solution to respect the physical laws modeled by these boundary interactions.

We propose a flexible local Robin-type boundary condition that adapts based on image information. This approach allows us to seamlessly transition between the commonly used homogeneous Neumann and homogeneous Dirichlet boundary conditions, and their non-homogeneous counterparts, by adjusting specific hyper-parameters. To tailor the boundary conditions to the unique features of each image, we introduce two methods for calculating an incoming/outgoing flow field map. This map helps guide the selection of appropriate boundary conditions at a local level. Additionally, we implement a hyper-parameter search process driven by the registration energy, which automatically optimises the two key hyper-parameters in our boundary condition model.

3.3.1 Mathematical formulation of locally adapted boundary conditions

To effectively address the issue of mis-registration at the boundary, which arises due to inappropriate boundary conditions, we implement a sophisticated Robin-type boundary condition:

$$A_s(\nabla T_s \cdot \mathbf{n}) + B_s T_s = C_s, \tag{3.3.2}$$

where $s \in \{x, y, z\}$, $\mathbf{T} = [T_x, T_y, T_z]$, $\mathbf{A} = [A_x, A_y, A_z], \mathbf{B} = [B_x, B_y, B_z], \mathbf{C} = [C_x, C_y, C_z]$ are vectors such that $\mathbf{A}$ and $\mathbf{B}$ have positive components, and $\mathbf{n}$ is the normal to the boundary $\delta\Omega$ of the image.

This approach, named after the mathematician Victor Gustave Robin, elegantly combines both Dirichlet and Neumann conditions, allowing for a more accurate and stable solution to boundary value problems. The trade-off is characterised by the vecors $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$. By incorporating this method, we can significantly enhance the precision of our computational models, ensuring that boundary behaviors are accurately represented and reducing the likelihood of errors caused by ill-designed constraints.

Our approach is further enhanced with an automatically computed flow field map, which dynamically guides the registration process in harmony with the specific characteristics of the image. This method ensures that the boundary conditions are tailored to the unique details of each image. By leveraging the flow field map's dynamic adjustments, we can significantly improve the overall accuracy and reliability of the registration. This precision ensures that every nuance of the image is considered, resulting in a highly refined and dependable boundary representation.

It is possible to recover the most prevalent boundary conditions in image registration, specifically the homogeneous Dirichlet and homogeneous Neumann conditions, by making selections for the vectors $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$:

- Homogeneous Dirichlet boundary conditions are achieved with $\mathbf{A} = \mathbf{0}, \mathbf{B} = \mathbf{I}$ and $\mathbf{C} = \mathbf{0}$:
$$T_s = 0. \tag{3.3.3}$$
- Homogeneous Neumann boundary conditions are similarly achieved with $\mathbf{A} = \mathbf{1}, \mathbf{B} = \mathbf{0}$ and $\mathbf{C} = \mathbf{0}$:
$$\nabla T_s \cdot \mathbf{n} = 0. \tag{3.3.4}$$

The non-homogeneous counterparts are achievable for non-null $\mathbf{C}$ vectors.

To streamline the complexity of our model, we propose a strategic reduction in the number of hyper-parameters. This approach will not only simplify the model but also enhance its efficiency and balance. By focusing on essential hyper-parameters and eliminating redundant or less impactful ones, we can achieve a more streamlined and optimised solution:

$$\mathbf{A} = \mathbf{1} - \mathbf{B}. \tag{3.3.5}$$

Furthermore, to significantly enhance the precision of our model, we integrate incoming/outgoing flow field information into the two remaining hyper-parameters. This refined approach ensures that we comprehensively account for the unique characteristics of the image. We incorporate an incoming/outgoing flow field map, denoted as $\mathbf{g}$, directly into the boundary condition, such that:

$$\beta(g_s)(\nabla \mathbf{T}_s \cdot \mathbf{n}) + (1 - \beta(g_s))\mathbf{T}_s = \gamma g_s \tag{3.3.6}$$

where γ is a scalar hyper-parameter and β (Fig 3.5) is such that:

$$\beta(x) = \begin{cases} 1 - tanh(a - |x|) & \text{if } |x| < a \\ 1 & \text{otherwise.} \end{cases} \tag{3.3.7}$$

for any $x \in \mathbb{R}$. a is thus the second hyper-parameter of the proposed local boundary condition.

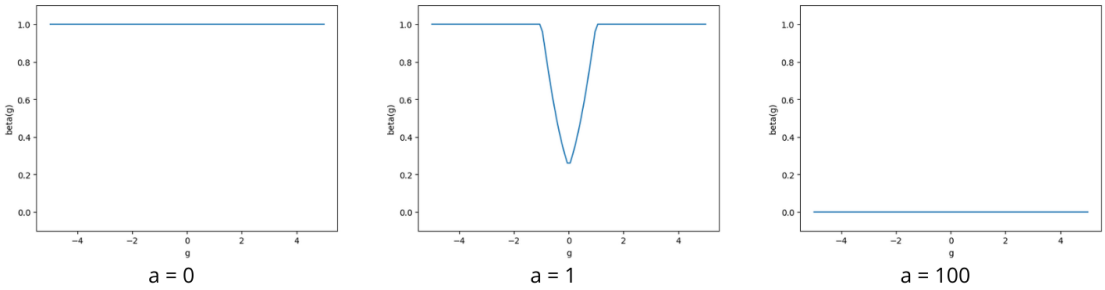


Figure 3.5: β function (eq. 3.3.7) describing the balance between Dirichlet ($\beta := 0$) and Neumann ($\beta := 1$) boundary conditions for different values of a.

The map **g** serves as an advanced guiding framework, fine-tuning the boundary conditions to precisely match the specific flow field dynamics observed within the image. By implementing this approach, we ensure that our boundary conditions are not only adaptive but also acutely responsive to the details present in the image. This method guarantees that our model achieves accuracy and reliability, yielding precise results that reflect the complexity of the observed motion.

3.3.2 Incoming/outgoing flow field detection

In this section, we introduce two innovative methods for detecting incoming/outgoing flow field. The first method offers a scalar detection approach by leveraging the mean absolute error between the moving image, unregistered, and the moving image, registered with an initial motion estimate. The second method enhances the detection process by providing detailed, vector field information, that captures the discrepancies between the forward and backward transformation fields. This approach delivers a comprehensive representation of flow field dynamics, offering nuanced insights into the intricate variations within the transformation fields.

Scalar-valued flow field estimation: Mean absolute error (MAE) is a commonly utilised evaluation metric in the field of image registration due to its simplicity and effectiveness in many scenarios. However, its performance tends to degrade significantly when confronted with severe distortion types and in multi-modality settings, where images from different modalities must be aligned. In such cases, MAE often fails to capture the nuanced differences and complex relationships between the images, leading to suboptimal registration results [38]. It is based on pixel statistics as follows:

$$MAE(\mathbf{r}, I_f, I_m) = \frac{\sum_{\mathbf{r}\in\Gamma} |I_f(\mathbf{r}) - T(I_m(\mathbf{r}))|}{L}, \tag{3.3.8}$$

where L is the number of pixel/voxel in the patch Γ centered around $\mathbf{r}$.

In scenarios involving incoming/outgoing flow field at the boundary, a voxel may be registered with missing information from beyond the image boundary. This missing data is typically substituted with zeroes, resulting in a conspicuous discrepancy between the current image and the moving image, as depicted in Figure 3.6. This discrepancy can be effectively detected using MAE. While MAE values do not convey specifics about the direction of the flow field, they can still be employed to identify the presence of incoming/outgoing flow field at the boundary. We hypothesize that the larger the difference, the more voxels were registered using outside information and thus, the larger the flow field.

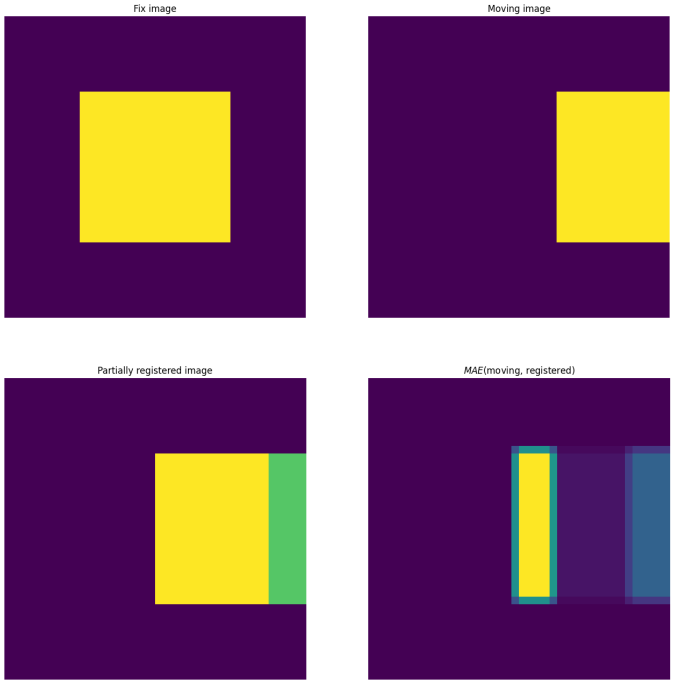


Figure 3.6: Computation of the scalar-valued flow field estimation on a translated square. The difference in intensities within the partially registered image are exacerbated for demonstration purposes: the voxels on the right are registered with information from outside the field of view which is commonly replaced by zero, therefore creating a local difference between the current image and the moving image. It is this difference that we interpret as flux and that is detected by the MAE. The image is only partially registered similarly to our implementation for time efficiency.

Since the MAE is computed between the registered and the moving image, it always measure flow field between information from the same modality. It can thus be used even in multi-modal tasks.

This first approximation of the flow field across the boundaries is advantageous in its simplicity. It is a quick computation, only necessitating the fix image and the moving image registered with the current estimation of the motion field. In most variational framework, both those variables are readily available.

Vector-valued flow field estimation: Registration is a symmetric task. That is, switching the fix and the moving image should lead to a transformation field $\mathbf{T}_{inv}$ such that:

$$\mathbf{T} = (\mathbf{T}_{inv})^{-1} \tag{3.3.9}$$

However, due to numerical approximations, problem formulation and incoming/outgoing flow field at the boundary, this relation is not always verified. Under the assumption that the contributions from numerical approximations and problem formulation are negligible with respect to discrepancies due to incoming/outgoing flow fields, we interpret inverse inconsistency as incoming/outgoing flow fields:

$$InvC(\mathbf{r}) = \frac{\mathbf{T} + \mathbf{T}_{inv}(\mathbf{T}) + \mathbf{T}_{inv} + \mathbf{T}(\mathbf{T}_{inv})}{2} \tag{3.3.10}$$

In fact, we average the discrepancies between $\mathbf{T}$ and $\mathbf{T}_{inv}$ and between $\mathbf{T}_{inv}$ and $\mathbf{T}$.

This method involves the computation of both forward and backward motion estimations, each of which is computationally intensive due to its iterative nature. As a result, to manage computational resources efficiently, the estimates are limited to a reduced number of iterations—specifically, only 10 iterations are performed in this study. This reduction is a strategic choice to balance the trade-off between computational expense and the accuracy of motion estimations, ensuring that the process remains feasible while still providing valuable insights.

3.3.3 Automatic hyper-parametrisation

The proposed boundary condition involves two critical hyper-parameters that must be precisely defined to optimise performance: the parameter a, which dictates the specific form of the function β, and the parameter γ, which balances the contribution of the source term. To tailor these parameters effectively for the given task, we employ an adaptive strategy that identifies the optimal combination by minimising the DIR energy (equation 3.3.1). This approach ensures that the boundary condition is finely tuned to reduce discrepancies and enhance the accuracy

of the registration process.

In our approach, we conduct a grid-search to optimise the boundary conditions by systematically varying the hyper-parameters γ and a. Specifically, γ is explored across a continuous range from 0 to 1, and a is varied from 0 to 100. This range effectively spans the spectrum of boundary conditions from Neumann to Dirichlet types. To enhance the accuracy and relevance of our results, we compute the energy over the entire image or, when available, focus exclusively on groundtruth elements such as landmarks or segmentation masks. This tailored approach ensures that our evaluation metric closely aligns with the specific features and constraints of the task at hand.

3.3.4 Numerical implementation

Integrating the proposed local boundary conditions within a pre-existing registration algortithm requires 2 additional steps.

Initially, it is essential to compute the incoming/outgoing flow fields map, which provides local guidance for the boundary conditions, by performing a preliminary motion estimation using homogeneous Neumann boundary conditions. In parallel, the vectorial guidance based on inverse consistency necessitates the estimation of the inverse transformation field, as detailed in Algorithm 2. While this step is optional, initialising the registration algorithm with this inverse estimate has been shown to enhance performance. This approach leverages a slightly imperfect starting point to provide momentum for the algorithm, thereby facilitating a more effective search for the accurate transformation field.

Evaluating the incoming/outgoing flow field can be performed in various ways depending on the specific requirements of the algorithm and the computational resources available. It may be executed just once at the beginning of the process, during each iteration of the algorithm, or at every resolution level in a multi-resolution scheme, as suggested in Algorithm 1 and visualised in Figure 3.7. The choice between these approaches involves a trade-off between computational efficiency and accuracy. Performing the step once can save computation time but may not capture all necessary details, while executing it at each iteration or resolution step provides more detailed results at the cost of increased computational demand.

Second, the application of boundary conditions to the motion estimate is crucial. In the proposed method, this is executed explicitly to ensure that the boundary conditions are rigorously applied and verified. From a numerical perspective, this new boundary condition is implemented by adjusting the boundary condition to fit the discrete grid of the image, thereby ensuring that the computational representation accurately reflects the physical constraints imposed by the boundary conditions. This approach guarantees that the boundary conditions are properly enforced throughout the domain, thus enhancing the precision and reliability of

Algorithm 1: Primary framework for registration

Input : $(I_f, I_l, \mathbf{T}_{init})$ = (fix image, moving image, initial condition)
Output : $\mathbf{T}$ = motion field
$\mathbf{T} \leftarrow$ Resample $\mathbf{T}_{init}$ to $(16 \times 16 \times 16)$
for $res \in \{(16 \times 16 \times 16), (32 \times 32 \times 32), ..., (256 \times 256 \times 256)\}$ **do**
  $I'_f \leftarrow$ Resample I_f to res
  $I'_m \leftarrow$ Resample I_m to res
  $\mathbf{T} \leftarrow$ Resample $\mathbf{T}$ to res
  $\mathbf{g}, \mathbf{T} \leftarrow$ Algorithm 2$(I'_f, I'_m, \mathbf{T})$
  $\mathbf{T} \leftarrow$ Algorithm 3$(I'_f, I'_m, \mathbf{T}, \mathbf{g})$
end

Algorithm 2: Calculation of incoming/outgoing flow field in the boundary

Input : $(I_f, I_m, \mathbf{T})$ = (fix image, moving image, motion field)
Output : $\mathbf{g}$ = Voxelwise in/out flow field in the boundary
$\mathbf{T}_{inv}$ = motion field
$D \leftarrow$ data fidelity term between I_f and I_m
$\mathbf{T}_{inv} \leftarrow \mathbf{T}$
$I^*_m \leftarrow I_m$ transformed with $\mathbf{T}$
$i \leftarrow 0$
while $i < 10$ **do**
  update $\mathbf{T}$ boundaries according to homogeneous Neumann boundary conditions
  $\mathbf{T} \leftarrow \mathcal{L}(I_f, I^*_m, D, \mathbf{T})$ where $\mathcal{L}$ is one update in the variational DIR algorithm
  $i \leftarrow i + 1$
end
$i \leftarrow 0$
while $i < 10$ **do**
  update $\mathbf{T}_{inv}$ boundaries according to homogeneous Neumann boundary conditions
  $\mathbf{T}_{inv} \leftarrow \mathcal{L}(I^*_m, I_f, D, \mathbf{T}_{inv})$
  $i \leftarrow i + 1$
end
Update $\mathbf{T}_{inv}$ boundaries according to homogeneous Neumann boundary conditions
$\mathbf{g} \leftarrow$ guidance metric from $\mathbf{T}$ and $\mathbf{T}_{inv}$

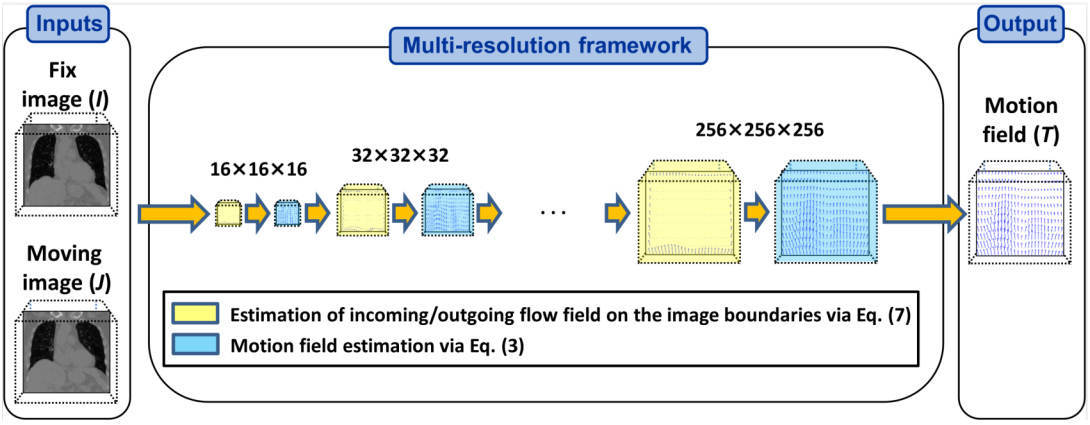


Figure 3.7: Generic multi-resolution scheme including the proposed framework for adaptable boundary conditions. The fix and moving image are used at various resolution. The resulting motion estimate are used as the starting point of the registration process (blue blocks) at the subsequent higher resolution. This process is further detailed in Algorithm 1. Furthermore, the incoming/outgoing flow field maps are updated (yellow blocks) at the begining of each resolution step for a balance between precision and efficiency. Typical inputs I and J of size 256x256x256 are used to illustrate the various resolution stages.

the motion estimate.

To apply the constraint in Equation 3.3.6, we first need to compute $\nabla T_s \cdot \mathbf{n}$, where $s \in \{x, y, z\}$. To do so, we make the following second order approximation:

$$\nabla T_s^0 \cdot \mathbf{n} = -\frac{-3T_s^0 + 4T_s^1 - T_s^2}{2}, \tag{3.3.11}$$

and on the other side of the boundary:

$$\nabla T_s^N \cdot \mathbf{n} = \frac{T_s^{N-2} - 4T_s^{N-1} + 3T_s^N}{2}, \tag{3.3.12}$$

where N is the number of voxels along the direction considered.

Substituting in equation 3.3.6 and solving for T_s^0 and for T_s^N:

$$T_s^0 = \frac{\beta_s^0(4T_s^1 - T_s^2) + 2\gamma g_s^0}{\beta_s^0 + 2}, \tag{3.3.13}$$

$$T_s^N = \frac{\beta_i^N(4T_s^{N-1} - T_s^{N-2}) + 2\gamma g_s^N}{\beta_s^N + 2}. \tag{3.3.14}$$

.

Again, we may recover homogeneous Dirichlet boundary condition for $\beta = 0$ and $\gamma = 0$:

$$T_s^0 = 0, \tag{3.3.15}$$

$$T_s^N = 0, \tag{3.3.16}$$

and homogeneous Neumann boundary condition for $\beta = 1$ and $\gamma = 0$:

$$T_s^0 = \frac{4T_s^1 - T_s^2}{3}, \tag{3.3.17}$$

$$T_s^N = \frac{4T_s^{N-1} - T_s^{N-2}}{3}. \tag{3.3.18}$$

Finally, the motion field is estimated with the registration algorithm. An overview of the steps included in such a procedure is proposed in Algorithm 3.

Algorithm 3: Calculation of the motion field estimate

Input : $(I_f, I_m, \mathbf{T}, \mathbf{g})$ = (fix image, moving image, motion field, incoming/outgoing flow)
Output : $\mathbf{T}$ = motion field
$i \leftarrow 0$
while $i < it$ **do**
  $I_m^* \leftarrow I_m$ transformed with $\mathbf{T}$
  $D \leftarrow$ data fidelity term between I_f and I_m^*
  update $\mathbf{T}$ boundaries according to the boundary condition and $\mathbf{g}$
  $\mathbf{T} \leftarrow \mathcal{L}(I_f, I_m^*, D, \mathbf{T})$
  update $\mathbf{T}$ boundaries according to the boundary condition and $\mathbf{g}$
  $i \leftarrow i + 1$
end

3.4 Validation with a variational method

3.4.1 EVolution: a multi-modal variational method

To rigorously evaluate the effectiveness of the proposed method, we integrate it into the established variational algorithm known as EVolution [86]. This approach allows us to leverage the robust framework of EVolution, ensuring a comprehensive and precise assessment of our method's performance. By doing so, we can observe its behavior and efficacy within a proven and reliable algorithmic structure, providing clear insights and valuable comparative data.

Variational algorithms present three significant advantages within the context of this study. Firstly, their deterministic nature ensures that identical inputs consistently yield the same outputs, offering a level of predictability and reproducibility not commonly found in deep learning methods. Secondly, these algorithms require minimal hyper-parameters beyond those intrinsic to our method, simplifying the optimisation process and reducing computational overhead associated with extensive hyper-parameter tuning. Lastly, despite involving iterative procedures, variational algorithms demonstrate commendable speed and efficiency, making them an optimal choice for rapid yet accurate computations.

EVolution is a multi-modal DIR algorithm employing Tikhonov regularisation. This ensures smooth and consistent transformations. The algorithm's data fidelity term, introduced in section 3.2.3, is designed to enhance edge alignment between the registered images, thereby achieving superior accuracy and clarity in image registration.

Hyper-parameters: The method comes with 2 hyper-parameters.

The first parameter, α, plays a pivotal role in determining the balance between the data fidelity term and the regularisation within the energy functional that is being minimised:

$$\mathbb{E}(\mathbf{T}) = \int_{\Omega} D(\mathbf{T}) + \frac{\alpha}{2}(||\nabla T_x||_2^2 + ||\nabla T_y||_2^2 + \nabla T_z||_2^2)\mathbf{dr}, \tag{3.4.1}$$

with $\mathbf{T} = (T_x, T_y, T_z)$.

By adjusting α, one can control the trade-off between fitting the data accurately and maintaining a smooth, regularised solution. This careful balancing act ensures that the model does not overfit while still capturing essential patterns in the data.

Overall, the EVolution algorithm demonstrates considerable robustness to variations in the parameter α. This resilience ensures that minor adjustments in α do not significantly impact the algorithm's performance or outcomes. Consequently, for the purpose of this study, we will consistently employ a value of 0.5 for α, striking a balance between regularisation and data fidelity that has proven effective in preliminary testing.

The second critical hyper-parameter to consider is the patch size employed in the computation of the data fidelity term. This parameter significantly enhances the algorithm's robustness to structural variations in the data, simultaneously providing implicit regularisation. While a smaller patch size can markedly improve the quality of registration by capturing finer details, it comes with a trade-off. If the patch size is reduced excessively, it may introduce numerical instabilities that could compromise the overall stability of the computation. For this work, a patch size of 5x5x5 has been selected, balancing the need for high registration accuracy with the necessity of maintaining computational stability.

Numerically solving the optimisation problem: To obtain an iterative scheme for finding $\mathbf{T}$, we first apply the Euler-Lagrange equation to the energy (Eq. 3.4.1) in every dimension $s \in \{x, y, z\}$:

$$\frac{\delta D(\mathbf{T})}{\delta T_s} - \nabla \cdot (\frac{2\alpha}{2} \nabla T_s) = \frac{\delta D(\mathbf{T})}{\delta T_s} - \alpha \Delta T_s = 0 \tag{3.4.2}$$

Let us now introduce a time variable such that $\frac{dT_s}{dt} = 0$:

$$\frac{dT_s}{dt} = \alpha \Delta T_s - \frac{\delta D(\mathbf{T})}{\delta T_s}. \tag{3.4.3}$$

Using an explicit fixed-point scheme, we obtain:

$$\frac{T_s^{t+1} - T_s^t}{dt} = \alpha \Delta T_s - \frac{\delta D(\mathbf{T})}{dT_s}, \tag{3.4.4}$$

where the superscript correspond to the time index.

Solving for T_s^{t+1}:

$$T_s^{t+1} = T_s^t + dt(\alpha \Delta T_s - \frac{\delta D(\mathbf{T})}{\delta T_s}). \tag{3.4.5}$$

The derivatives of the data fidelity term are computed as follows:

$$\frac{\delta D(\mathbf{T})}{\delta s} = \frac{D(S_s^+ \cdot \mathbf{T}) - D(S_s^- \cdot \mathbf{T})}{2}, \tag{3.4.6}$$

with S_s^+ the forward shifted transformation in the s direction and S_s^- the backward shifted transformation.

Moreover, considering that the variational equation 3.4.1 is only applicable for small displacements, a multi-resolution scheme is employed to effectively capture movement at different scales (see Fig. 3.7). This approach enables the algorithm to identify global motions at lower resolutions while resolving finer movements at higher resolutions. The motion estimate calculated at each resolution level is subsequently used as the initial point for the optimisation process at the next, finer resolution. This hierarchical method ensures a comprehensive and accurate motion analysis across various scales, enhancing the robustness and precision of the overall motion estimation.

3.4.2 Hardware and implementation

The proposed method is implemented under the Python language with the Cupy library to allow for the use of GPUs. The following results are obtained on a 24-core AMD Zen2 EPYC 7402 CPU paired with an Nvidia A100 GPU with 16 GB of memory from the Plafrim cluster.

3.4.3 Assessment on a mono-modal task

The DIR-Lab dataset: Our initial focus is on a mono-modal image registration task utilising the DIR-Lab dataset, which comprises 10 pairs of lung CT scans from patients undergoing treatment for esophageal cancer. Each scan pair captures the lung at two extreme phases of respiration: maximal inhalation and maximal exhalation. This setup introduces substantial diaphragm-induced movement, presenting unique challenges. We focus on the intricate issue of motion discontinuities that manifest at the interface between the lung and the rib cage. This area is particularly sensitive to the diaphragm's movements, which can induce abrupt alterations in both the shape and position of the lungs. Such dynamic changes cause notable deformation in delicate structures, including small airways and blood vessels, which are susceptible to significant distortion from the respiratory process. These distortions present considerable challenges in achieving precise alignment during registration procedures. Moreover, variations in image contrast are frequently observed due to the compressive forces on the lungs during inhalation and exhalation, which affect the clarity and accuracy of the images captured [86].

The primary characteristics of the CBCT scans and their corresponding ground truth are detailed in Table 3.1. Notably, the magnitude of displacement caused by breathing motion varies significantly between patients, introducing a broad spectrum of complexity across the dataset. This variability is particularly advantageous, as it presents a diverse range of challenges, enabling a comprehensive evaluation of the proposed method. By confronting these varying levels of difficulty, the robustness and adaptability of our approach can be thoroughly assessed, ensuring that it is capable of handling real-world scenarios with fluctuating and unpredictable patient-specific respiratory motion.

Patient	Dimension	Spacing (mm)	Displacement (mm)	Observer uncertainty (mm)
1	256x256x94	0.97x0.97x2.5	4.01 ± 2.91	0.85±1.24
2	256x256x112	1.16x1.16x2.5	4.65 ± 4.09	0.70±0.99
3	256x256x104	1.15x1.15x2.5	6.73 ±4.21	0.77±1.01
4	256x256x99	1.13x1.13x2.5	9.42± 4.81	1.13±1.27
5	256x256x106	1.10x1.10x2.5	7.10 ± 5.14	0.92±1.16
6	512x512x128	0.97x0.97x2.5	11.10±6.98	0.97±1.38
7	512x512x136	0.97x0.97x2.5	11.59±7.78	0.8±1.32
8	512x512x128	0.97x0.97x2.5	15.16±9.11	1.03 ± 2.19
9	512x512x128	0.97x0.97x2.5	7.82±3.99	0.75±1.09
10	512x512x120	0.97x0.97x2.5	7.63±6.54	0.86 ± 1.45

Table 3.1: Data characteristics for the DIR-Lab dataset. The wide range of displacement poses an interesting challenge for the registration method.

The preprocessing steps involve two key operations: scaling the intensity values to a normalised range between 0 and 1, and cropping any artificial padding. The

intensity scaling standardizes the data, facilitating consistent input across different scans, while the cropping refines the dataset by preserving only the meaningful regions of interest. This ensures that the boundaries of the scan precisely align with the actual field of view, eliminating extraneous elements that could interfere with the registration task.

To thoroughly evaluate the registration algorithm's performance for scientific and clinical applications, we utilised a comprehensive ground-truth comprising 3,000 expertly placed landmarks. These landmarks are integral for assessing the algorithm's accuracy in real-world scenarios. Lung CT scans were selected due to their high contrast and rich anatomical details, including critical landmarks such as vessels and bronchial bifurcations.

The manual registration process, for producing the landmark pairs, was conducted using the Matlab-based software, Assisted Point Registration of Internal Landmarks (APRIL). Initially, a primary expert manually placed landmarks across the entire lung image (Fig 3.8). To evaluate intra-reader reproducibility, the same expert re-placed a subset of these landmarks. Similarly, extra-reader reproducibility was assessed by having two additional experts independently place a subset of landmarks. The results showed intra-reader reproducibility ranging from 0.61 to 1.11 mm and extra-reader reproducibility from 0.74 to 1.14 mm. The landmarks were strategically distributed throughout the lung volume to effectively capture the complex tissue motions induced by breathing.

Assessment: The evaluation of registration solutions is conducted by comparing them against the ground truth data provided. This process involves several key steps to ensure accuracy and reliability.

Each registration result is assessed by aligning it with the ground truth data, which serves as the benchmark for correctness. This comparison helps in quantifying how closely the registration results match the expected outcomes. In general, various metrics are used to measure the performance of the registration algorithms, depending on the ground truth available. These metrics provide a quantitative measure of the accuracy and quality of the registration.

In addition to quantitative metrics, visual inspection is performed to assess the quality of the registration. This involves overlaying the registered images with the ground truth to visually inspect alignment and detect any discrepancies.

The ground truth provided being under the form of landmarks, we compute a Target Registration Error (TRE) as the Euclidean distance between the moving landmarks, registered with the motion estimate, and the fix landmarks:

$$TRE(\mathbf{T}) = ||\mathbf{r}_f + (T_x(\mathbf{r}_f), T_y(\mathbf{r}_f), T_z(\mathbf{r}_f)) - \mathbf{r}_m||_2, \tag{3.4.7}$$

where $\mathbf{r}_f$ are the landmark coordinates on the fix image and $\mathbf{r}_m$ are the landmark

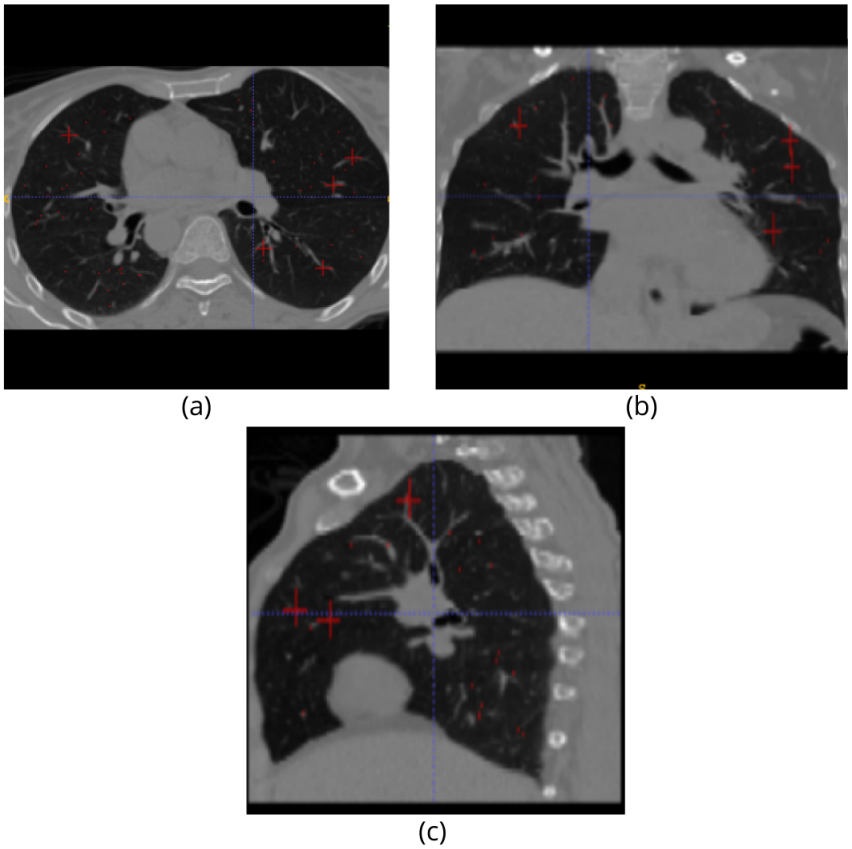


Figure 3.8: Slices of patient 1 from the DIR-lab dataset with corresponding landmarks in red. (a) is the transversal plane, (b) the coronal plane and (c) the sagittal plane.

coordinates on the moving image.

The TRE effectively reflects the accuracy of the registration by quantifying the alignment error between corresponding landmarks in the moving and fixed images, ensuring that the spatial transformations applied are accurate and reliable.

To locally assess the impact of the proposed boundary conditions, the error data at individual landmarks is interpolated using trilinear interpolation. This method maps the discrete error measurements onto a continuous, regular grid, thereby enhancing the spatial resolution and accuracy of the error representation. This technique helps in capturing the nuances of the error variation within the volumetric space.

For effective visualisation, a mean intensity projection of the interpolated error map is generated. This projection aggregates the error values to create a 2D representation, facilitating interpretation and analysis of the error distribution

Table 3.2: Mean TRE (mm) obtained for each case of the DIR-Lab data set (mono-modal CT thorax registration) and for each tested boundary conditions, with mean and standard deviation (std). Best scores are highlighted with bold characters.

Case	Homogeneous	Homogeneous	Inverse Consistency		MAE	
	Neumann	Dirichlet	Auto	Best	Auto	Best
1	**1.05**	1.06	**1.05**	**1.05**	1.06	**1.05**
2	1.02	1.02	1.02	**1.01**	1.02	1.02
3	**1.20**	1.23	**1.20**	**1.20**	**1.20**	**1.20**
4	1.43	**1.42**	**1.42**	**1.42**	**1.42**	**1.42**
5	1.57	1.56	**1.51**	**1.51**	1.56	1.56
6	1.49	1.53	1.42	**1.41**	1.46	1.49
7	1.67	1.68	**1.47**	**1.47**	1.67	1.67
8	4.56	4.61	4.39	**3.83**	4.56	4.56
9	1.26	1.26	**1.24**	**1.24**	**1.24**	**1.24**
10	1.71	1.73	1.71	1.71	1.71	**1.70**
Mean ± std	1.70 ± 1.04	1.71 ± 1.05	1.64 ± 0.99	**1.59 ± 0.82**	1.70 ± 1.03	1.69 ± 1.04

across different slices or sections of the model.

This approach ensures that the localisation of errors due to boundary conditions is both detailed and visually accessible, providing a clear understanding of their spatial impact.

Results: To benchmark the method's performance, we compared the proposed adaptive local boundary conditions against two widely utilised global boundary conditions: homogeneous Neumann and homogeneous Dirichlet. Each boundary condition variant was assessed using two different results.

Optimal parameterisation: This result reflects the configuration that minimises the registration energy, as detailed in the automatic hyper-parameterisation section. This configuration aims to automatically optimise the registration process by fine-tuning the parameters to achieve the best fit.

Minimal achievable error: This result is obtained from identifying the configuration that results in the lowest possible registration error during the grid-search. This approach allows to verify the validy of the fully optimised framework for the determination of hyper-parameters.

Among the widely used global boundary conditions, homogeneous Neumann boundary conditions tend to yield superior results (Tab. 3.2). This advantage arises from the inherently flexible formulation, which imposes constraints on the normal shear tensor of the transformation rather than directly on the transformation itself. By focusing on the rate of change normal to the boundary, Neumann conditions allow for a more adaptable and nuanced approach to managing boundary constraints, enhancing their effectiveness compared to homogeneous Dirichlet.

The inverse consistency guidance demonstrates superior performance in this

context. Specifically, the regularisation achieved through DIR energy minimisation results in significantly lower errors compared to global boundary conditions, with a p-value of 0.03 as compared with homogeneous Neumann via a paired t-test, and an average reduction of 4% and reaching up to 12% in case 7. Given that the proposed method exclusively targets the boundary, these findings are promising, particularly for scenarios where the fields of view are even more constrained than those examined in this study. This indicates a robust potential for enhancing accuracy in highly limited observational conditions.

However, the MAE guidance provides results very similar to homogeneous Neumann, the best global boundary condition, with a p-value of 0.30.

This is explained by the sensitivity of each guidance and the level of information. The inverse consistency based flow field map provides vectorial information of larger magnitude than the MAE based map as seen in Figure 3.9 and 3.10.

While the MAE based flow field map successfully detect the upward flow field near the boundary of the lung, the information does not propagate well to the border where it is required to guide the boundary conditions. Indeed, values near the diaphragm, where there is the strongest incoming/outgoing flow fields, are about 1e-3.

On the contrary, the flow field map computed through inverse consistency detects well the incoming/outgoing flow fields at the borders. In the front to back direction (g_u) some movement due to the lung expansion during inhalation are detected in both organs. In the right to left direction (g_v), some movement, located mainly in the outercorners of the lungs are detected for the same reasons. Finally, in the top bottom direction (g_w), large motion is signalled near the diaphragm, as expected.

As anticipated, the primary discrepancies in the transformations estimated using different boundary conditions predominantly manifest near the edges of the image. However, a closer examination of the error maps reveals that these variations extend inward, affecting the regions of interest at the center of the image. This propagation of errors highlights the significant influence that boundary conditions exert on the accuracy of the transformation across the entire image, not just its periphery. Such insights underscore the necessity for meticulous selection and application of boundary conditions to ensure precise and reliable transformations in critical regions.

We observe a significant reduction in the error near the diaphragm when comparing the global boundary conditions (Fig. 3.11 a, b) with the local boundary condition guided by the inverse consistency-based flow field map (Fig. 3.11 c). This improvement underscores the efficacy of the local boundary condition approach in enhancing accuracy in critical regions. By leveraging the inverse consistency-based flow field map, the method adapts more precisely to local anatomical variations,

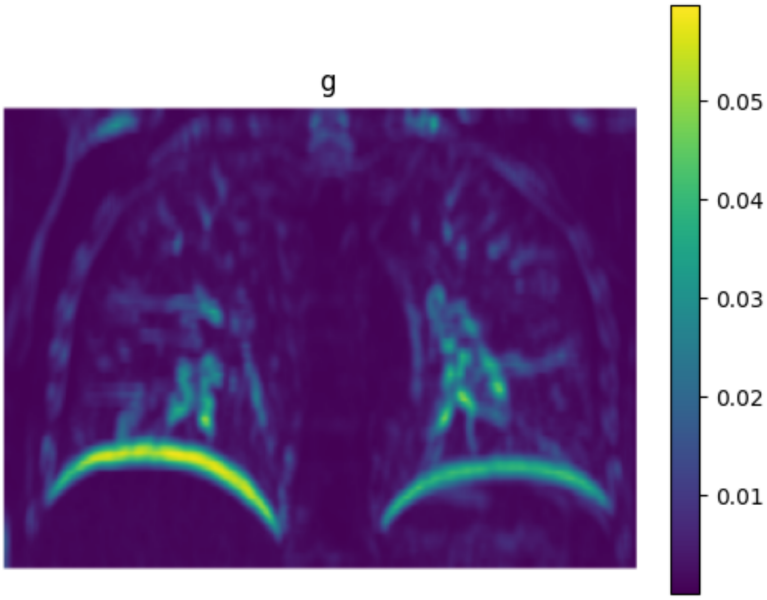


Figure 3.9: A coronal slice of the flow field map, computed using the MAE metric, is presented for Patient 1 from the DIR-Lab dataset. This visualisation showcases the distribution of the flow field across the coronal plane, represented in arbitrary units, allowing a closer examination of the accuracy and performance of this estimation technique. Notably, the map reveals key insights into the flow patterns, highlighting both incoming and outgoing flow field dynamics. However, one significant observation is the method's inability to capture the full extent of the flow near the diaphragm region. This limitation directly correlates with the minimal improvements observed when utilising this approach, particularly compared to homogeneous Neumann boundary conditions, where the method's lack of sensitivity in this area limits its overall effectiveness.

thereby reducing discrepancies and yielding more reliable results in areas of complex movement, such as near the diaphragm. For this patient, the MAE guidance leads back to homogeneous Neumann boundary condition. The two corresponding error maps are thus identical.

As demonstrated in Figure 3.12, the automation of hyper-parameter tuning is achievable for both guidance methods. This approach efficiently aligns the energy minimum calculated from the landmarks with the minimum of the TRE. The close correlation between these two minima underscores the method's effectiveness in optimising parameters and achieving accurate alignment. The streamlined process of hyper-parameter adjustment not only simplifies the calibration but also ensures that the resulting model closely mirrors the optimal TRE, thereby enhancing overall registration performance.

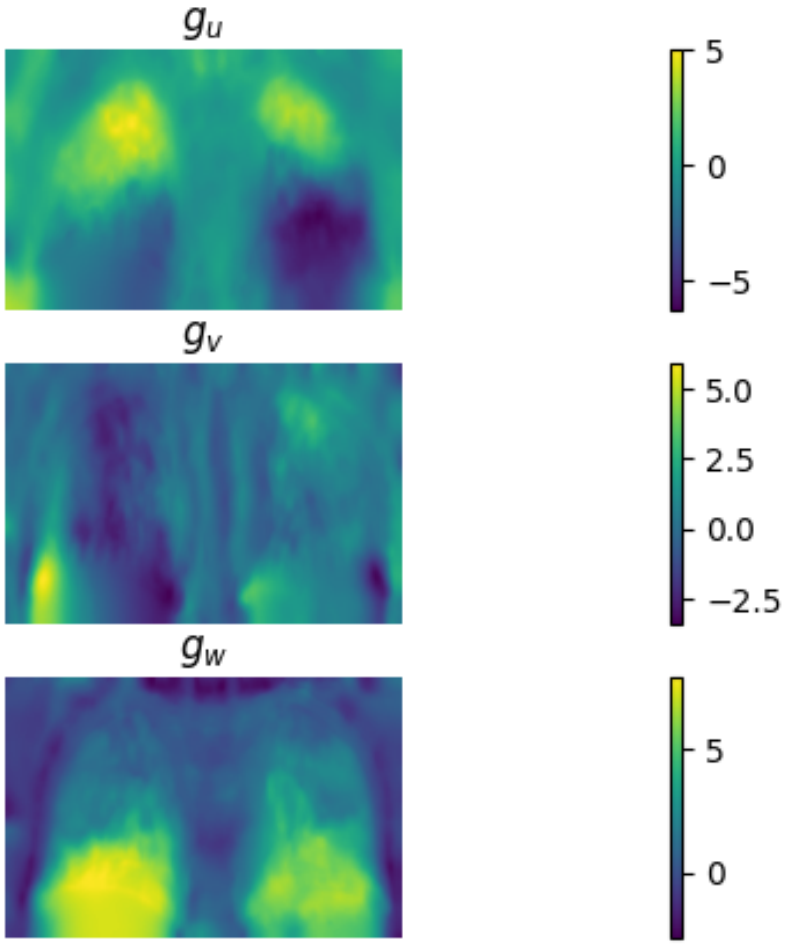


Figure 3.10: A coronal slice of each component of the flow field map, which has been computed using the inverse consistency for Patient 1 from the DIR-Lab dataset where g_u represent the flow field in the in-out direction, g_v in the left-right direction and g_w in the up-down direction, in voxels. This method successfully detects a significant upward motion (g_w) near the diaphragm, due to breathing.

We note that the MAE guidance primarily allows a tradeoff between Dirichlet and Neumann, and not between homogeneous and non-homogeneous boundary conditions. Due to the fact that the so computed flow field map does not contain any information regarding the flow field direction, the source term is not relevant for the registration process. Over this dataset, the improvement is thus negligeable.

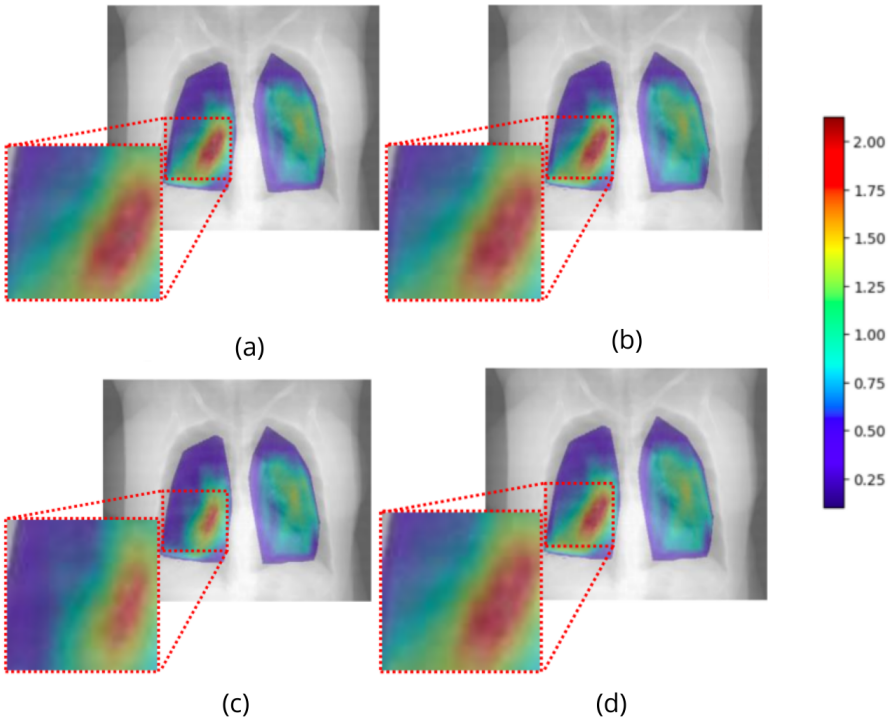


Figure 3.11: The MIP of the error map, which has been interpolated from errors associated with each landmark, is overlaid on the MIP of the fixed image for Case 7 in the DIR-Lab dataset. The detailed views are as follows: (a) shows the error map under homogeneous Neumann boundary conditions, (b) under homogeneous Dirichlet boundary conditions, (c) presents the result of applying inverse consistency guidance, and (d) of applying MAE guidance, both with automatic computation of the hyper-parameters.

3.4.4 Multi-modal task

The Learn2Reg dataset: The method is then evaluated using a multi-modal dataset featured in the Learn2Reg 2021 challenge. This dataset comprises 122 scans, including both MRI and CT modalities, with most scans being unpaired. For a comprehensive assessment of the registration process, the dataset includes segmentation masks of key anatomical structures—namely the liver, spleen, and both kidneys—available for 7 patients. These well-defined segmentation masks are crucial for evaluating the accuracy and effectiveness of the registration technique, ensuring that the method performs robustly across various anatomical regions and imaging conditions.

The imaging data is characterised by a voxel spacing of 2 mm uniformly across all three spatial dimensions, ensuring high-resolution and consistency in the volumetric

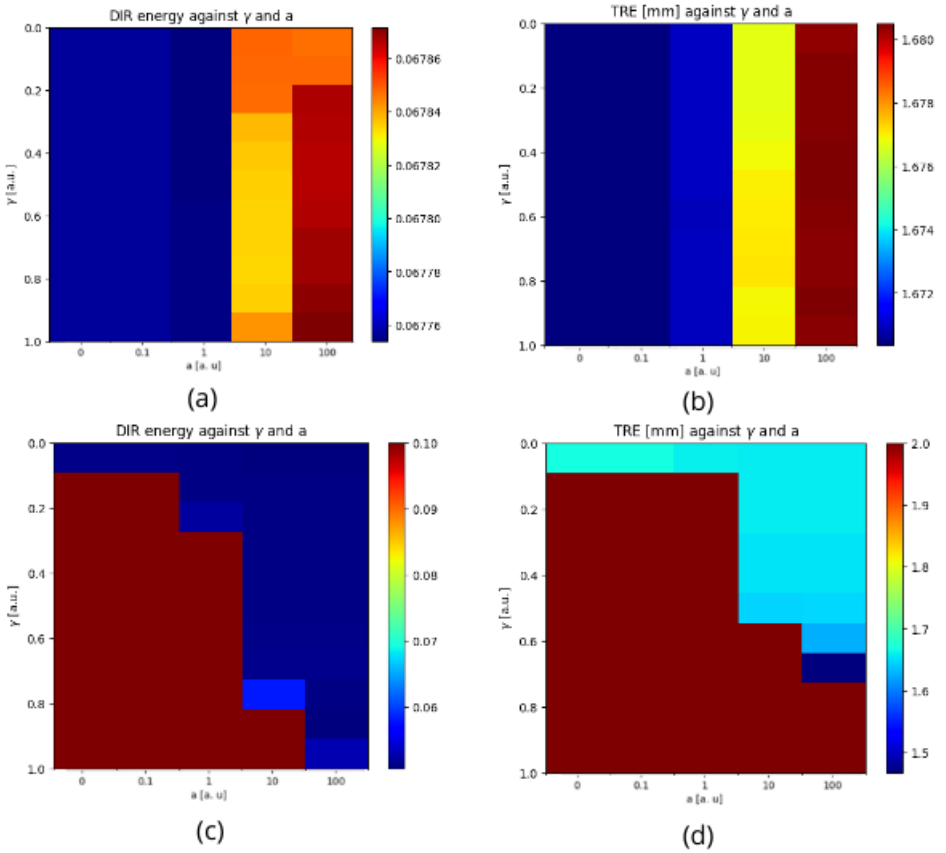


Figure 3.12: The outcomes of the parameter search for Case 7 from the DIR-Lab dataset are presented as follows: (a) displays the DIR energy calculated using MAE guidance, (b) shows the TRE with MAE guidance, (c) represents the DIR energy assessed with inverse consistency guidance, and (d) illustrates the TRE obtained using inverse consistency guidance. This experiment shows that minimising the DIR energy effectively maximises the segmentation quality. In the absence of ground-truth, it can thus be used to optimise the hyper-parameters.

representation. This isotropic voxel size facilitates precise anatomical modeling and accurate image analysis. The detailed dimensions of the scanned volume are provided in the accompanying Table 3.3, which outlines the extent and size of the dataset in each direction, enabling comprehensive spatial understanding. The differences of dimensions between the MRI and the CT lead to an additional preprocessing step compared to the preprocessing pipeline for the DIR-Lab dataset: the largest image is cropped to the smallest image dimension for a successful registration.

Patient	MRI Dimension	CT Dimension
1	$192 \times 159 \times 124$	$181 \times 143 \times 175$
2	$192 \times 160 \times 180$	$177 \times 150 \times 136$
3	$192 \times 130 \times 189$	$192 \times 149 \times 142$
4	$180 \times 125 \times 173$	$192 \times 141 \times 192$
5	$191 \times 146 \times 117$	$192 \times 160 \times 149$
6	$184 \times 160 \times 155$	$192 \times 160 \times 167$
7	$189 \times 158 \times 135$	$192 \times 160 \times 146$

Table 3.3: Data characteristics for the Learn2Reg dataset. The different dimensions between the moving and fix image require an additional cropping step.

Assessment: In this context, the ground truth data is represented as segmentation masks, which serve as the reference standard for evaluating image registrations. To quantitatively assess the similarity between the registered masks and the fixed reference mask, we employ the Dice coefficient:

$$Dice(M_f, \mathbf{T}(M_m)) = \frac{2|M_f \cap \mathbf{T}(M_m)|}{|M_f||\mathbf{T}(M_m)|}, \tag{3.4.8}$$

where M_f is the segmentation mask on the fix image I_f, and M_m is the segmentation mask on the moving image I_m.

This metric is used to gauge the degree of overlap between two binary masks. The Dice coefficient ranges from 0 (no overlap) to 1 (perfect overlap), providing a robust measure of the accuracy of the registration by comparing the overlap of the segmented regions against the ground truth.

For qualitative assessment of the registration accuracy, we overlay the contours of the registered masks onto the contours of the fixed reference masks. This visualisation technique allows us to directly compare the alignment of the segmented regions between the registered and fix masks. By examining these overlaid contours, we can visually assess discrepancies and determine how well the registered masks conform to the fixed masks. This method provides an intuitive understanding of registration performance, revealing areas of alignment and misalignment through clear, visual comparison.

Results: To validate the method's efficacy, we present results aligned with the minimum energy computations, which highlight the successful automation of hyperparameter tuning. Additionally, we showcase the optimal registration outcomes, demonstrating the method's capacity to achieve precise alignments under the

Table 3.4: Dice coefficients (averaged over 4 labels) obtained for each case of the Learn2Reg 2021 data set (multi-modal CT to MRI abdomen registration) and for each tested boundary conditions, with mean and standard deviation (std). Best scores are highlighted with bold characters

Case	Homogeneous Neumann	Homogeneous Dirichlet	Inverse consistency Auto	Inverse consistency Best	MAE Auto	MAE Best
1	0.846	0.793	**0.847**	**0.847**	0.846	0.846
2	**0.695**	0.608	**0.695**	**0.695**	**0.695**	**0.695**
3	0.855	0.818	**0.856**	**0.856**	**0.856**	**0.856**
4	**0.866**	0.760	0.852	**0.866**	0.852	**0.866**
5	0.722	0.556	0.722	**0.756**	0.722	**0.756**
6	**0.810**	0.529	**0.810**	**0.810**	**0.810**	**0.810**
7	**0.849**	0.829	**0.849**	**0.849**	**0.849**	**0.849**
Mean ± std	0.806 ± 0.069	0.699 ± 0.130	0.804 ± 0.068	**0.811 ± 0.064**	0.804 ± 0.068	**0.811 ± 0.064**

given constraints. This thorough examination underscores the robustness and adaptability of the approach in handling diverse and challenging registration tasks.

In this scenario, none of the flow field map yield improvements over the global boundary conditions as seen in Table 3.4. Despite its potential, the hyper-parametrisation frequently converges back to homogeneous Neumann boundary conditions. This indicates that the anticipated enhancements from the locally adaptable boundary conditions approach do not materialize, and the system reverts to the simpler, established boundary conditions, underscoring the fact that Neumann boundary conditions may be the most suitable for the motion exhibited in the abdomen.

To visualise the impact of different boundary conditions on the registration of segmentation masks, we display the contours of the registered masks alongside the fixed mask on a slice of the images. This approach clarifies why the homogeneous Dirichlet boundary condition underperforms: organs such as the liver, as seen in Figure 3.13, are positioned near the image borders. Despite experiencing some movement, the homogeneous Dirichlet condition enforces a null displacement, impeding accurate motion field estimation. Conversely, this visualisation confirms that the three other tested boundary conditions effectively register the organs of interest, demonstrating their superior performance in these scenarios.

As seen in figure 3.14, the DIR energy effectively reflects the registration quality in terms of the ground truth provided, with both the MAE and inverse consistency guidance. This confirms the validity of the automatic hyper-parametrisation.

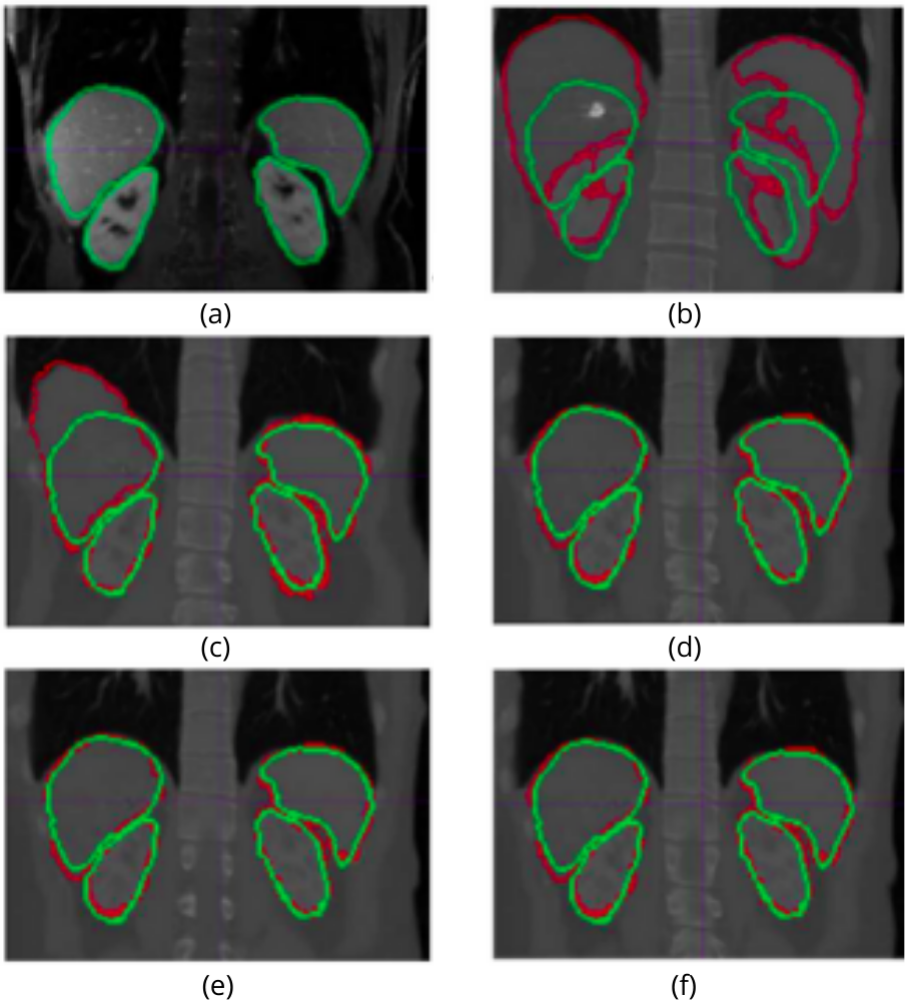


Figure 3.13: The edges of the segmentation masks for Case 4 of the Learn2Reg dataset are displayed on a coronal slice of the corresponding images, as follows: (a) the fixed MRI (the corresponding mask is shown in green in all 5 subsequent images), (b) the moving CT scan, (c) the CT scan registered using homogeneous Dirichlet boundary conditions, (d) using homogeneous Neumann boundary conditions, (e) using inverse consistency guidance, and (f) using MAE guidance. The hyper-parametrisation leading to the minimum energy is used for the last two cases.

3.5 Limitations and perspectives for the auto-adaptive boundary condition framework

The introduced locally adaptable boundary conditions offer significant enhancements over conventional boundary conditions like homogeneous Neumann and

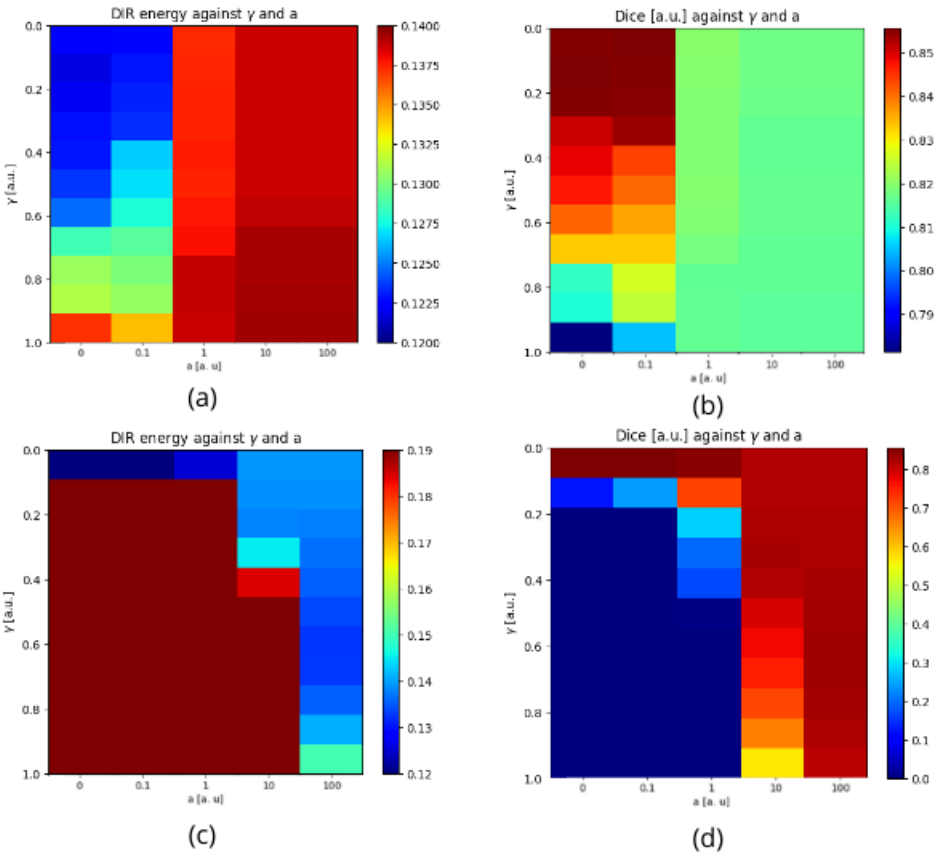


Figure 3.14: The outcomes of the hyper-parameter search for Case 4 from the Learn2Reg dataset are presented as follows: a. displays the DIR energy calculated using MAE guidance, b. shows the TRE with MAE guidance, c. represents the DIR energy assessed with inverse consistency guidance, and d. illustrates the TRE obtained using inverse consistency guidance.

homogeneous Dirichlet. These new conditions dynamically adjust to local image characteristics, potentially offering better performance for specific scenarios. However, they also confirm that, under certain flow field conditions, the traditional global boundary conditions might still be preferable. One reason that justifies their common use in the literature is the ease and effectiveness of implementing these global conditions using cosine and sine transforms, respectively. Despite this advantage, it is important to note that such an approximation of the boundary conditions can affect the overall quality of the image registration, as evidenced by our analysis.

The inverse consistency method, designed to enhance flow field estimation accuracy, performs well in both mono-modal and multi-modal scenarios. Nonetheless, it

is not without limitations. For example, in environments with poor or insufficient tissue contrast, the technique may fail to detect true flow field accurately, often resulting in false negatives. This is because the registration is heavily influenced by the regularisation term, which may obscure subtle flow field signals. On the other hand, in areas with transient or rapidly changing objects, the method might produce false positives due to variations in the data fidelity term. Moreover, when compared to homogeneous Neumann boundary conditions, this approach does not demonstrate any significant improvement in the multi-modal case.

MAE based flow field estimation has similar limitations. Indeed, the values are also influenced by flow field along the boundary and not only across it. To smooth this behaviour, the flow field map is computed patch-wise. Furthermore, this method of incoming/outgoing flow detections exhibits a much lower sensibility, as highlighted by the fact that inhomogeneous boundary conditions are never favored in the hyper-parameters search.

One common advantage between those two flow field estimation method is the fact that no a priori information is required. Also, they are both usable in multi-modal tasks as well. However, they are not universal: the inverse consistency based flow field estimation performs better on the DIR-Lab dataset but do not permit any improvement on the Learn2Reg dataset. We thus expect that more task specific evaluation of the flow field at the boundary could show better improvement than those tested here.

The proposed hyper-parameter optimisation demonstrates a correlation between the minimisation of DIR energy and the maximisation of image registration quality. Remarkably, when compared to ground truth—whether landmark-based or region-based—the DIR energy consistently exhibits an inverse relationship with the computed error values. This clear anti-correlation underscores the effectiveness of using DIR energy as a proxy for error minimisation. As the DIR energy decreases, image registration accuracy improves, making it a reliable indicator for fine-tuning hyper-parameters. This insight is particularly valuable in clinical settings where ground truth data is unavailable. In such cases, the use of DIR energy as a guiding metric for optimisation becomes not only practical but essential, providing a robust mechanism for enhancing registration quality without direct error measurements.

The calibration method employed in this study for boundary conditions, while effective for detailed analysis, is not optimised for clinical environments due to its time-consuming nature. Specifically, conducting a comprehensive grid-search for a $192\times149\times142$ image can take up to 3 hours, which is impractical for real-time clinical applications. This exhaustive parameter search is crucial for meticulously documenting how different boundary conditions affect outcomes. However, the findings indicate several promising approaches to streamline the hyper-parameter tuning process. These methods aim to significantly cut down the calibration time,

potentially making the approach more feasible for clinical use and improving overall efficiency in boundary condition optimisation.

First, given similar data acquisition settings, both guidances leads to very similar boundary conditions. For example, in the mono-modal task, the optimal boundary conditions with inverse consistency guidance were often a non-homogeneous Dirichlet boundary conditions with γ between 0.6 and 0.7. This parametrisation is well suited for the motion to be estimated: the breathing cycle involves a strong upward motion on the diaphragm side of the CBCT with only little magnitude motions on the other sides. An inhomogeneous Dirichlet boundary conditions represents this configuration. In the multi-modal task, the best hyper-parametrisation nearly always corresponded to a homogeneous Neumann boundary condition, due to minor motions on every part of the boundary, inherent to a partial field of view of the abdominal region. In clinical settings, it could thus be possible to estimate the best boundary conditions on one patient and then use it for every similar cases. If time allows it, the boundary condition could even be further optimised with a reduced grid-search around the previously estimated boundary conditions, or both parameters, a and γ could be estimated in each direction individually.

To enhance the efficiency of the grid-search process, one approach is to minimise the time required for each registration task. Currently, registering a 192x149x142 image takes approximately 4 minutes per iteration. Leveraging deep learning-based methods, which are recognized for their superior speed compared to traditional iterative variational approaches, could significantly cut down this registration time. These advanced methods, such as those utilising model-driven variational networks, can expedite the optimisation process, thus allowing a more comprehensive and efficient grid-search in a shorter period of time. The incorporation of these fast, data-driven techniques can streamline the overall procedure, making it more practical for real-time applications and extensive parameter exploration.

In the experiments presented, non-homogeneous Neumann boundary conditions consistently did not correspond to the minimal energy state. This discrepancy likely stems from two hypothetical factors. First, the datasets selected for these experiments may not be representative of scenarios where such boundary conditions play a significant role. Specifically, none of the depicted motions exert a direct influence on the shear tensor, unlike in interventional radiology. In that context, the insertion of surgical tools induces tissue compression, a phenomenon that can be effectively modeled using non-homogeneous Neumann boundary conditions.

Second, the methods employed for estimating the flow fields may not be well-suited to capturing this particular type of boundary condition. However, the inverse consistency-based flow field map demonstrates overall satisfactory performance, suggesting that the approach remains robust in general applications. Nevertheless, refining the flow field estimation process by experimenting with alternative

techniques more tailored to this specific context could yield better results. This would allow for the full range of boundary conditions, as defined by the current formulation, to be potential candidates, ultimately enhancing performance.

3.6 Impact of registration boundary conditions on the electric field estimation

3.6.1 Visualising the electric field

Many medical treatments rely on precise dose delivery, and IRE ablation is no exception. IRE uses electric fields to destroy cancerous cells by creating permanent pores in their membranes, leading to cell death. To maximise effectiveness and minimise harm to surrounding healthy tissues, determining the appropriate dose of electric pulses is crucial. This involves careful evaluation of the delivered electric field and real-time adjustments of the procedure.

If the electric dose is insufficient, the ablation may fail to fully eradicate the tumor. In such cases, clinicians might increase the treatment intensity by applying an additional set of electric pulses or performing a pull-back, adjusting the depth of the electrodes to ensure the entire tumor is encompassed by the electric field. This flexibility allows for precise targeting of the tumor while protecting nearby structures, making IRE a valuable, minimally invasive tool in cancer treatment

However, during an IRE ablation procedure, the interventional radiologist faces the challenge of not being able to directly visualise the electric field applied or its immediate effects on the surrounding tissues. Unlike thermal ablation techniques where heat can be monitored via imaging, the nearly invisible nature of the electric fields makes it difficult to predict the precise extent of tissue ablation. This limitation can result in uneven treatment, with the possibility of under-treating portions of the tumor or damaging nearby healthy tissue. Indeed, real-time feedback remains a significant obstacle in IRE, making careful online evaluation and follow-up imaging crucial to confirm complete tumor destruction.

On the one hand, the direct imaging of electric fields remains at an experimental stage and poses significant challenges. The primary issue lies in the fact that the materials used in most electric field sensors tend to interfere with the fields they aim to measure. This distortion occurs due to factors like surface charging, dielectric polarisation, and free carrier polarisation, which alter the electric field's natural behavior. Moreover, the sensitivity of many current electric field sensors is insufficient to accurately detect and analyse fields deep within materials or tissues [30].

If such sensors were applied in a medical context, such as IRE ablation, the process of embedding them into the tissue could be more invasive than the ablation

itself. This undermines their practical utility, as the very act of measuring the field could damage the surrounding tissue, negating the minimally invasive nature of IRE. Thus, until significant advancements are made, imaging electric fields in such a scenario remains impractical.

On the other hand, the mechanism behind IRE induces cell death primarily through apoptosis, a process that unfolds over several hours after treatment. Unlike thermal ablation techniques, which cause immediate tissue destruction, apoptosis triggered by IRE is a gradual process that cannot be immediately detected using standard imaging techniques. While initial effects such as cell membrane disruption can be observed, the full scope of cell death and tissue response becomes apparent only after a delay, complicating real-time assessment through imaging modalities like ultrasound or MRI.

As a result, while IRE is effective in sparing critical structures like vessels and ducts, its delayed therapeutic impact makes it challenging to monitor the full treatment effect in real time. Follow-up imaging is often required to confirm the complete success of the ablation after the apoptotic process has taken place

The only feasible approach to visualising the treated area during IRE procedures relies on estimating the electric field distribution. This estimation is made using key information, such as the precise localisation of the needles, which are inserted into the tissue, and the segmentation of the tumor, typically overlaid onto the same imaging scan that includes the needle placements. Additionally, the chronograms (timing and intensity) of the electric pulses are crucial to predict the electric field spread. By combining these data points, we can model the electric field, thereby approximating the ablation zone. This method, though indirect, enables some level of control and planning during the procedure, ensuring the electric field reaches the desired treatment area without damaging surrounding tissue.

In this section, we begin by presenting the model employed for estimating the electric field, a fundamental aspect of accurately assessing treatment areas. The next step involves experimenting with different boundary conditions during the registration process, using data from a selected patient in our database. Specifically, we test the effects of homogeneous Neumann boundary conditions, which have been applied in previous studies, as well as homogeneous Dirichlet boundary conditions. Additionally, we use the previously introduced novel approach investigating boundary conditions designed to minimise the registration energy. These experiments allow us to explore the impact of boundary condition choices on the precision of dose calculations. Ultimately, our findings emphasise how carefully selected boundary conditions can significantly enhance the accuracy of dose distribution, improving treatment efficacy.

3.6.2 Electric field model

When it comes to electric field simulations, choosing the right model depends heavily on the context. Static models, which assume a steady-state system, are often simpler and easier to implement. They offer a snapshot of the system at a given moment but fail to capture changes over time. This can be a limitation when the interactions or the system's state evolves, such as in biological or physical processes where time-dependent variables matter.

On the other hand, dynamic models introduce time as a factor, allowing for the modeling of complex behaviors, such as interactions between different components, feedback loops, and temporal variations in the electric field. While dynamic models can offer deeper insights and more accurate representations of real-world phenomena, their complexity increases computational demand. This tradeoff can make dynamic models less practical or necessary in clinical or industrial settings where simpler, more immediate insights may suffice.

In the case of IRE ablation, the model used is the static linear model [28]:

$$-\nabla \cdot (\sigma \nabla \phi) = 0 \tag{3.6.1}$$

where σ is the conductivity, that depends on the tissue characteristics, and ϕ is the electric potential.

Then, the electric field E is related to the electric potential ϕ as follows:

$$E = -\nabla \phi \tag{3.6.2}$$

This model plays a critical role in medical procedures, particularly those involving IRE, where precision is key. Since it tends to underestimate the actual electric field strength, clinicians can be confident that the threshold necessary for effective IRE is surpassed during treatment. This ensures that cells targeted for destruction are indeed affected. The ability of the model to provide this conservative estimate leads to more accurate treatment planning and a better understanding of the effective treatment area, ultimately improving patient outcomes.

Boundary conditions: To ensure the accuracy and stability of the electric field model, appropriate boundary conditions must be imposed. The first step is applying boundary conditions at the active needles, which serve as the primary sources of the electric field. Since the electric potential at these locations is predefined—originating directly from the source—homogeneous Dirichlet boundary conditions are employed:

$$\phi = \phi_{src}, \tag{3.6.3}$$

where ϕ is the electric potential and ϕ_{src} is the electric potential delivered from the electrode.

These conditions effectively fix the potential to known values—typically a constant, which simplifies the mathematical treatment of the problem. This approach ensures that the electric field distribution is precisely controlled in regions where the field is injected, allowing for reliable modeling of the interaction between the needles and the surrounding tissue.

Second, the presence of passive needles plays a critical role in shaping the electric field distribution, primarily attributable to their high conductivity. These conductive elements are influential modifiers of the surrounding electric landscape, affecting how the electric field propagates through the tissue. To accurately account for this interaction, we implement floating potential boundary conditions at these needles:

$$\phi = \phi_b \text{ such that } \int_{active} \sigma \partial_n \phi = 0, \tag{3.6.4}$$

where ϕ_b is the potential of the boundary.

This approach allows the potential to adjust freely in response to the electric field, ensuring that the passive needles do not impose fixed values that could skew the simulation results. By using floating potential conditions, we can capture the nuanced behavior of the electric field around these conductive components, leading to more precise and realistic modeling outcomes.

Lastly, at the simulation boundaries, we implement homogeneous Fourier-Robin boundary conditions to effectively account for the limitations of our field of view:

$$\sigma \nabla \phi \cdot \mathbf{n} + \alpha \phi = 0, \tag{3.6.5}$$

where $\mathbf{n}$ is the normal to the boundary.

This consideration is crucial, as it acknowledges the existence of additional tissue beyond the defined boundary, which can significantly influence the electric field distribution within the simulation. By employing these boundary conditions, we can model the interactions between the electric field and the surrounding biological environment more accurately. This approach not only enhances the fidelity of our simulations but also ensures that any potential effects from the adjacent tissue are incorporated into the overall analysis, thereby providing a more comprehensive understanding of the electric field behavior in a clinical context.

3.6.3 Impact of registration on the procedure evaluation

In order to efficiently use the electric field model, numerous information needs to be extracted during the procedure.

Firstly, it is essential to meticulously document the electric pulses generated during the process. To achieve this, both voltage and current chronograms are captured with precision, providing a detailed time-based representation of the electrical signals. These chronograms serve as critical inputs, as they are directly incorporated into the boundary conditions for the active needles. By doing so, we can accurately model the entry point of electrical energy within the simulation framework. This ensures that the dynamics of the electric field are properly represented, allowing for a realistic depiction of the interactions between the electrical stimuli and the surrounding medium. This thorough documentation and integration not only enhance the fidelity of the simulation but also facilitate more effective analysis and optimisation of the electroporation process.

Secondly, the successful application of the previously discussed boundary conditions, along with the floating potential conditions, hinges on the precise localisation of the needles. This critical task is accomplished through the innovative pipeline outlined in Chapter 2, where deep learning techniques are synergistically integrated with the Hough transform to provide an analytical representation of the electrodes. The active needles play a pivotal role in delivering the electric field, while the passive needles, characterised by their high conductivity, interact with the field in significant ways. Therefore, accurately pinpointing the positions of both needle types is essential for a comprehensive evaluation of the electric field distribution. This precision not only enhances the effectiveness of the electroporation process but also ensures the reliability of the simulation outcomes, ultimately leading to more informed clinical decisions.

Lastly, it is imperative to address the differences in electrical conductivity between the tumor and the surrounding liver tissue. To achieve this, test pulses are administered prior to the IRE procedure to accurately measure the conductivity values in both tissue types. Utilizing established values from the literature as a foundation, these conductivities are meticulously adjusted by hand to ensure precision.

Moreover, precise localisation remains crucial for effectively capturing the dynamics of this phenomenon. Thus, the accurate registration that aligns the tumor within the same reference frame as the electrodes responsible for delivering the electric field is essential. As depicted in Figure 3.15, a mis-registration as small as 1mm impacts significantly the electric field estimation, leading to errors in the evaluation of the procedure. The alignment is thus vital for a thorough estimation of the treated area, enabling the IRE procedure to be conducted with optimal efficacy and safety. By ensuring these factors are considered, we can enhance the overall outcomes of the treatment.

Additionally, ensuring that the tumor, along with a sufficient safety margin, is effectively electroporated is paramount. This necessitates precise localisation of the

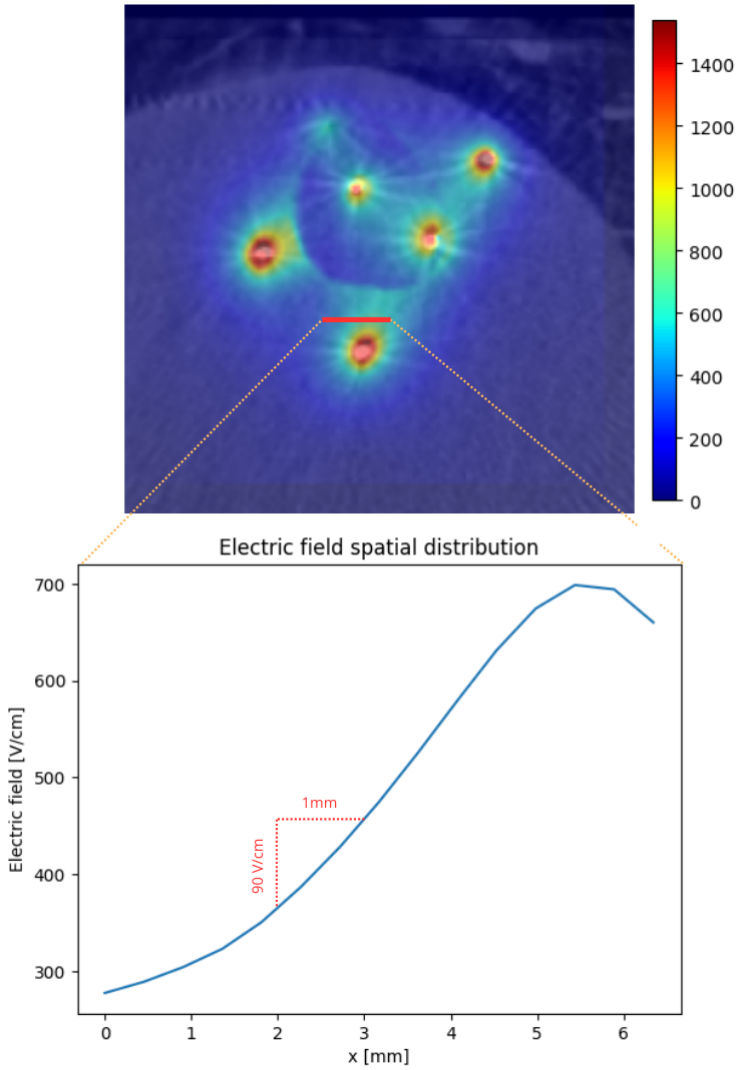


Figure 3.15: Electric field distribution as delivered by electrodes in the liver during IRE. A displacement of 1mm resulting in a change in the electric field magnitude of 90 V/cm can significantly affect the accuracy of electroporation treatments. This could cause areas that were not irreversibly electroporated (non-treated) to appear as if they were treated, and vice versa.

tumor in relation to the electrodes. Accurate positioning allows for optimal delivery of the electric field, which is critical for maximising the treatment's efficacy. By confirming that the tumor is situated correctly with respect to the needles, we can guarantee that the electroporation process targets not only the tumor but also the surrounding tissue, thereby enhancing the therapeutic window. Ultimately, this

attention to localisation plays a crucial role in the success of the electroporation procedure and contributes to improved patient outcomes.

Hence, the registration process significantly impacts the estimation of the electric field during electroporation procedures. Accurate tumor localisation is essential for adjusting the conductivity parameters, which directly influence the equations governing the electric field potential and the boundary conditions at the passive electrodes. This precise mapping ensures that the electric field is appropriately configured to target the tumor effectively. Furthermore, the successful assessment of the procedure hinges on this registration; interventional radiologists must ascertain whether the entirety of the tumor has undergone irreversible electroporation. This information is critical for evaluating treatment efficacy and determining subsequent patient management strategies. Ultimately, the accuracy of the registration process not only enhances the effectiveness of the electroporation but also supports optimal patient outcomes through informed clinical decision-making.

3.6.4 Experimental validation

To document the impact of boundary conditions for the registration on the electric field simulation, we test 3 boundary conditions, namely homogeneous Dirichlet, homogeneous Neumann, as used previously, and boundary conditions minimising the DIR energy, as guided by the inverse consistency. We do so on a patient of the database provided by the interventional radiologists at AP-HP Avicenne Hospital. The patient was selected after consideration of the tumor location: we chose a tumor close to the field of view boundary in order to observe the effect of the boundary conditions best.

Cylindrical field of view: In this step, we adjust the boundary condition application to accommodate the cylindrical field of view, which significantly alters how spatial data is considered at the boundary compared to the cubic field of view used in datasets like DIR-Lab and Learn2Reg. By tailoring the boundary conditions to this specific geometry, we capture variations in shape and volume more effectively, improving the precision of medical imaging registration and the subsequent analysis such as the estimation of the delivered electric field.

In this process, artificial padding is added to match the cubic array dimensions of a cylindrical CBCT scan. The padding, which is assigned a distinct value of -999, is segmented using a thresholding technique to differentiate it from the actual scan data. To reduce memory usage, only the central slice of this segmented mask, in the vertical direction, is retained for the boundary condition application at the edges of the field of view. As the multi-resolution scheme proceeds, both the segmentation mask and the input images are downsampled at each stage. To isolate

the perimeter, the mask is eroded once, and the difference between the original and eroded masks is calculated. This technique also helps eliminate some noise in the segmentation mask, as the intensity of -999 can also be present within the CBCT field of view.

For each point along the perimeter of the boundary, the normal vector is calculated by taking the difference between the coordinates of the point and the center of the disk in the mask slice. This approach ensures that the normal vector points towards the center, defining the direction normal to the boundary. Afterward, this vector is normalised, meaning its magnitude is adjusted to 1. Once normalised, this vector is subtracted either once or twice, depending on the required position, to determine two key points along the normal. These two locations are used to compute the transformation field at the boundary, which is essential for applying the boundary conditions:

$$T_s^{\mathbf{r}} = \frac{\beta_s^{\mathbf{r}}(4T_s^{\mathbf{r}-\mathbf{n}_1} - T_s^{\mathbf{r}-2\mathbf{n}_1}) + 2\gamma g_s^{\mathbf{r}}}{\beta_s^{\mathbf{r}} + 2}, \tag{3.6.6}$$

where $\mathbf{r}$ is the location of the voxel on the boundary and $\mathbf{n}_1$ is the normalised vector normal to the boundary.

The top and bottom slice of the cylinder are treated seperately following equations 3.3.13 and 3.3.14.

Automatic adaptation of the boundary condition: We document the optimisation of the boundary conditions and the resulting transformation field. The same grid-search strategy is used here as the purpose of this study is to establish the impact of optimal boundary conditions prior to introducing this approach in clinical settings. In the previous section, it was shown that the DIR energy correlates with the registration quality as measured with different kinds of ground-truth: landmarks and segmentation masks. The hyper-parametrisation results in a non-homogeneous Dirichlet boundary conditions with $\gamma = 0.1$ as seen in Figure 3.16.

Different transformation fields: The spatial transformation obtained with the optimal boundary conditions is compared to those resulting from the application of the two most commonly used boundary conditions in the field of image registration, namely homogeneous Neumann and homogeneous Dirichlet boundary conditions. Prior to the implementation of the proposed framework for automatically adaptable boundary conditions, Neumann boundary conditions were applied at the edges of the multi-dimensional array representing the CBCT. It is worth noting that we hereforth consider the edges of the cylindrical field of view.

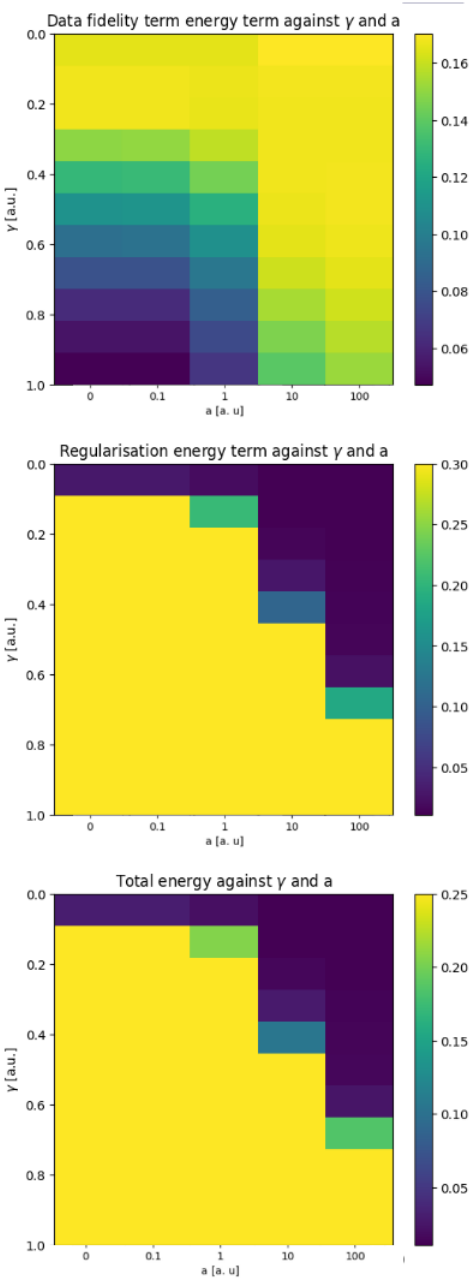


Figure 3.16: The outcomes of the hyper-parameter search for Patient 3 of the IRE database. The data fidelity term, the regularisation term, and their sum, i.e. the total registration energy, are displayed.

The Euclidean distance between transformations reveals deviations exceeding 10mm (Fig. 3.17), notably at both the perimeter of the cylinder and towards the

center of the image. This suggests that the choice of boundary conditions—whether Neumann, Dirichlet, or optimised adaptive conditions—plays a critical role in influencing tumor localisation within medical imaging. These disparities are particularly significant when estimating the electric field, as they can lead to substantial errors in treatment.

Treated areas comparison: To evaluate the impact of boundary conditions in the registration process on IRE procedure assessment, we focus on the 400 V/cm isoline, which defines the effective zone for IRE ablation. By analysing different metrics, we can quantitatively determine how different boundary conditions, namely homogeneous Dirichlet, homogeneous Neumann and locally adapted boundary conditions, influence the quality of the registration and its resulting effect on the estimation of the IRE treatment zone.

First, we use the Dice coefficient (Eq. 3.4.8) to assess the degree of overlapping between the estimated treated area. A value close to 1 signifies a strong similarity.

Second, we compute the Hausdorff distance:

$$D = max_{a \in A} min_{b \in B} ||a - b||_2, \tag{3.6.7}$$

where A and B are the region to compare.

The boundary-based metric quantifies disparities by measuring how the edges of an object deviate under different boundary conditions. Since boundary conditions significantly impact the edges of the field of view image registration, this metric is crucial for ensuring accuracy. Specifically, the boundaries of the estimated treated area in IRE must be carefully evaluated, as even a small number of surviving cancerous cells can become the origin of tumor recurrence. Ensuring precision in the determination of the IRE-affected area helps improve treatment efficacy and prevent tumor regrowth. By accurately defining the boundaries of the ablation zone, clinicians can minimise the risk of leaving behind viable cancer cells, leading to a more complete and effective treatment.

Accurate volume estimation also helps in assessing the procedural outcome and enables better planning for follow-up interventions if necessary, further improving patient outcomes in cancer treatment.

The Dice coefficient, while useful for measuring the overall overlap between segmented areas, often fails to capture fine differences in the boundaries of treated areas, as seen in Table 3.5. In contrast, the Hausdorff distance provides deeper insights as it measures the maximum deviation between boundary points, which is crucial when comparing different boundary conditions.

For instance, in scenarios involving homogeneous Neumann boundary conditions versus locally adapted boundary conditions, the Hausdorff distance reveals a more pronounced difference, than between homogeneous Dirichlet boundary conditions

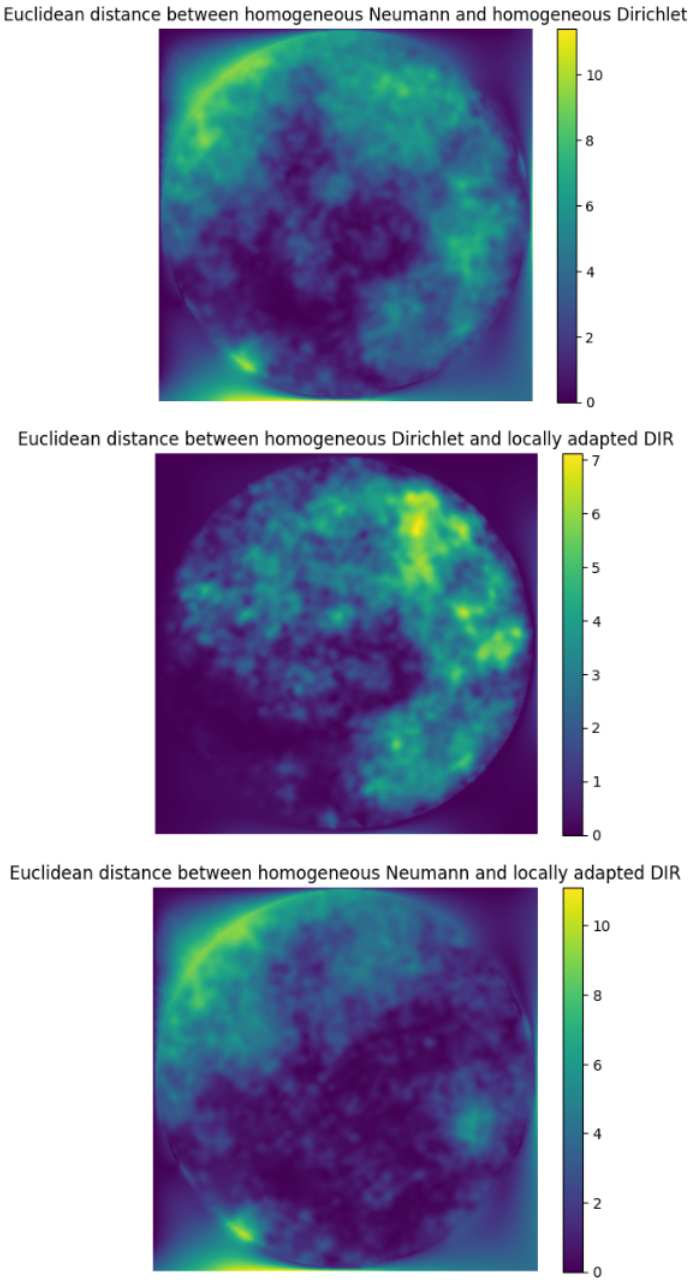


Figure 3.17: Euclidian distance between a slice of the transformation fields estimated with homogeneous Neumann, homogeneous Dirichlet, and locally adapted boundary conditions

and locally adapted boundary conditions. This is due to the hyper-parameterisation, where the automatically adapted boundary conditions are set to non homogeneous

Table 3.5: Comparison of the treated areas, computed as the 400V/cm isoline of the delivered electric field, with different boundary conditions for the registration tasks. In blue are the Hausdorff distances in mm, in red the Dice coefficients and in black the treated volumes in mm^3. The Hausdorff distance highlights significant differences in the estimated electric fields delivered under varying boundary conditions for the task of image registration. When comparing the fields generated with homogeneous Neumann boundary conditions to those from other boundary conditions, the Hausdorff distance reveals substantial deviations. This is further supported by the analysis of treated volumes, where the minimum discrepancy reaches $33mm^3$. Although seemingly minor, this difference in volume is clinically significant, as it could be enough to cause cancer recurrence by allowing untreated or insufficiently treated areas to remain, posing a risk of the disease returning.

	Homogeneous Neumann	Homogeneous Dirichlet	Locally adapted
Homogeneous Neumann	44117	1.4911	1.4953
Homogeneous Dirichlet	0.9938	43652	0.7829
Locally adapted	0.9936	0.9988	43685

Dirichlet boundary conditions with $\gamma = 0.1$, allowing local variations with respect to the estimated flow fields $\mathbf{g}$. This enables a more detailed comparison, especially in the context of cancer treatment, where boundary accuracy is critical.

The discrepancies in treated volumes, ranging from $33mm^3$ to $465mm^3$, are clinically relevant, especially since even $33mm^3$ of untreated tissue could serve as a sufficient space for new tumor growth. This emphasises the critical need to carefully apply accurate boundary conditions during treatment evaluation. Inaccurate or improperly chosen boundary conditions can directly influence the estimation of the treated area, leading to either over- or underestimation of the region affected by IRE. By refining boundary conditions, clinicians and researchers can ensure a more precise understanding of the treated volume, reducing the risk of recurrence.

Accurate tumor localisation is essential in IRE, where precise field targeting is crucial. The findings underline the importance of boundary condition optimisation to minimise such disparities, ultimately enhancing the reliability of medical image registration and related treatment outcomes.

3.7 Conclusion

Registration is an essential pre-processing step in medical imaging, pivotal for ensuring accurate image analysis and interpretation. Radiologists frequently encounter images captured from various sources and at different times, each potentially featuring slight variations in anatomical positioning or patient movement. To provide a coherent view and facilitate accurate diagnosis, it is imperative to align

these images. This process involves estimating and compensating for any movement or misalignment between images, thereby ensuring that anatomical features are correctly matched across different scans. By aligning anatomical structures across images, registration enhances the clarity and consistency of the images, significantly aiding in the diagnostic process and improving the reliability of clinical assessments.

This study highlights the pivotal role of boundary conditions in achieving high-quality image registration. Specifically, the absence of adequate information beyond the borders of the field of view can lead to significant mis-registration. This issue arises because the registration process relies on boundary conditions to inform the alignment of images. When boundary data is lacking, errors introduced at the edges can propagate inward, exacerbating mis-registration across the entire image. This phenomenon is particularly pronounced due to the influence of the regularisation term, which attempts to smooth out discrepancies but can inadvertently amplify errors originating from the boundaries. Consequently, ensuring accurate and well-defined boundary conditions is essential for maintaining registration fidelity and minimising distortions throughout the image.

We introduce an innovative approach utilising a Robin-type boundary condition to enhance image registration. This technique involves incorporating flow field information tailored to each specific image pair, which significantly improves the accuracy of registration at the image borders. The proposed method is efficient and straightforward, requiring only two hyper-parameters to fine-tune.

Prior to initiating the registration process, the flow field map that captures the relevant boundary information for the image pair are calculated.

The computed boundary conditions are integrated into the registration algorithm. This step ensures that the algorithm accurately aligns the images by using the flow field information to guide the registration at the borders.

This method seamlessly integrates with most existing registration algorithms, requiring minimal adjustments, thus offering a practical solution to enhance registration accuracy and efficiency.

For the flow field map, we investigated two methods. The first one, MAE-based, relies on the voxel intensity to detect flow field. It requires an initial motion estimate, with the homogeneous Neumann boundary condition as it is the least restrictive. However, it does not allow to take advantage of the source term in the boundary conditions. Indeed, the lack of direction information limits the use of this term. Hence, we do not recommend this particular guidance when little is known about the motion.

The second method, based on inverse consistency, involves both forward and backward motion estimates to identify discrepancies. These discrepancies are interpreted as flow field, which provides a deeper understanding of the motion to estimate. This technique is notable for its potential to enhance registration

accuracy beyond traditional boundary conditions. By comparing the results of inverse consistency with standard boundary conditions, such as homogeneous Neumann and homogeneous Dirichlet conditions, this method either demonstrates improvements or validates the efficacy of the conventional choices. This comparative analysis reveals whether more complex boundary conditions offer a significant advantage or if the simpler, established methods remain optimal.

In our numerical workflow for IRE, we have chosen to integrate only the latter approach to enhance the registration algorithm's effectiveness. This deliberate choice aims to refine the precision of electric field estimations, which are crucial for accurately assessing the treatment area in IRE ablation, and potentially adapt the procedure as it is performed. By incorporating this refined method, we show a significant improvement in the overall quality of electric field mapping. This enhancement facilitates a more accurate and reliable evaluation of the treatment zone, thereby optimising the efficacy of IRE procedures and ensuring better outcomes.

In brief

Boundary conditions are a crucial yet often overlooked factor in image registration accuracy. To address this, we introduce a novel framework that automatically adapts Robin-type boundary conditions, accounting for the complex flow field dynamics at the registration boundaries. Unlike conventional approaches that neglect or inadequately handle boundary behavior, our method aligns boundary conditions with the true physiological or physical aspects of the problem. This results in improved registration outcomes and enhanced model robustness, emphasising the significant role boundary conditions play in the overall accuracy and reliability of image registration.

This method is integrated into a variational registration algorithm, though its design is inherently versatile, making it adaptable to a wide range of other registration algorithms. Its implementation necessitates two key modifications: first, the computation of the flow field that governs the boundary conditions, and second, the update of the transformation field specifically at the boundaries. Additionally, the method introduces two new hyper-parameters that, crucially, are shown to be optimisable by leveraging the DIR energy as a guiding metric. Minimising the DIR energy correlates strongly with reducing the registration errors when ground truth is available, demonstrating the efficacy of this approach. Importantly, in scenarios where ground truth information is absent—such as in clinical settings—these hyper-parameters can still be reliably optimised based on DIR energy minimisation alone. This capability ensures that the method remains practical and effective in real-world applications, where direct validation data is often unavailable, while still maintaining a high level of registration accuracy.

Our approach has yielded promising results, both enhancing and validating the effectiveness of commonly employed global boundary conditions. By directly addressing the often-overlooked aspect of boundary conditions, our framework makes a substantial contribution to improving the accuracy and reliability of image registration outcomes. This novel approach provides a critical advancement in the field, ensuring that boundary behaviors are no longer an afterthought but a key factor in optimising registration quality. By refining how boundaries are managed within registration algorithms, our method paves the way for more precise, robust solutions across a variety of applications with or without subsequent analysis such as dose computation, offering a significant step forward in the accuracy of medical and computational imaging practices.

Chapter 4

Conclusion

Contents

In the realm of contemporary medical science, computer vision stands as a cornerstone, revolutionising the way we approach patient care by delivering vital insights with minimal risk to individuals. This technology has become indispensable, particularly in the field of oncology, where precision and accuracy are paramount. Medical imaging, powered by advanced computer vision techniques, plays a critical role in diagnosing cancer, devising tailored treatment plans, and monitoring the progression of the disease.

Among the various applications, certain therapies necessitate real-time image guidance to ensure optimal outcomes. One such advanced procedure is IRE, a focal point of this thesis. IRE exemplifies how image-guided techniques can enhance the efficacy of cancer treatments, highlighting the transformative impact of computer vision in modern medical practices. By integrating these cutting-edge technologies, we can achieve unparalleled precision in cancer care, ultimately improving patient outcomes.

In close collaboration with interventional radiologists conducting IRE on patients with liver and pancreatic cancers, this work has significantly advanced the existing numerical workflow. The enhancements made were crafted to seamlessly integrate with the clinical protocols employed by the physicians, thereby optimising the overall efficacy and precision of the treatment process. This integration not only refined the procedural workflow but also ensured that the improvements were perfectly aligned with the practical needs and routines of the healthcare professionals, leading to a more streamlined and effective clinical practice.

Our initial contribution addresses the critical task of localising the electrodes responsible for delivering electric pulses during IRE as captured in CBCT scans. Recognising the exceptional efficacy of deep learning in visual recognition tasks, we have integrated a deep neural network into our localisation algorithm. This advanced network is designed to perform a preliminary, coarse segmentation of the objects of interest—specifically, the extremely thin electrodes used in the procedure.

To tailor the network's performance to the unique challenges of this task, including the segmentation of highly slender structures, we optimised the network architecture, learning strategy, and post-processing techniques, drawing on established methodologies from the latest research. Given the significant extent of artifacts present in the CBCT images, the initial segmentation provided by the neural network is subsequently refined using a Hough transform, producing the analytical representation of the active part of the electrode, invisible on the CBCT due to artefact, and ensuring accurate placement within the clinical setting.

Secondly, this thesis makes a significant contribution to the expansive field of image registration through the introduction of dynamically adaptable boundary conditions. The challenge of boundary conditions in image registration is well-documented as one of the most intricate and demanding aspects of the process, yet

there have been relatively few advancements in addressing this complexity. In this work, we propose a novel approach utilising Robin-Fourier type boundary conditions, where flow field information is seamlessly integrated to tailor the registration process to the specific requirements of the task at hand. This innovative method represents a crucial advancement, offering a sophisticated solution to a long-standing problem.

The hyper-parameters derived from the boundary conditions are fixed through an exhaustive grid search process aimed at minimising registration energy. However, this approach is not without its drawbacks, particularly its time-consuming nature.

These innovative boundary conditions offer significant advantages across various application domains by effectively mitigating mis-registration at the peripheries, which can otherwise propagate towards the central regions of the image. More specifically, they represent a substantial enhancement in the processing of partial fields of view—a common challenge in medical imaging. In clinical settings, where it is often necessary to limit the captured zone to reduce radiation exposure or address other practical constraints, these boundary conditions ensure greater accuracy and reliability in image registration. By addressing the inherent issues at the boundaries, these advancements facilitate more precise and effective image analysis, even within restricted capture areas, thereby improving diagnostic and treatment outcomes in the medical field.

4.1 Perspectives

4.1.1 On the segmentation of fine objects

Preliminary experiments suggest that simpler neural network architectures often yield better results for segmentation tasks. However, there are still numerous untapped avenues for improvement. For instance, incorporating attention gates could enhance the network's ability to focus on critical regions during both training and inference. These gates would help the model prioritise relevant areas, leading to better segmentation outcomes. Additionally, deeper supervision could refine segmentation accuracy by extracting information from multiple layers, which would optimise the use of hierarchical features throughout the network.

Nevertheless, given the current U-Net architecture with just two resolution levels, the impact of deeper supervision might be limited. The simplicity of this design, prioritised to limit overfit, could be restricting its potential, and increasing complexity through multi-resolution supervision might not lead to significant improvements. Yet, these techniques—attention mechanisms and deeper supervision—have shown success in similar tasks, such as segmenting thin structures like needle, suggesting their potential benefits even in this scenario [103].

Moreover, it is essential to consider the time constraints that limit the complexity

of algorithms in real-world applications. In clinical settings, any modifications or enhancements to the simple U-Net architecture must be time-efficient to ensure that they do not hinder the speed of image processing and decision-making. While more advanced techniques, such as attention mechanisms or deeper supervision, may enhance segmentation accuracy, they must be balanced with the requirement for fast inference times. Maintaining this equilibrium is critical to ensuring that the model remains practical and usable in the fast-paced environment of clinical practice, where real-time or near-real-time decisions are necessary.

In clinical settings, the proposed coarse-to-fine approach has not yet been adopted. The interventional radiologists favor the semi-automatic needle detection algorithm. Though it meets the time and memory constraints, the accuracy and precision are still questionable. The limitations of deep learning approaches, such as CNNs, arise from the lack of a theoretical error bound. The learned convolutional filters in CNNs are not always fully understood, adding to the uncertainty. Many error metrics provided in research are empirical and often based on small datasets, which limits their generalisability. To increase the trust in these algorithms for clinical use, further work is needed. An entirely automatic algorithm for the electric field modelisation, including the segmentation of the tumor, the localisation of the electrodes, and the registration of the CBCT would be ideal.

4.1.2 On local boundary conditions for image registration

Flow field estimation utilising spatial transformation estimates calculated during preliminary registration offers notable advantages in generality and versatility. This methodology estimates flow fields indirectly through transformations, enabling its application across diverse scenarios and datasets. However, this versatility is accompanied by critical limitations. One significant challenge is the occurrence of false positives, particularly in areas characterised by sudden intensity contrasts, such as sharp edges or boundaries. In these cases, the data fidelity term in registration algorithms may misinterpret rapid changes, failing to differentiate between genuine object motion and intensity variations caused by lighting or noise, and to account for transcient objects. Another limitation is the tendency for regularization, that can result in the loss of subtle flow patterns, which are essential in contexts like medical imaging or fine object detection.

Despite these drawbacks, the reliance on spatial transformation estimates ensures broad applicability, making the approach adaptable across various contexts. Ultimately, the choice lies in balancing the need for enhanced registration quality tailored to specific tasks against the advantages of maintaining a more universal application.

The calibration method used for boundary conditions in this study, while effective for detailed analysis, poses challenges in clinical settings due to its lengthy

process—taking up to three hours for a comprehensive grid search on a 192×149×142 image. This extensive parameter search is crucial for assessing how different boundary conditions affect outcomes, but not pertinent in the operating room when the main focus is on efficiency and precision. The experiment results indicate promising approaches to streamline hyper-parameter tuning and reduce calibration time, thereby improving clinical applicability.

For similar data acquisition processes, inverse consistency guidance often yields comparable boundary conditions, with non-homogeneous Dirichlet boundaries favored for the breathing motion and homogeneous Neumann boundaries preferred for the abdominal region. Given a certain type of aquisition, a lung CT for instance, the same set of hyper-parameters could thus be used for every instances. If time allows, the boundary conditions could be further refined thanks to a grid search restricted around the usual hyper-parameters for the type of acquisition at hand.

To enhance grid search efficiency, we also suggest minimising registration time, currently around four minutes per iteration. Employing deep learning techniques, recognised for their speed compared to traditional methods, could significantly reduce this registration time and optimise the overall process. This would make the calibration procedure more practical for real-time applications and facilitate broader parameter exploration, tailored to each patient.

The experiments indicate that non-homogeneous Neumann boundary conditions do not align with the minimal energy state, likely due to two factors. Firstly, the selected datasets may not accurately reflect scenarios where these boundary conditions are critical, as none of the motions analysed affect the shear tensor. Secondly, the methods used to estimate flow fields might not effectively inform this boundary condition type. Despite this, the inverse consistency-based flow field map shows satisfactory overall performance, suggesting robustness in general applications. However, exploring alternative flow field estimation techniques tailored to this specific context may improve results, allowing a wider range of boundary conditions defined by the current formulation to be considered, thereby enhancing performance. Additionally, we could devise a synthetic dataset where non-homogeneous Neumann boundary conditions are the correct constraint to apply. This would allow to fully investigate the reason why this boundary condition do not effectively minimise the DIR energy.

4.1.3 Evaluating both contributions with respect to IRE efficacy

To better highlight the advancements brought forth by the contributions in this thesis, notably the integration of deep learning into needle localisation and the development of a framework for automatic adaptive boundary conditions in

the registration task, further experimentation could leverage post-operative MRI imaging. These scans, where both reversible and irreversible electroporation zones are clearly visible, provide an ideal setting to quantitatively assess the performance of each method individually, as well as their combined effectiveness. This approach would offer a more robust analysis of the improvements achieved by deep learning for accurate needle placement and by adaptive boundary conditions for more precise image registration on the evaluation of the treated area. Together, these tools would enable a better understanding of how both innovations contribute to more effective medical procedures, improving precision and outcomes in complex interventions like electroporation-based therapies.

4.1.4 Generalisability

Given the rapid evolution of medical sciences—encompassing advancements in treatment methodologies, clinical protocols, and medical imaging technologies—it is imperative that the proposed methods remain adaptable to these continual changes within clinical environments.

The proposed enhancements to the neural network for the segmentation of thin objects are not limited to the specific case of electrode localisation but hold promise for broader applications involving the segmentation of thin objects. For example, patch-based datasets can be crafted in a similar fashion to address the inherent dataset imbalances, thereby improving segmentation accuracy for similarly challenging scenarios. However, it is important to note that deep learning models are intrinsically data-dependent; thus, a neural network trained for one specific application may not be directly applicable to different setups. Each new task necessitates a comprehensive learning process, including the creation of new ground truth data, to fully leverage the proposed methodology.

To navigate this challenge for related tasks, transfer learning offers a valuable solution. By utilising a pre-trained network as a starting point, one could significantly streamline the development of a new model. In this approach, the pre-trained network is partially retrained to adapt to a new task, especially in cross-domain transfer learning scenarios. When the source and target tasks differ significantly, modifications are often necessary. For example, in classification tasks, the pre-trained model might have a fully connected layer as the head, with output neurons corresponding to the number of classes. However, if the new task is image segmentation, the fully connected layer may be replaced by a convolutional layer. This adjustment ensures that the model can handle the requirements of the new task appropriately.

The pre-trained network may also more simply be partially or entirely fine-tuned, choosing the pre-trained weights as a starting point for the optimisation on the target task and/or dataset. Considering cross-modal transfer learning where

the same tasks knowledge needs to be transferred to a different medical imaging technique, one could choose to retrain the encoder to accommodate the unique characteristics of the new data, while the decoder's weights are preserved, given that they have already been optimised for segmentation construction. This strategic reuse of existing knowledge would not only reduce the time and computational resources required to develop a new network but also alleviate the need for extensive labeled dataset while ensuring that the results remain precise and relevant to the new context.

The method has been successfully used before in cross-modal settings, significantly improving the quality of the model when both modalities contain enough information for the task [1], and cross-domain, experimenting with other medical tasks like the segmentation of a different organ [61] or tasks on natural images [1].

Conversely, transfer learning could substantially improve the needle segmentation tasks by using pre-trained models from large, diverse datasets like FLARE 2021 [65]. This dataset includes 360 labeled CT scans, which focus on abdominal organ segmentation, and its modality is quite similar to CBCT. Fine-tuning the parameters of a pre-trained network on a large dataset can significantly enhance segmentation quality in medical images. The network, already trained on a diverse set of images, gains valuable insight into anatomical structures, allowing it to generalise better to the new clinical dataset. This results in a reduced risk of over-fitting, as the model does not need to learn from scratch, and speeds up the training process. Such fine-tuning can lead to improved segmentation performance, even when working with smaller datasets typical in clinical environments. Leveraging knowledge from the FLARE dataset allows the model to better capture the nuances of fine object segmentation near the liver, facilitating more accurate and efficient clinical outcomes.

Similarly, the novel boundary conditions proposed here are designed with flexibility in mind. The modular structure of this approach—characterised by the separation between the highly parametric boundary conditions and the prior information injected—facilitates straightforward enhancements and adaptations to specific clinical scenarios. Future developments could include tailoring the estimation of incoming and outgoing flow fields at the boundaries to more specific tasks, as well as exploring even more comprehensive boundary condition frameworks. This adaptability ensures that the methods can evolve alongside technological advancements, maintaining their relevance and effectiveness in an ever-changing medical landscape.

Bibliography

[1] K Aderghal, K Afdel, J Benois-Pineau, and G Catheline. "Improving Alzheimer's stage categorization with Convolutional Neural Network using transfer learning and different magnetic resonance imaging modalities". In: *Heliyon* 6.12 (2020).

[2] I Agirrezabal, M Bouattour, D. J Pinato, A D'Alessio, V. K Brennan, P. L Carion, S Shergill, N Amoury, and V. Vilgrain. In: *European Journal of Cancer* 196 (2024). DOI: https://doi.org/10.1016/j.ejca.2023.113427.

[3] G Allasia, R Cavoretto, and A. De Rossi. "A class of spline functions for landmark-based image registration". In: *Mathematical Methods in the Applied Sciences* 35.8 (2012), pp. 923–934.

[4] M.M Assifi and O.J. Hines. "Anti-angiogenic agents in pancreatic cancer: a review". In: *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)* 11.5 (2011), pp. 464–469.

[5] E Aydin, Jia Y, S Alessandro, C Kejia, Y Vahid, and Z. Zhuoli. "Combination of natural killer cell-based immunotherapy and irreversible electroporation for the treatment of hepatocellular carcinoma". In: *Annals of Translational Medicine* 9.13 (2021). ISSN: 2305-5847. URL: https://atm.amegroups.org/article/view/72885.

[6] S Baker, D Scharstein, J.P Lewis, S Roth, M.J Black, and R. Szeliski. "A database and evaluation methodology for optical flow". In: *International journal of computer vision* 92 (2011), pp. 1–31.

[7] H Bay, T Tuytelaars, and L. Van Gool. "Surf: Speeded up robust features". In: (2006), pp. 404–417.

[8] S Bharati, M Mondal, P Podder, and V.B. Prasath. "Deep learning for medical image registration: A comprehensive review". In: *arXiv preprint arXiv:2204.11341* (2022).

[9] N Bhardwaj, G Gravante, A.D Strickland, F Ahmad, J Dormer, and D.M. Dennison A.R andLloyd. "Cryotherapy of the liver: A histological review". In: *Cryobiology* 61.1 (2010), pp. 1–9.

[10] A Blazevski, M.J Scheltema, A Amin, J.E Thompson, and P.D. Lawrentschuk N. Stricker. “Irreversible electroporation (IRE): a narrative review of the development of IRE from the laboratory to a prostate cancer treatment”. In: *BJU international* 125.3 (2020), pp. 369–378.

[11] S.G Bown, A.A Rogowska, D.E Whitelaw, W.R Lees, L.B Lovat, L Ripley P. Jones, P Wyld, A Gillams, and A.W.R. Hatfield. “Photodynamic therapy for cancer of the pancreas”. In: *Gut* 50.4 (2002), pp. 549–557.

[12] U.-D Braumann and J.-P. Kuska. “Influence of the boundary conditions on the result of non-linear image registration”. In: 1 (2005), pp. I–1129.

[13] E Castillo, R Castillo, D Fuentes, and T. Guerrero. “Computing global minimizers to a constrained B-spline image registration problem from optimal l1 perturbations to block match data”. In: *Medical physics* 41.4 (2014), p. 041904.

[14] G Chan and U. Pua. “Irreversible electroporation of the pancreas”. In: 36.03 (2019), pp. 213–220.

[15] W Chen, C.-L Chiang, and L. A. Dawson. “Efficacy and safety of radiotherapy for primary liver cancer”. In: *Chinese Clinical Oncology* (2020).

[16] J Cheng and B. Hofmann. “Regularization Methods for Ill-Posed Problems”. In: (2011). Ed. by Otmar Scherzer, pp. 87–109. DOI: 10.1007/978-0-387-92920-0_3. URL: https://doi.org/10.1007/978-0-387-92920-0_3.

[17] C Dalitz, T Schramke, and M. Jeltsch. “Iterative Hough transform for line detection in 3D point clouds”. In: *Image Processing On Line* 7 (2017), pp. 184–196.

[18] C Dalitz, T Schramke, and M. Jeltsch. “Iterative Hough transform for line detection in 3D point clouds”. In: *Image Processing On Line* 7 (2017), pp. 184–196.

[19] R.V Davalos, I. L Mir, and B. Rubinsky. “Tissue ablation wit irreversible electroporation”. In: *Annals, of biomedical engineering* 32 (2005), pp. 223–231. DOI: https://doi.org/10.1007/s10439-005-8981-8.

[20] Y Dong, Z Wan, X Gao, and L. Yang G. Liu. “Reprogramming immune cells for enhanced cancer immunotherapy: targets and strategies”. In: *Frontiers in Immunology* 12 (2021), p. 609762.

[21] G Falksom, C. G Moertel, P Lavin, and Carbone P. P. Pretorius FJ. “Chemotherapy studies in primary liver cancer. A prospective randomized clinical trial”. In: *Cancer* 42.5 (1978), pp. 2149–2156.

[22] T Fechter and D. Baltas. “One-shot learning for deformable medical image registration and periodic motion tracking”. In: *IEEE transactions on medical imaging* 39.7 (2020), pp. 2506–2517.

[23] L.A Feldkamp, L.C Davis, and J.W. Kress. “Practical cone-beam algorithm”. In: *J. Opt. Soc. Am. A* 1.6 (1984), pp. 612–619. DOI: 10.1364/JOSAA.1.000612. URL: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-1-6-612.

[24] G.-S Feng, K. L Hanley, Y Liang, and X. Lin. “Improving the efficacy of liver cancer immunotherapy: the power of combined preclinical and clinical studies”. In: *Hepatology* 73 (2021), pp. 104–114.

[25] B Fischer and J. Modersitzki. “A unified approach to fast image registration and a new curvature based registration technique”. In: *Linear Algebra and its applications* 380 (2004), pp. 107–124.

[26] Y Fu, Y Lei, T Wang, W.J Curran, T Liu, and X. Yang. “Deep learning in medical image registration: a review”. In: *Physics in Medicine & Biology* 65.20 (2020), 20TR01.

[27] O Gallinato, B Denis de Senneville, O Seror, and C. Poignard. “Numerical workflow of irreversible electroporation for deep-seated tumor”. In: *Physics in medicine and biology* 64 (2019). DOI: https://doi.org/10.1088/1361-6560/ab00c4.

[28] Olivier Gallinato, Baudouin Denis de Senneville, Olivier Seror, and Clair Poignard. “Numerical modelling challenges for clinical electroporation ablation technique of liver tumors”. In: *Mathematical Modelling of Natural Phenomena* 15 (2020), p. 11.

[29] P Ganesan and L. M. Kulik. “Hepatocellular Carcinoma: New Developments”. In: *Clinics in liver disease* (2023), pp. 85–102.

[30] ER Generazio. “Electric potential and electric field imaging”. In: 1806.1 (2017).

[31] W. Girindra. “Irreversible Electroporation: Maximizing Treatment Efficacy Through Optimization Strategies and Robotic Assistance”. In: (June 2023). DOI: 10.3990/1.9789036556750.

[32] X Gong, Z Chen, J. J Hu, and C. Liu. “Advances of electroporation-related therapies and the synergy with immunotherapy in cancer treatment”. In: *Vaccines* 10.11 (2022), p. 1942.

[33] V Granata, R Fusco, M Piccirillo, R Palaia, A Petrillo, S Lastoria, and F. Izzo. “Electrochemotherapy in locally advanced pancreatic cancer: Preliminary results”. In: *International Journal of Surgery* 18 (2015), pp. 230–236.

[34] D.M Gress, S.B Edge, F.L Greene, M.K Washington, E.A Asare, J.D Brierley, D.R Byrd, Jessup J.M Compton C.C, and D.P. Winchester. "Principles of cancer staging". In: *AJCC cancer staging manual* 8 (2017), pp. 3–30.

[35] T. F Greten and B. Sangro. "Targets for immunotherapy of liver cancer". In: *Journal of hepatology* 68.1 (2018), pp. 157–166.

[36] X Gu, H Pan, Y Liang, R Castillo, D Yang, D Choi, E Castillo, A Majumdar, T Guerrero, and S.B. Jiang. "Implementation and evaluation of various demons deformable image registration algorithms on a GPU". In: *Physics in Medicine & Biology* 55.1 (2009), p. 207.

[37] P Hadjicostas, N Malakounides, C Varianos, E Kitiris, F Lerni, and P. Symeonides. "Radiofrequency ablation in pancreatic cancer". In: *Hpb* 8.1 (2006), pp. 61–64.

[38] S Hao and S Li. "A weighted mean absolute error metric for image quality assessment". In: *2020 IEEE international conference on visual communications and image processing (VCIP)*. IEEE. 2020, pp. 330–333.

[39] J He and G.E. Christensen. "Large deformation inverse consistent elastic image registration". In: (2003), pp. 438–449.

[40] K He, X Zhang, S Ren, and J. Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[41] M.P Heinrich, H Handels, and I.J.A. Simpson. "Estimating large lung motion in COPD patients by symmetric regularised correspondence fields". In: (2015), pp. 338–345.

[42] M.P Heinrich, M Jenkinson, M Bhushan, T Matin, F.V Gleeson, M Brady, and J.A. Schnabel. "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration". In: *Medical image analysis* 16.7 (2012), pp. 1423–1435.

[43] J Hong, M.T Stewart, Daniel S Cheek, D.E Francischelli, and N. Kirchhof. "Cardiac ablation via electroporation". In: (2009), pp. 3381–3384.

[44] G Huang, Z Liu, L van der Maaten, and K. Q. Weinberger. "Densely connected convolutional networks". In: *arXiv* (2018). DOI: https://doi.org/10.48550/arXiv.1608.06993.

[45] G Huang, Z Liu, L Van Der Maaten, and K.Q. Weinberger. "Densely Connected Convolutional Networks". In: (2017), pp. 2261–2269. ISSN: 1063-6919. DOI: 10.1109/CVPR.2017.243. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.243.

[46] E Inacio, L Lafitte, L Facq, C Poignard, and B. Denis de Senneville. “Adaptive local boundary conditions to improve deformable image registration”. In: *Physics in Medicine and Biology* (2024).

[47] E Inacio, L Lafitte, O Sutter, O Seror, B Denis de Senneville, and C. Poignard. “Automated needle localisation for electric field computation during an electroporation ablation”. In: (2022), pp. 1279–1284.

[48] F Isensee, J Petersen, A Klein, D Zimmerer, P. F Jarger, S Kohl, J Wasserthal, G Koehler, T Norajitra, S Wirkert, and K. H. Maier-Hein. “nnU-Net: Self-adapting framework for U-Net-based medical image registration”. In: *arXiv* (2018). DOI: https://doi.org/10.48550/arXiv.1809.10486.

[49] F Isensee, T Wald, C Ulrich, M Baumgartner, S Roy, K Maier-Hein, and P. F. Jaeger. “nnU-Net revisited: a call for rigorous validation in 3D medical image segmentation”. In: *arXiv* (2024). DOI: https://doi.org/10.48550/arXiv.2404.09556.

[50] S. Jadon. “A survey of loss functions for semantic segmentation”. In: (2020), pp. 1–7. DOI: 10.1109/CIBCB48159.2020.9277638.

[51] E. M Jaffee, R.H Hruban, M Canto, and S.E. Jern. “Focus on pancreas cancer”. In: *Cancer cell* (2002), pp. 25–28. DOI: https://doi.org/10.1016/s1535-6108(02)00093-4.

[52] X Jiang and S. Zhang R. Nie. “Image Segmentation Based on Level Set Method”. In: *Physics Procedia* 33 (2012). 2012 International Conference on Medical Physics and Biomedical Engineering (ICMPBE2012), pp. 840–845. ISSN: 1875-3892. DOI: https://doi.org/10.1016/j.phpro.2012.05.143.

[53] N Jourabchi, K Beroukhim, B.A Tafti, S.T Kee, and E.W. Lee. “Irreversible electroporation (NanoKnife) in cancer treatment”. In: *Gastrointestinal Intervention* 3.1 (2014), pp. 8–18.

[54] M Kaibori, H Kosaka, K Matsui, M Ishizaki, H Matsushima, T Tsuda, T Hishikawa H. Okumura, and M. Sekimoto. “Near-infrared fluorescence imaging and photodynamic therapy for liver tumors”. In: *Frontiers in oncology* 11 (2021), p. 638327.

[55] U. S Kamilov, C.A Bouman, G.T Buzzard, and B. Wohlberg. “Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications”. In: *IEEE Signal Processing Magazine* 40.1 (2023), pp. 85–97.

[56] L Ke, M Danelljan, Y.-W Li X. Tai, T, C.-K, and F. Yu. “Mask transfiner for high-quality instance segmentation”. In: *arXiv* (2021). DOI: https://doi.org/10.48550/arXiv.2111.13673.

[57] J. W Kim, S. S Shin, S. H Heo, J. H Hong, H. S Lim, H. J Seon, Y. H Hur, C. H Park, Y. Y Jeong, and H. K. Kang. "Ultrasound-huided percutaneous radiofrequency ablation of liver tumors: how we do it safely and completely". In: *Korean Journal of Radiology* (2015), pp. 1226–1239. DOI: https://doi.org/10.3348/kjr.2015.16.6.1226.

[58] N.N. Korpan. "Cryosurgery: ultrastructural changes in pancreas tissue after low temperature exposure". In: *Technology in cancer research & treatment* 6.2 (2007), pp. 59–67.

[59] M Kranjc and D. Miklavčič. "Electric field distribution and electroporation threshold". In: *Handbook of Electroporation, Springer* (2017). DOI: https://doi.org/10.1007/978-3-319-32886-7_4.

[60] J Larrey-Ruiz, J Morales-Sánchez, and R. Verdú-Monedero. "Generalized regularization term for non-parametric multimodal image registration". In: *Signal processing* 87.11 (2007), pp. 2837–2842.

[61] J Lee and R M Nishikawa. "Cross-organ, cross-modality transfer learning: feasibility study for segmentation and classification". In: *IEEE Access* 8 (2020), pp. 210194–210205.

[62] X Li, H Ding, H Yuan, W Zhang, J Pang, G Cheng, K Chen, Z Liu, and C. C. Loy. "Transformer-based visual segmentation: a survey". In: *arXiv* (2023). DOI: https://doi.org/10.48550/arXiv.2304.09854.

[63] D.G. Lowe. "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60 (2004), pp. 91–110.

[64] P.V Lukashevich, B.A Zalesky, and S.V. Ablameyko. "Medical image registration based on SURF detector". In: *Pattern Recognition and Image Analysis* 21 (2011), pp. 519–521.

[65] Jun Ma, Yao Zhang, Song Gu, Cheng Zhu, Cheng Ge, Yichi Zhang, Xingle An, Congcong Wang, Qiyuan Wang, Xin Liu, Shucheng Cao, Qi Zhang, Shangqing Liu, Yunpeng Wang, Yuhui Li, Jian He, and Xiaoping Yang. "AbdomenCT-1K: Is Abdominal Organ Segmentation A Solved Problem?" In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021). DOI: 10.1109/TPAMI.2021.3100536.

[66] S McDermott and D. A. Gervais. "Seminars in Interventional Radiology". In: (2013), pp. 46–55. DOI: https://doi.org/10.1055/s-0033-1333653.

[67] M.F Meloni, J Chiang, P.F Laeseke, C.F Dietrich, A Sannino, M Solbiati, E Nocerino, C.L Brace, and F.T. Lee. "Microwave ablation in primary and secondary liver tumours: technical and clinical approaches". In: *International journal of hyperthermia* 33.1 (2017), pp. 15–24.

[68] S Minaee, Y Boykov, F Porikli, A Plaza, N Kehtarnavaz, and D. Terzopoulos. "Image segmentation using deep learning: a survey". In: *arXiv* (2020). DOI: https://doi.org/10.48550/arXiv.2001.05566.

[69] G Narayanan and M.H. Doshi. "Irreversible electroporation (IRE) in renal tumors". In: *Current urology reports* 17 (2016), pp. 1–7.

[70] J. A Ottesen, D Yi, E Tong, M Iv, A Latysheva, C Saxhaug, K.D Jacobsen, A Helland, K.E Emblem, D. L Rubin, A Bjornerud, G Zaharchuk, and E. Grovik. "2.5D and 3D segmentation of brain metastases with deep learning on multinational MRI data". In: *Frontiers in Neuroinformatics* 16 (2023). DOI: https://doi.org/10.3389/fninf.2022.1056068.

[71] J Ouyang, S Liu, H Peng, H Garg, and D.N.H. Thanh. "LEA U-Net: a U-Net-based deep learning framework with local feature enhancement and attention for retinal vessel segmentation". In: *Complex & Intelligent Systems* 9.6 (2023), pp. 6753–6766.

[72] S Petrolini M. Cagnoni and M. Mordonini. "Automatic Detection of Sensitive Data Using Transformer- Based Classifiers". In: *Future Internet* 14 (July 2022), p. 228. DOI: 10.3390/fi14080228.

[73] M.A Phillips, R Narayan, T Padath, and B. Rubinsky. "Irreversible electroporation on the small intestine". In: *British journal of cancer* 106.3 (2012), pp. 490–495.

[74] R Qasrawi, L Silve, F Burdio, Z Abdeen, and A. Ivorra. "Anatomically realistic simulations of liver ablation by irreversible electroporation: impact of blood vessels on ablation volumes and undertreatment". In: *Technology in cancer research & treatment* 16.6 (2017), pp. 783–792.

[75] G Ramadori and S Cameron. "Effects of systemic chemotherapy on the liver". In: *Annals of hepatology* 9.2 (2010), pp. 133–143.

[76] F Richard and L. Cohen. "A New Image Registration Technique With Free Boundary Constraints: Application To Mammography". In: ().

[77] K.S. Roberts. "A new representation for a line". In: (1988), pp. 635–636.

[78] O Ronneberger, P Fischer, and T. Brox. "U-Net: Convolutional networks for biomedical image segmentation". In: *arXiv* (2015).

[79] C Rosazza, S Haberl Meglic, A Zumbusch, M.-P Rols, and D. Miklavcic. "Gene electrotransfer: a mechanistic perspective". In: *Current gene therapy* 16.2 (2016), pp. 98–129.

[80] H.R Roth, L Lu, N Lay, A.P Harrison, Farag A, A Sohn, and R.M. Summers. "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation". In: *Medical Image Analysis* 45 (2018), pp. 94–107. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2018.01.006.

[81] N. Rousselle J.-J. Vincent and N. Verbeke. "Genetic Algorithm to Set Active Contour". In: 2756 (Aug. 2003), pp. 345–352. DOI: 10.1007/978-3-540-45179-2_43.

[82] R Sacco, V Mismas, S Marceglia, A Romano, L Giacomelli, M Bertini, G Federici, S Metrangolo, G Parisi, E Tumino, G Bresci, A Corti, M Tredici, M Piccinno, L Giorgi, C Bartolozzi, and I. Bargellini. "Transarterial radioembalization for hepatocellular carcinoma: an update and perspectives". In: *World Journal of Gastroentology* (2015), pp. 6518–6525. DOI: https://doi.org/10.3748/wjg.v21.i21.6518.

[83] Y Sang and D. Ruan. "4D-CBCT registration with a FBCT-derived plug-and-play feasibility regularizer". In: (2021), pp. 108–117.

[84] D Schizas, N Charalampakis, C Kole, P Economopoulou, E Koustas, E Gkotsis, D Ziogas, A Psyrri, and M.V. Karamouzis. "Immunotherapy for pancreatic cancer: A 2020 update". In: *Cancer treatment reviews* 86 (2020), p. 102016.

[85] M Schmittbuhl, D Turgeon, D Matenine, and J.F. Matern. "Principes de l'imagerie cone beam CT". In: *Journal d'imagerie diagnostique et interventionnelle* 2.6 (2019), pp. 294–299. ISSN: 2543-3431. DOI: https://doi.org/10.1016/j.jidi.2019.07.009. URL: https://www.sciencedirect.com/science/article/pii/S2543343119301162.

[86] B Denis de Senneville, C Zachiu, M Ries, and C. Moonen. "EVolution: an edge-based variational method for non-rigid multi-modal image registration". In: *Physics in Medicine & Biology* 61.20 (2016), p. 7377.

[87] American Cancer Society. "Cancer Facts and Figures 2024". In: (2024).

[88] H Spallek, P Bischoff, W Zhou, F de Terlizzi, F Jakob, and A. Kovàcs. "Percutaneous electrochemotherapy in primary and secondary liver malignancies–local tumor control and impact on overall survival". In: *Radiology and Oncology* 56.1 (2022), pp. 102–110.

[89] C Springfeld, D Jäger, M.W Büchler, O Strobel, T Hackert, D.H Palmer, and J.P. Neoptolemos. "Chemotherapy for pancreatic cancer". In: *La Presse Medicale* 48.3 (2019), e159–e174.

[90] H.D Tagare, D Groisser, and O. Skrinjar. “Symmetric non-rigid registration: A geometric theory and some numerical techniques”. In: *Journal of Mathematical Imaging and Vision* 34 (2009), pp. 61–88.

[91] F.E.F Timmer, B Geboers, S Nieuwenhuizen, E.A.C Schouten, M Dijkstra, J.J.J de Vries, M.P van den Tol, M.R Meijerink, and H.J. Scheffer. “Locoregional treatment of metastatic pancreatic cancer utilizing resection, ablation and embolization: a systematic review”. In: *Cancers* 13.7 (2021), p. 1608.

[92] B Trotovšek, M Djokić, M Čemažar, and G. Serša. “New era of electrochemotherapy in treatment of liver tumors in conjunction with immunotherapies”. In: *World Journal of Gastroenterology* 27.48 (2021), p. 8216.

[93] V Varatharasan, H.-S Shin, A Tsourdos, and N. Colosimo. “Improving learning effectiveness for object detection and classification in cluttered backgrounds”. In: (2019), pp. 78–85.

[94] A Vincent, J Herman, R Schulick, R. H Hruban, and M. Goggins. “Pancreatic cancer”. In: *Lancet* (2011), pp. 607–620. DOI: https://doi.org/10.1016/S0140-6736(10)62307-0.

[95] V Vishnevskiy, T Gass, G Szekely, C Tanner, and O Goksel. “Isotropic total variation regularization of displacements in parametric image registration”. In: *IEEE transactions on medical imaging* 36.2 (2016), pp. 385–395.

[96] W Wei, A Chengfeng, Y Zhao, and G. Zhang. “Image registration algorithm based on super pixel segmentation and SURF feature points”. In: (2018), pp. 1–5.

[97] J Wulf, U Guckenberger M. Haedinger, U Oppitz, G Mueller, K Baier, and M. Flentje. “Stereotactic radiotherapy of primary liver cancer and hepatic metastases”. In: *Acta oncologica* 45.7 (2006), pp. 838–847.

[98] L Xu, W Ouyang, M Bennamoun, F Boussaid, and D. Xu. “Multi-class token transformer for weakly supervised semantic segmentation”. In: *arXiv* (2022). DOI: https://doi.org/10.48550/arXiv.2203.02891.

[99] Y Yuan, X Chen, X Chen, and J. Wang. “Segmentation transformer: object-contextual representations for semantic segmentation”. In: *arXiv* (2021). DOI: https://doi.org/10.48550/arXiv.1909.11065.

[100] Piven J Hazlett H. C Smith R. G Ho S Gee J.C Gerig G. Yushkevich P. A. “User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability”. In: *Neuroimage* 31.3 (2006), pp. 1116–1128.

[101] H Zhang Y. Liu and Q. Hu. "TransFuse: fusing transformers and CNNs for medical image segmentation". In: *arXiv* (2021). DOI: https://doi.org/10.48550/arXiv.2102.08005.

[102] T Zhang, N.P Orton, T.R Mackie, and B.R. Paliwal. "A novel boundary condition using contact elements for finite element based deformable image registration". In: *Medical physics* 31.9 (2004), pp. 2412–2415.

[103] Y Zhang, Y Lei, R. L. J Qiu, T Wang, H Wang, A. B Jani, W. J Curran, P Patel, T Liu, and X. Yang. "Multi-needle localization with attention U-Net in US guided HDR prostate brachytherapy". In: *Medical physics* 47 (2020), pp. 2735–2745. DOI: https://doi.org/10.1002/mp.14128.

[104] Y Zhang K. Li, W Zuo, L Zhang, L Van Gool, and R. Timofte. "Plug-and-play image restoration with deep denoiser prior". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2021), pp. 6360–6376.

[105] Y Zheng, X Sui, Y Jiang, T Che, S Zhang, J Yang, and H. Li. "SymReg-GAN: symmetric image registration with generative adversarial networks". In: *IEEE transactions on pattern analysis and machine intelligence* 44.9 (2021), pp. 5631–5646.

[106] X.-D Zhou and Z.-Y. Tang. "Cryotherapy for primary liver cancer". In: 14.2 (1998), pp. 171–174.

[107] X.-D Zhu, Z.-Y Tang, and H.-C. Sun. "Targeting angiogenesis for liver cancer: past, present, and future". In: *Genes & Diseases* 7.3 (2020), pp. 328–335.

[108] H Zou, F Wang, J.-J Zhou, X Liu, Q He, C Wang, Y.-W Zheng, Y Wen, and L. Xiong. "Application of photodynamic therapy for liver malignancies". In: *Journal of gastrointestinal oncology* 11.2 (2020), p. 431.