

*ÉCOLE DOCTORALE MATHÉMATIQUES, SCIENCES DE L'INFORMATION ET DE
L'INGÉNIEUR*

Laboratoire ICube – UMR 7357

THÈSE présentée par :
William Brice NDZIMBONG

soutenue le: 17 Décembre 2024

pour obtenir le grade de :
Docteur de l'Université de Strasbourg

Discipline/S spécialité : Informatique

**Recalage Automatique des Images
Echographiques Tridimensionnelles et
Tomodensitométries du Rein**

THÈSE dirigée par :

M. GEORGE Daniel
M. THOME Nicolas

Professeur des universités, Université de Strasbourg
Professeur des universités, Sorbonne Université

CO-ENCADRANT :

M. COLLINS Toby

PhD - Directeur de la recherche, IRCAD

RAPPORTEURS :

Mme CHAMBON Sylvie
M. SDIKA Michael

Professeure des universités, Université de Toulouse
Ingénieur de Recherche - HDR, Université de Lyon

EXAMINATEUR :

M. BARTOLI Adrien

Professeur des universités, Université Clermont Auvergne

Remerciements

Aux membres de mon jury de thèse :

- Daniel George et Nicolas Thome, mes directeurs de thèse, dont l'orientation, le soutien et les conseils à mon endroit, n'ont jamais fait défaut.
- Toby Collins, qui a co-encadré cette thèse avec un immense dévouement, et a fortement contribué à maintenir ma motivation. Merci beaucoup.
- Sylvie Chambon et Michael Sdika, qui ont chacun manifesté un grand intérêt pour mon travail de thèse, et ont accepté de l'évaluer à travers leurs expertises dans le domaine du recalage. Merci beaucoup.
- Adrien Bartoli, qui s'est intéressé à ce travail depuis le début et me fait aussi l'honneur de l'évaluer, du haut de sa grande expérience dans le domaine du recalage. Je me sens honoré.

Je remercie les représentations France et Africa de l'Institut de Recherche contre les Cancers de l'Appareil Digestif (IRCAD), pour le financement de cette thèse, le soutien matériel, et l'accueil au sein de l'équipe Surgical Data Science.

Je remercie la Région Grand-Est de France, pour le financement de mes travaux au cours de cette thèse.

Je remercie le laboratoire ICUBE de l'Université de Strasbourg pour l'accueil et l'encadrement de cette thèse.

Je remercie l'équipe Surgical Data Science (IRCAD France - IRCAD Africa) et son Responsable, Alexandre Hostettler. Travailler parmi vous a été très stimulant et plaisant.

Je remercie Flavien Bridault, David Sixela, Alex Ancel, Erwan Duhamel, Yvonne, Günther Saibro, Mathieu Haller, Loïc Themyr, Josiane, Marc, Didier, Richard, Grâce, Bereket, Michael Kugler, Patrick, Stephanie, Jean, Florian, Gael, François, Sierra, ..., J'en oublie certainement.

Je remercie Ma Famille, pour la chaleur au delà de la distance.

Je remercie Mariane Tchudje, Louis Fippo, Nathalie Nkwuimi, Xavier Ouogue, Marcelle Ngo Momha, Christian Tayou, Cyrille Chopkeng, Gabin Nguenang, Ibrahim Mbouandi.

Je remercie Emilie Peynaud et Wilson Toussile.

Je remercie Charles Awono Onana, Jacques Tagoudjeu, Jean-Jules Tewa.

Je remercie Joseph Ngwene.

Résumé

Titre : Recalage Automatique des Images Echographiques Tridimensionnelles et Tomodensitométriques du Rein.

Mots clés : Recalage Inter-modal ; Echographie ; Apprentissage Profond ; Guidage Interventionnel ; Rein

Résumé : Le recalage automatique des images échographiques abdominales et des images tomodensitométriques (scanner) est essentiel pour améliorer le guidage interventionnel en chirurgie rénale. Toutefois, il représente encore un défi de recherche majeur. Une limitation importante est le manque de jeu de données publiques contenant des images des deux modalités pour un même patient (jeu de données appariées). Cette absence freine les avancées méthodologiques et empêche une comparaison systématique des méthodes de pointe. Une autre limitation importante est qu'il existe peu de méthodes robustes capables de réaliser un recalage sans intervention manuelle, notamment en ce qui concerne les méthodes dites « globales ». Cette thèse vise à surmonter ces obstacles par plusieurs contributions de recherche. La première est la création d'un nouveau jeu de données, composé d'images échographiques tridimensionnelles transabdominales et scanner appariées, provenant de reins de 48 patients humains. Ce jeu de données inclut des annotations de segmentation et de points de repère anatomiques, validées par deux radiographes expérimentés. En plus de fournir ces données, la concordance des annotations est analysée, et la valeur du jeu de données est démontrée en évaluant des méthodes qui traitent deux tâches fondamentales : la segmentation automatique des reins et le recalage d'images inter-modales. Les résultats montrent que ces deux défis restent ouverts, et ce jeu de données constitue une ressource importante pour faire progresser ces domaines. La seconde contribution principale est une méthode automatique de recalage global pour les images échographiques tridimensionnelles et scanner des reins. Cette méthode surmonte l'ambiguïté de recalage due à la symétrie naturelle de l'organe et, combinée à un algorithme d'affinement, permet un recalage précis, robuste et sans initialisation manuelle. Cette méthode a également d'autres applications importantes, telles que la conversion inter-modale d'images, la synthèse d'images, et le transfert d'annotations entre modalités.

Title : Automatic Registration of Three-Dimensional Ultrasound and Computed Tomography Images of the Kidney.

Keywords : Inter-modal Registration ; Ultrasound ; Deep Learning ; Interventional guidance ; Kidney

Abstract : Automatic registration between abdominal ultrasound (US) and computed tomography (CT) images is needed to enhance interventional guidance in kidney surgery. However, it remains an open research challenge. One striking limitation is the lack of public datasets that comprise images of the same patient in both modalities (paired datasets). This has hindered methodological progress, as well as prevented a systematic comparison of state-of-the-art methods. Another limitation is the lack of robust methods capable of solving registration without manual initialization (‘global’ methods). This thesis aims to overcome these challenges with several research contributions. The first contribution is a novel dataset with paired transabdominal 3D US and CT kidney images from 48 human patients that includes segmentation and anatomical landmark annotations from two expert radiographers. In addition to the dataset, annotation consistency is analyzed, and its value assessed by benchmarking methods that tackle two fundamental tasks : automatic kidney segmentation and inter-modal image registration. The findings show that both challenges are still open, and the dataset should serve as an important resource for advancing both topics. As a second main contribution, an automatic method for global registration of kidneys in 3D US and CT images is proposed. This method handles registration ambiguity caused by the organ’s natural symmetry. Combined with a registration refinement algorithm, it achieves robust and accurate kidney registration while avoiding manual initialization. The method has several other important applications, including inter-modal image translation and image synthesis, as well as label transfer between modalities.

Publications

"Global registration of kidneys in 3D ultrasound and CT images". **W. Ndzimbong**, N. Thome, C. Fourniol, Y. Keeza, B. Sauer, J. Marescaux, D. George, A. Hostettler et T. Collins. In : International Journal of Computer Assisted Radiology and Surgery (2024), pages 1-11. (Disponible sur : <https://rdcu.be/dTffd>)

"TRUSTED : The Paired 3D Transabdominal Ultrasound and CT Human Data for Kidney Segmentation and Registration Research". **W. Ndzimbong**, C. Fourniol, L. Themyr, N. Thome, Y. Keeza, B. Sauer, P.-T. Piechaud, A. Mejean, J. Marescaux, D. George, D. Mutter, A. Hostettler, T. Collins. In : Nature - Scientific Data. (Accepté pour publication).

"Global Registration of Kidneys in 3D Ultrasound and CT images", **W. Ndzimbong**, N. Thome, C. Fourniol, Y. Keeza, B. Sauer, J. Marescaux, D. George, A. Hostettler, T. Collins. In : CARS - Computer Assisted Radiology and Surgery, 38th International Congress and Exhibition, June 18 - 21, 2024, Barcelona, Spain.

Table des matières

Remerciements	3
Résumé	5
Publications	7
Table des matières	8
Liste des figures	11
Liste des tableaux	13
1 INTRODUCTION GÉNÉRALE	15
1.1 La problématique principale	15
1.2 Motivations et enjeux	16
1.3 Méthodes et défis	16
1.4 Objectifs et contributions de la thèse	17
1.5 Le plan de la thèse	18
2 CONTEXTE ET ÉTAT DE L'ART	21
Résumé du chapitre	21
2.1 Contexte clinique et motivation	22
2.1.1 Procédures percutanées guidées par échographie - Fusion d'images	22
2.1.2 Le contexte spécifique de la chirurgie rénale percutanée guidée par échographie	26
2.1.3 Recalage multimodal et applications cliniques	28
2.2 Considérations et défis dans le recalage d'images médicales . . .	33
2.2.1 Comment modéliser la transformation spatiale de recalage ?	33
2.2.2 Comment gérer les variabilités anatomique et pathologique	37
2.2.3 Comment obtenir des jeux de données pour le développe- ment de modèles d'apprentissage ?	38
2.2.4 Comment modéliser l'information dans les images médicales ?	39
2.2.5 Recalage global ou affinement du recalage ?	43
2.2.5.1 Recalage global et principaux défis.	44
2.2.5.2 Affinement du recalage et principaux défis.	45

2.2.6	Recalage automatique ou recalage interactif ?	45
2.2.7	Comment valider la précision du recalage ?	46
2.3	Méthodes de pointe pour le recalage multimodal d'images échographiques	48
2.3.1	Méthodes traditionnelles de recalage multimodal d'images échographiques	48
2.3.1.1	Les méthodes globales	48
2.3.1.2	Les méthodes d'affinement du recalage	51
2.3.2	Méthodes d'apprentissage profond appliquées au recalage multimodal	55
2.3.2.1	Les méthodes globales	56
2.3.2.2	Les méthodes d'affinement du recalage	61
2.3.2.3	Les métriques de similarité apprises	67
2.3.3	Conclusion	68
3	LE JEU DE DONNÉES TRUSTED	71
	Résumé du chapitre	71
3.1	Introduction	72
3.2	Le jeu de données TRUSTED	75
3.2.1	Acquisition des images	75
3.2.2	Annotation	76
3.2.3	Disponibilité du jeu de données	79
3.3	Analyse de la concordance	81
3.4	Validation du jeu de données pour la formation de modèles de segmentation automatique	84
3.4.1	Les modèles et la méthodologie d'évaluation	84
3.4.2	Résultats quantitatifs de segmentation automatique	86
3.4.3	Résultats qualitatifs de segmentation automatique	88
3.5	Validation du jeu de données pour la formation de modèles de recalage	92
3.5.1	Méthodologie d'évaluation	92
3.5.2	Résultats quantitatifs de recalage	94
3.5.3	Résultats qualitatifs de recalage	102
3.6	Conclusion	104
4	RECALAGE GLOBAL AUTOMATIQUE DES IMAGES ÉCHOGRAPHIQUES 3D ET TOMODENSITOMÉTRIQUES DU REIN	105
	Résumé du chapitre	105
4.1	Recalage global automatique utilisant les surfaces	106
4.1.1	L'information de surface	106

4.1.2	Recalage global automatique utilisant l'Analyse en Composantes Principales (ACP) des géométries	108
4.1.3	Recalage global automatique utilisant des descripteurs locaux de surface	109
4.1.4	Expériences	110
4.1.4.1	Les métriques d'évaluation	111
4.1.4.2	Résultats quantitatifs	112
4.1.4.3	Résultats qualitatifs	117
4.1.5	Conclusion et discussion	120
4.2	Une approche de recalage global automatique par points de repère étiquetés	121
4.2.1	Objectifs et méthode générale	121
4.2.2	Les données	123
4.2.3	Localisation et étiquetage des points de repère	124
4.2.4	Génération d'hypothèses de recalage global	127
4.2.5	Résolution des ambiguïtés des hypothèses de recalage	131
4.2.6	Expériences	133
4.2.6.1	Configuration des expérimentations	133
4.2.6.2	Résultats quantitatifs	135
4.2.6.3	Résultats qualitatifs	140
4.3	Conclusion	142
5	CONCLUSIONS ET PERSPECTIVES	145
	CONCLUSIONS ET PERSPECTIVES	145
5.1	Synthèse des contributions principales	145
5.2	Perspectives de recherche	146
	Bibliographie	149

Liste des figures

Note : Sauf mention contraire (référence indiquée), toutes les figures de ce rapport sont issues des travaux réalisés dans le cadre de ce doctorat en collaboration avec l'IRCAD.

2.1	Station échographique	24
2.2	Échographe sans fil	24
2.3	Échographe de poche	24
2.4	Guidage échographique "freehand"	24
2.5	Dispositif de guidage d'aiguille	24
2.6	Séquence d'ablation guidée par échographie	25
2.7	Limitations de l'échographie rénale	26
2.8	Représentation simplifiée du rein [28].	27
2.9	Les reins dans la cavité abdominale vue de dos	27
2.10	Incidence mondiale du cancer du rein en 2020	28
2.11	Images échographiques 3D avec balayage mécanique	29
2.12	Reins en imagerie CT et US 3D	30
2.13	Prostate en IRM et US 3D	31
2.14	Lésion du côlon en PET et échographie	31
2.15	Recalage rigide des vaisseaux hépatiques	35
2.16	Transformations rigides, affines et déformables	36
2.17	Séries de formes simulées de prostate	38
2.18	Caractéristiques FPFH pour recalage	49
2.19	Recalage global de volumes de foie avec ACP	50
2.20	Génération d'un jeu de données simulant des positions de sonde échographique et celles des structures vasculaires du foie	51
2.21	Pipeline de segmentation et recalage de lignes de vaisseaux hépatiques	53
2.22	Sélection de points caractéristiques du foie	55
2.23	Correspondances de points clés dans le jeu de données 3Dircadb	57
2.24	Modèle LiverMatch pour appariement de points	57
2.25	Pipeline d'entraînement du modèle DIP	58
2.26	Transformer pour l'extraction et l'appariement de caractéristiques abstraites	60
2.27	Points de repère anatomiques de la prostate pour recalage	61
2.28	Préparation de données pour ProRegNet	62
2.29	Architecture de ProRegNet	62

2.30 Architecture Free Point Transformer	63
2.31 Modèle général guidé par labels	64
2.32 Réseau de prédiction de transformations globale et locale	65
2.33 Architecture de VoxelMorph	66
2.34 Architecture de TransMorph	67
2.35 CNN pour apprentissage de métrique de similarité	68
3.1 Paire IRM-US avec annotations de prostate (MuRegPro)	73
3.2 Paire IRM-US avec points de repère anatomiques de prostate	74
3.3 Marquage de points de repère dans le volume échographique	77
3.4 Annotations des images dans le jeu de données TRUSTED	79
3.5 Segmentation automatique du rein dans les images CT de TRUSTED	89
3.6 Segmentation automatique du rein dans les images US 3D de TRUSTED	91
3.7 Positions relatives des reins après recalage initial perturbé	99
3.8 Sensibilité des méthodes d’affinement aux erreurs d’initialisation	101
3.9 Résultats qualitatifs de recalage avec le jeu de données TRUSTED	103
4.1 Segmentation automatique du rein et nuage de points issus d’images US 3D	107
4.2 Nuages de points CT-US avec repères géométriques	108
4.3 Recalage de nuages de points par ACP	109
4.4 Appariement de surfaces du rein CT et US avec DIP	110
4.5 Performances des méthodes de recalage global utilisant les surfaces, combinées à l’affinement par BCPD	113
4.6 Visualisations en damier des résultats de recalage, partie 1	118
4.7 Visualisations en damier des résultats de recalage, partie 2	119
4.8 Pipeline de recalage global automatique du rein	123
4.9 Vues 3D des points de repère rénaux dans les images CT et US 3D	124
4.10 Coupes sagittales du rein et étiquetage des points de repère	125
4.11 Points de repère rénaux détectés et étiquetés dans le CT	126
4.12 Points de repère rénaux détectés et étiquetés dans l’US	127
4.13 Illustration du fonctionnement de FlipNet	133
4.14 Performances des méthodes automatiques de recalage global, com- binées à BCPD	136
4.15 Résultats qualitatifs de recalage global automatique + BCPD, partie 1	141
4.16 Résultats qualitatifs de recalage global automatique + BCPD, partie 2	142

Liste des tableaux

2.1	Forces et faiblesses des différentes approches de recalage d'images médicales.	42
3.1	Statistiques de concordance inter-annotateurs pour la segmentation des reins	83
3.2	Statistiques de concordance inter-annotateurs pour les points de repère rénaux	83
3.3	Résultats quantitatifs de segmentation automatique de reins dans les images CT et US 3D	87
3.4	Résultats quantitatifs des méthodes d'affinement du recalage	97
4.1	Comparaison quantitative des méthodes automatiques de recalage global utilisant les surfaces, combinées avec l'affinement BCPD . . .	115
4.2	Comparaison quantitative des méthodes automatiques de recalage global utilisant les surfaces, mais sans affinement	116
4.3	Comparaison quantitative des méthodes de recalage global combinées avec l'affinement BCPD	138
4.4	Comparaison quantitative des méthodes de recalage global sans affinement	139

1. INTRODUCTION GÉNÉRALE

La chirurgie percutanée guidée par l'imagerie a transformé de manière significative les pratiques cliniques (accès aux structures internes à travers la peau), dont les procédures rénales. L'accès mini-invasif qu'elle offre améliore les résultats pour les patients en réduisant les temps de récupération et en minimisant les complications post-opératoires. Cependant, la précision des procédures percutanées dépend en grande partie de la capacité des praticiens à visualiser correctement les organes et les structures anatomiques, afin de guider précisément les instruments dans les zones ciblées. Parmi les techniques d'imagerie disponibles, l'échographie est particulièrement prisée pour la sécurité qu'elle offre, son coût abordable et son accessibilité en temps réel. Toutefois, l'échographie présente des limitations inhérentes en termes de qualité d'image, ainsi que des dépendances vis-à-vis de l'opérateur, en particulier, pour le rein qui est sujet à des variations de position et de morphologie. Ce qui rend parfois difficile une visualisation optimale de ce dernier dans un contexte chirurgical. Afin de pallier ces contraintes, l'utilisation de l'imagerie de fusion – technique qui consiste à combiner différentes images grâce à leur alignement spatial ou recalage précis – est devenue un axe de recherche prometteur. Les images échographiques peuvent par exemple être recalées avec des images de tomodensitométrie (ou scanner). Ce qui permet de combiner les atouts des deux modalités. En effet, alors que l'échographie fournit des images en temps réel, la tomodensitométrie apporte une vision anatomique précise et globalement invariable. Pourtant, le recalage de ces deux modalités d'imagerie reste un défi technique.

1.1 La problématique principale

Les techniques de recalage échographie-scanner ont démontré une certaine maturité dans le cadre d'applications cliniques comme le diagnostic ou la planification. Cependant, ces approches peinent encore à s'imposer dans des interventions nécessitant une imagerie en temps réel, comme une ablation par radiofréquence ou une biopsie rénale ciblée. En effet, les méthodes actuelles de recalage précis du rein dépendent encore, au moins partiellement, de l'intervention humaine. Leur intégration telle quelle dans des applications d'imagerie en temps réel ralentirait l'exécution et générerait un inconfort pour le patient et le praticien. Ceci pose le problème de l'automatisation complète

des approches de recalage précis d'images échographiques et scanner de rein. Apporter des solutions à cette problématique est essentiel pour l'adoption des méthodes dans ces contextes cliniques spécifiques.

1.2 Motivations et enjeux

L'automatisation complète du recalage précis des images échographiques et scanner du rein constituerait une avancée majeure dans l'amélioration des soins apportés aux patients. Notamment avec l'échographie tridimensionnelle (3D), qui offre l'information de profondeur, comme un scanner. Un recalage automatisé rapide permet de réduire le temps de l'intervention. Il augmente également l'efficacité pour la localisation des petites lésions lors d'interventions ciblées guidées en temps réel, comme la biopsie des masses suspectes. Ce qui permet d'éviter les prélèvements non représentatifs. En termes de potentiel pour la recherche sur les interventions assistées par intelligence artificielle, l'automatisation du recalage permettrait d'intégrer des systèmes intelligents capables d'analyser les images en temps réel, de suggérer des trajectoires optimales et de signaler les risques en fonction de la topographie des structures. De plus, l'automatisation permet la standardisation du recalage et augmente la reproductibilité des résultats. Enfin, en termes de démocratisation de l'échographie, l'automatisation complète du recalage, réduirait la forte dépendance des pratiques de guidage par imagerie, aux compétences avancées des opérateurs, pour un guidage précis. Même des praticiens moins spécialisés en imagerie pourraient obtenir des visualisations précises et fiables, élargissant l'usage de l'échographie dans des contextes plus variés, y compris des centres de soins moins équipés. Cela contribuerait ainsi à la démocratisation de cette technologie, rendant des diagnostics et des interventions mini-invasives plus accessibles et de haute qualité pour un plus grand nombre de patients.

1.3 Méthodes et défis

Afin d'automatiser les systèmes de recalage précis d'images échographiques et scanner du rein, cette thèse propose d'y intégrer des modèles d'intelligence artificielle à base de réseaux de neurones profonds. Ainsi, des modèles neuronaux ont été formés pour remplacer les étapes où l'intervention humaine était encore nécessaire, et permettre une transition automatique fluide entre les étapes de ces systèmes.

L'intégration de réseaux de neurones profonds dans les systèmes de recalage d'images échographiques et scanner du rein nécessite des données d'entraînement. Il s'agit d'exemples fournis aux modèles afin qu'ils apprennent la tâche que l'on souhaite résoudre. Or, il n'existe pas, dans le domaine public, de jeu de données pouvant être utilisé à cette fin. Le second défi est le risque important qu'une méthode de recalage conduise à des solutions erronées. La complexité du problème du recalage précis de modalités différentes, la nature bruitée des images échographiques et la forme symétrique du rein, sont autant d'éléments augmentant la possibilité même pour un algorithme performant de se tromper. Le troisième défi est celui de l'évaluation des approches de recalage. Quand bien même des données seraient disponibles pour l'évaluation, il est en général très difficile de les avoir en état d'alignement précis pouvant servir de vérité terrain. Voilà les défis principaux auxquels les travaux conduits dans le cadre de cette thèse, font face.

1.4 Objectifs et contributions de la thèse

L'objectif principal de cette thèse est de concevoir, développer et évaluer des méthodes entièrement automatiques pour aligner précisément des images échographiques tridimensionnelles (3D) et des images scanner du rein. Ce travail identifie les défis spécifiques de cette tâche et propose des pistes de recherche pour les surmonter. De plus, cette thèse ouvre la voie à de nouvelles approches méthodologiques en mettant en lumière les obstacles et en suggérant des stratégies pour les résoudre.

L'atteinte de cet objectif principal se décline en plusieurs objectifs spécifiques qui représentent les contributions méthodologiques de cette thèse :

1. Stimuler la recherche pour les méthodes automatiques de segmentation et de recalage multimodal, en proposant un jeu de données publiques unique, développé en collaboration avec des radiologues et des radiologistes. Il est constitué d'images échographiques transabdominales tridimensionnelles et scanner de reins de 48 patients, ainsi que d'annotations précises. Les images brutes ont été collectées, anonymisées, inspectées et stockées dans un format standard. Un protocole strict a été élaboré pour définir des annotations et décrire comment les obtenir. Elles contiennent des segmentations et des points de repère anatomiques de rein dans les images. Les vérités terrain ont été estimées à partir de doubles annotations indépendantes, puis vérifiées et évaluées quantitativement. L'utilité du jeu de données pour la recherche en segmentation automatique a été me-

surée en entraînant et évaluant 4 modèles de pointe dont 3 réseaux de neurones convolutifs et un "Vision Transformer". L'utilité du jeu de données pour la recherche en recalage multimodal a été mesurée en évaluant des méthodes de pointe, dont 2 utilisant les surfaces, une iconique, et un "Vision Transformer". Enfin, une analyse de la sensibilité des méthodes de recalage aux erreurs d'initialisation a été conduite, mettant en évidence les défis spécifiques liés à l'initialisation imprécise du recalage.

2. Proposer le premier système complètement automatique de recalage précis d'images échographiques tridimensionnelles et scanner de rein, en combinant une initialisation (recalage global) automatique et un affinement robuste. Dans une première expérimentation, des segmentations automatiques du rein ont été utilisées pour extraire des caractéristiques de surface et générer des hypothèses de recalage initial, avec un succès limité. Cette expérimentation a révélé la difficulté d'obtenir une initialisation stable, en raison des ambiguïtés causées par la forme symétrique du rein, et des limitations des modèles de segmentation automatique. Ces derniers ne fournissent pas toujours les détails nécessaires, de plus, les vérités terrain servant à leur entraînement sont difficiles à obtenir. Pour surmonter ces obstacles, une nouvelle approche robuste aux ambiguïtés de recalage et nécessitant moins de supervision a été proposée. La méthode s'appuie sur le jeu de données précédemment évoqué. Elle utilise un petit nombre de points de repère anatomiques coplanaires (7 points), localisés automatiquement dans les images, pour générer deux hypothèses de recalage global correspondant aux ambiguïtés dues à la symétrie des pôles supérieur et inférieur du rein. Un réseau de neurones de segmentation a ensuite été adapté pour distinguer l'hypothèse correcte parmi les deux. Cette initialisation, combinée à un affinement robuste, constitue un système complètement automatique pour le recalage précis des images échographiques tridimensionnelles et scanner du rein, avec des performances comparables aux systèmes semi-automatiques (initialisation manuelle et affinement automatique).

1.5 Le plan de la thèse

La suite de ce document est organisée ainsi :

Le Chapitre 2 : Contexte et état de l'art, présente le cadre général du recalage des images échographiques et scanner, avec comme applications les procédures rénales percutanées. Puis, il analyse la littérature en matière de

recalage multimodal appliqué aux images échographiques.

Le Chapitre 3 : Le jeu de données TRUSTED, présente et analyse le premier jeu de données publiques constitué d'images échographiques 3D transabdominales de reins de 48 patients humains, d'images scanner appariées, ainsi que d'annotations précises dans les deux modalités destinées aux méthodes de recalage et de segmentation. Il est ensuite démontré l'utilité du jeu de données TRUSTED pour la recherche sur les systèmes automatiques de recalage multimodal précis d'images médicales, et pour la recherche sur les méthodes de segmentation automatique.

Le Chapitre 4 : Recalage global automatique du rein dans les images échographiques et tomodensitométriques, propose un système complètement automatique pour le recalage multimodal précis des images échographiques. L'approche associe à un affinement, une méthode globale (d'initialisation) faiblement supervisée et robuste aux ambiguïtés de recalage causées par la symétrie naturelle du rein.

Le Chapitre 5 : Conclusion et perspectives, rappelle brièvement les contributions de la thèse, présente les limitations identifiées au cours de ce travail et propose des pistes de recherche envisageables pour approfondir les contributions de cette thèse.

2. CONTEXTE ET ÉTAT DE L'ART

Résumé du chapitre

Ce chapitre présente le contexte clinique de cette recherche, en mettant en lumière l'importance du recalage multimodal et ses applications cliniques. Il se concentre sur les procédures guidées par échographie et les techniques de fusion d'images, en particulier pour la chirurgie rénale percutanée. Un alignement spatial précis entre l'imagerie échographique interventionnelle et d'autres modalités permet d'améliorer les diagnostics et les traitements.

Le chapitre aborde les considérations et les défis critiques en matière de recalage d'images médicales. Il relève l'importance de la modélisation de l'information contenue dans les images médicales, et discute des stratégies pour gérer les variabilités anatomique et pathologique. En décrivant les défis associés aux méthodes de recalage global et aux méthodes d'affinement du recalage, il met en avant les questions de recherche ouvertes, propres au recalage global. La discussion se poursuit vers les méthodes de recalage automatiques et interactives. Elle souligne également la nécessité de valider la précision du recalage afin de garantir des résultats fiables.

Dans l'avant-dernière section de ce chapitre (Section 2.3), les méthodes de pointe pour le recalage multimodal d'images échographiques sont analysées, incluant les techniques de recalage traditionnelles fondées sur des approches artisanales, et les méthodologies plus récentes utilisant l'apprentissage profond à différentes étapes des systèmes proposés.

Enfin, les limites des méthodes existantes applicables au recalage d'images rénales sont abordées, mettant en avant le fait que le recalage automatique du rein reste un défi non résolu. Ces limites sont particulièrement dues à la fenêtre d'acquisition restreinte des images échographiques et à la symétrie naturelle du rein ; ce qui rend difficile l'obtention d'un recalage global unique et stable. Le besoin urgent de disposer de jeux de données à grande échelle et annotés avec précision, pour la formation et l'évaluation de modèles d'apprentissage profond est également mis en avant. La recherche dans le domaine du recalage inter-modal d'images échographiques par apprentissage profond, souffre profondément de la rareté de base de données communes pouvant servir à l'entraînement, l'évaluation et la comparaison des modèles. Cette limitation met plus spécifiquement en avant l'importance des contributions apportées

dans cette thèse.

2.1 Contexte clinique et motivation

2.1.1 Procédures percutanées guidées par échographie - Fusion d'images

La prise en charge des cancers a beaucoup progressé avec l'arrivée de traitements percutanés mini-invasifs (accès aux structures internes à travers la peau), tels que la radiothérapie ou la cryothérapie. Ces approches permettent de mieux préserver les tissus sains, offrent un meilleur confort au patient, réduisent la durée d'hospitalisation et entraînent moins de complications que la chirurgie traditionnelle [1].

Une conséquence de l'usage de procédures percutanées est la perte d'une visualisation directe de la région d'intérêt. Cela implique, non seulement l'apprentissage de nouveaux gestes et repères pour les chirurgiens et radiologues interventionnels, mais aussi la nécessité d'un guidage. L'imagerie médicale simplifiait déjà le diagnostic, en informant sur une présence potentielle de tumeur, avant confirmation par biopsie (prélèvement d'échantillon de tissu organique pour analyse), en cas de risque avéré. Il était donc normal de penser à un guidage des procédures percutanées par l'imagerie médicale.

Le développement de procédures mini-invasives et la nécessité de leur guidage par l'imagerie médicale ont stimulé dans la communauté scientifique une grande variété d'activités de recherche autour de thèmes tels que :

- la génération automatique et en temps réel de modèles biomécaniques d'organes abdominaux, adaptés à chaque patient [2, 3, 4];
- la simulation numérique des déformations des tissus pour la planification d'interventions chirurgicales [5, 6];
- la prédiction de positions d'organes abdominaux pendant l'activité physiologique afin de réduire l'irradiation des tissus sains en radiothérapie [7];
- le développement de systèmes de guidage par réalité augmentée [8, 9], ou par recalage d'images [10, 11, 12], afin de fusionner l'information visuelle temps réel avec une complémentaire;
- plus récemment, l'introduction de modèles d'apprentissage profond (voir section 2.3.2) [13] pour l'exécution rapide et automatique de tâches importantes dans ces processus, comme la segmentation d'organes [14, 15] ou le diagnostic [16];

- et un ensemble de procédures chirurgicales robotiques guidées par l'échographie [17]

La sonde échographique est souvent qualifiée de "stéthoscope du chirurgien", car elle offre à ce dernier la capacité de visualiser les structures internes à des fins de diagnostic : examen de grossesse, imagerie des organes abdominaux (foie, reins, vésicule biliaire), évaluation cardiaque (échocardiogramme), détection de pathologies comme les tumeurs, les kystes, les abcès [17]. De plus, l'essor de la chirurgie mini-invasive a conduit à l'intégration de l'échographie dans les procédures percutanées comme le guidage d'aiguille (biopsie), de canule (drainage ou introduction d'instruments) ou de sonde (exploration, mesure, drainage), regroupées sous le terme d'échographie interventionnelle.

L'idée d'utiliser l'échographie pour guider des aiguilles a émergé dans les années 1960. En 1961, Berlyne [18] a proposé l'utilisation d'une sonde échographique pour guider une biopsie rénale. Les premières publications décrivant ces procédures sont apparues en 1972, avec les travaux de Holm et Goldberg [19, 20], qui ont conçu des transducteurs permettant l'insertion d'aiguilles via une perforation centrale. Depuis lors, l'échographie interventionnelle a connu un développement rapide avec une gamme de rôles de plus en plus large.

L'un des principaux avantages de l'échographie est la visualisation en temps réel, qui permet de contrôler l'insertion. De plus, le dispositif échographique est portable, qu'il s'agisse de station (Figure 2.1), de modèle sans fil (Figure 2.2) ou de modèle de poche (Figure 2.3) ; ce qui permet son utilisation même dans des conditions d'espace restreint ou en déplacement. Enfin, le dispositif échographique est moins coûteux et ne présente pas de risques de radiation en comparaison aux autres modalités d'imagerie telles que la tomodensitométrie encore appelée scanner (TDM ou CT), ou l'imagerie par Résonance Magnétique (IRM) [21]. Ce qui fait de l'échographie une modalité pratique en radiologie interventionnelle et largement accessible.

L'échographie permet deux méthodes de guidage : indirecte et directe. Le guidage indirect consiste à localiser le site d'intervention par échographie avant d'insérer l'aiguille sans assistance en temps réel. Ce procédé est moins précis que le guidage direct, lequel permet de visualiser l'aiguille tout au long de son trajet jusqu'à la cible. Cette seconde méthode est préférée en échographie interventionnelle. Le guidage direct peut lui-même être réalisé de deux manières distinctes. La première est la technique dite "à main levée" ou "freehand" (Figure 2.4) qui laisse une liberté et une flexibilité totales au praticien pour le positionnement et l'orientation, ce qui a également pour conséquence une variabilité dans la précision du geste. La seconde utilise un dispositif de guidage d'aiguille fixé (Figure 2.5), ce qui accorde moins de degré de liberté et moins

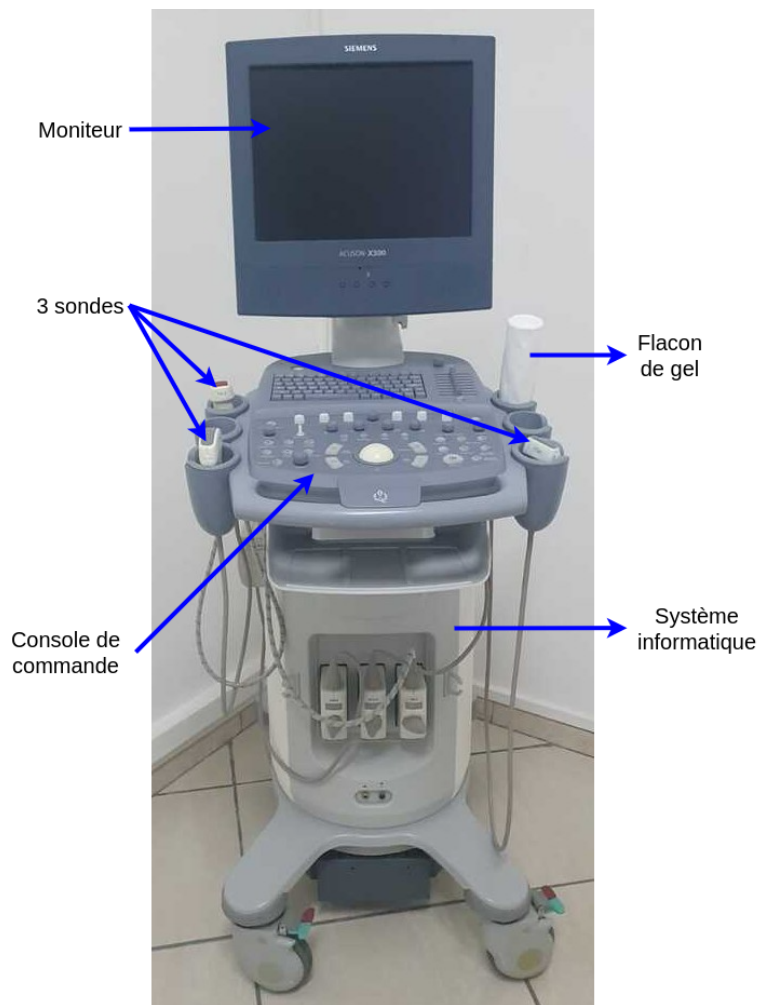


FIGURE 2.1 – Station échographique [22].



FIGURE 2.2 – Échographe avec sonde sans fil [23].



FIGURE 2.3 – Échographe de poche [24].

de variabilité dans le geste.



FIGURE 2.4 – Guidage échographique "freehand" [25].

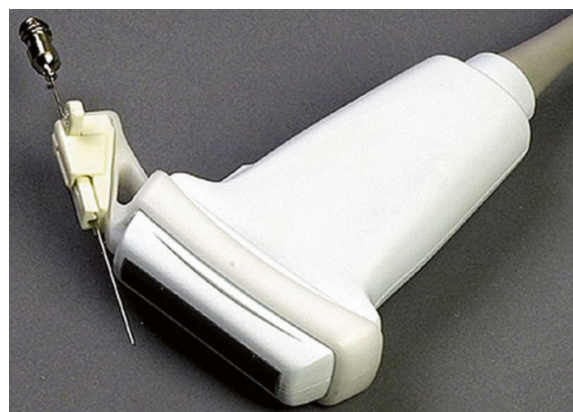


FIGURE 2.5 – Dispositif de guidage d'aiguille fixé sur la sonde [25].

L'échographie interventionnelle est utilisée dans plusieurs cas d'application. Elle permet la réalisation de biopsies. Il s'agit d'insérer une aiguille à travers la

peau jusqu'à la zone ciblée. Une fois l'aiguille en place, une seringue est utilisée pour aspirer un petit échantillon. Ce dernier est ensuite analysé au laboratoire pour poser un diagnostic, par exemple pour déterminer si une masse est bénigne ou maligne. L'échographie permet le drainage en guidant le placement du drain dans une cavité corporelle cible afin d'évacuer des fluides. Le drainage est généralement réalisé pour traiter des collections anormales de liquide, comme des abcès ou d'autres accumulations de liquide dans les tissus. L'échographie guide également l'ablation de tumeurs, comme le carcinome hépatocellulaire (Figure 2.6). Parmi les techniques d'ablation, il y a l'injection d'éthanol, la cryothérapie et l'ablation par micro-ondes. Ces procédures sont de plus en plus utilisées pour le traitement des tumeurs.

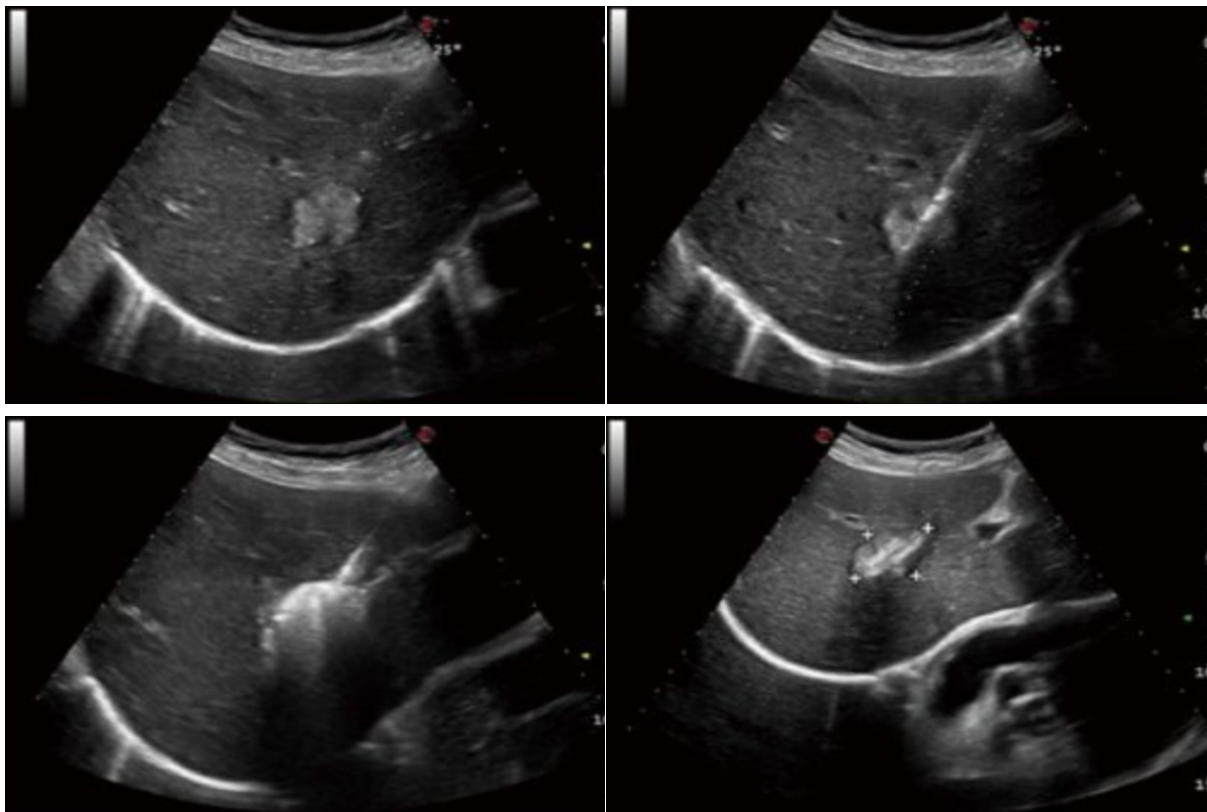


FIGURE 2.6 – Séquence d'ablation par micro-ondes percutanée guidée par échographie d'un carcinome hépatocellulaire de taille moyenne du lobe droit. De haut en bas et de gauche à droite : Évaluation échographique avant l'ablation ; Insertion de l'aiguille ; Effet d'ébullition hyperéchogène dans la zone d'ablation pendant la procédure ; Évaluation échographique un mois plus tard, la trace hyperéchogène interne correspond à la position de la sonde active[26].

L'échographie interventionnelle offre aux praticiens un moyen efficace et économique pour effectuer des diagnostics et des traitements. Son utilisation continue de croître, si bien qu'elle est devenue indispensable dans la gestion des maladies intra-abdominales. Cependant, l'image échographique présente

des limites. Tout d'abord, sa dépendance à la fenêtre acoustique rend parfois difficile la visualisation de structures profondes ou situées derrière des obstacles osseux, en raison de l'ombre acoustique. C'est par exemple le cas pour la visualisation des pôles supérieurs des reins (fenêtre limitée et ombre acoustique, voir Figures 2.7-A et B), ou la visualisation exhaustive du foie (fenêtre limitée). De plus, l'image échographique est bruitée, a un faible contraste, et est sensible à la présence de gaz intestinaux. Ce qui complique l'évaluation de certaines régions anatomiques, notamment au niveau de l'abdomen, et peut rendre difficile la visualisation précise de certaines lésions (Figure 2.7-C).

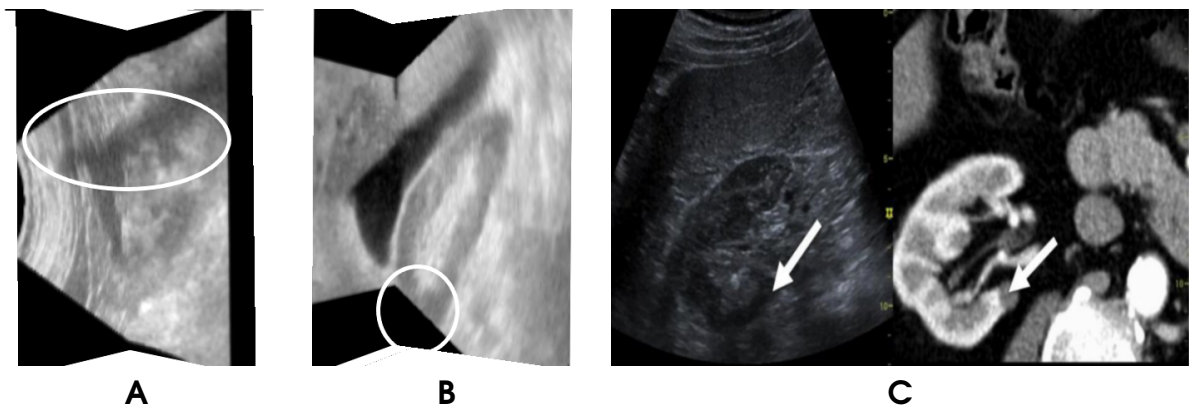


FIGURE 2.7 – Limitations de l'échographie. A- Pôle supérieur de rein occlus par une côte. B- Pôle inférieur du rein "coupé" par la fenêtre acoustique. C- Kyste peu visible à l'échographie, mais clairement visible au scanner.

Pour surmonter ces limitations, l'imagerie de fusion s'est imposée comme une avancée majeure. Il s'agit de combiner les informations pertinentes de deux ou plusieurs images en une résultante [27]. Cette technique permet par exemple de combiner les avantages de l'échographie en temps réel avec ceux d'autres modalités d'imagerie (scanner, IRM), offrant une meilleure visualisation des structures anatomiques profondes, facilitant la planification et la réalisation des gestes interventionnels. Le recalage d'images est une étape clé dans la fusion d'images. Il consiste à aligner spatialement les différentes images à fusionner, malgré les différences de contraste, de résolution et de perspective éventuelles. Lorsque les images mises en jeu sont issues de différentes modalités, le recalage est dit multimodal ou inter-modal.

2.1.2 Le contexte spécifique de la chirurgie rénale percutanée guidée par échographie

Plusieurs des pathologies mentionnées ci-avant sont susceptibles d'attaquer le rein.

Les reins sont des organes du système urinaire, de forme dite "en haricot", avec une symétrie marquée entre ses pôles supérieur et inférieur, qui sera désignée par "symétrie pôle supérieur - pôle inférieur" dans le reste de cette thèse (voir Figure 2.8). Ils sont situés de part et d'autre de la colonne vertébrale, sous le foie pour le rein droit et sous la rate pour le gauche (voir Figure 2.9). Les reins filtrent le sang pour éliminer les déchets et l'excès d'eau sous forme d'urine, régulent la pression artérielle, équilibrent les quantités de sels minéraux et d'acide dans le sang. Un dysfonctionnement rénal compromet ces processus essentiels et augmente le risque d'hypertension artérielle, de déshydratation ou de complications cardiaques.

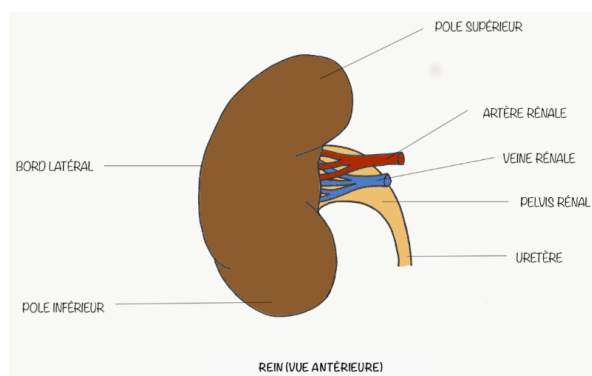


FIGURE 2.8 – Représentation simplifiée du rein [28].

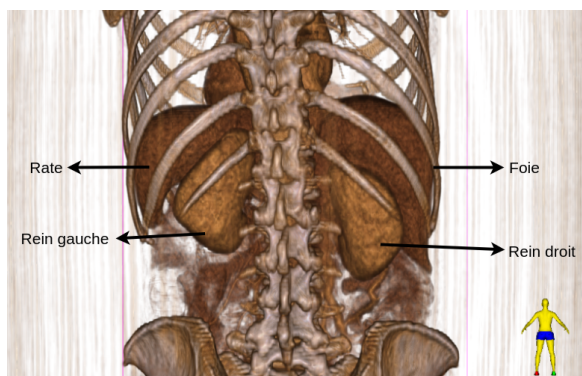


FIGURE 2.9 – Les reins dans la cavité abdominale vue de dos

Parmi les pathologies rénales se trouvent les abcès rénaux, qui se forment en réponse à une infection et apparaissent sous forme de collections liquides. Il y a également les calculs rénaux, petites masses solides de minéraux cristallisés qui se forment dans les reins lorsque certaines substances comme le calcium, l'oxalate ou l'acide urique ne sont pas suffisamment diluées dans l'urine. Les kystes rénaux, quant à eux, sont des poches de liquide, généralement bénignes, mais pouvant nécessiter une intervention dans certains cas. Et principalement, les tumeurs rénales, qui peuvent être bénignes comme l'angiomyolipome ou malignes comme le carcinome à cellules rénales. D'après le "Global Cancer Observatory", le cancer du rein se classe au 9^e rang des cancers les plus fréquents chez les hommes et au 14^e chez les femmes, à l'échelle mondiale, avec environ 431 288 nouveaux cas estimés en 2020 [29]. Les taux d'incidence publiés étant généralement plus élevés en Europe et en Amérique du Nord (Figure 2.10). Cependant, cette forte incidence par rapport aux autres continents est estimée biaisée par une prévalence accrue des petites masses rénales dans les régions où l'imagerie abdominale est plus couramment utilisée [30].

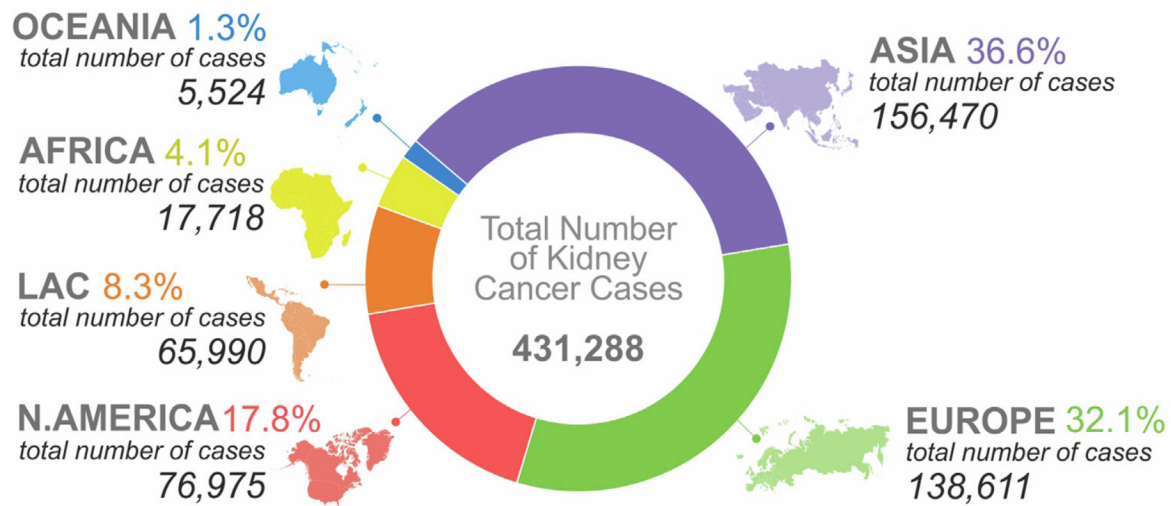


FIGURE 2.10 – Incidence du cancer du rein à travers les continents, représentée en pourcentage du total et le nombre de cas par région pour l’année 2020. Données obtenues de l’Agence internationale de recherche sur le cancer/Organisation mondiale de la Santé [30].

2.1.3 Recalage multimodal et applications cliniques

Modalités d’imagerie - Dimensions spatiales des images.

Les données d’imagerie médicale sont variées, englobant diverses modalités telles que l’échographie, l’IRM ou le scanner..., et diverses dimensions spatiales (images bidimensionnelles ou 2D, images tridimensionnelles ou 3D). Chaque type de données fournit des informations anatomiques ou pathologiques particulières et répond à des indications cliniques spécifiques. L’IRM fournit un excellent contraste des tissus mous, le scanner capture des images haute résolution des structures denses, et l’échographie peut offrir une imagerie en temps réel en 2D et 3D (selon le choix de la sonde). La modalité et la dimensionnalité spatiale des données d’imagerie influencent la conception et les caractéristiques des méthodes de recalage, impactant des aspects tels que l’extraction de caractéristiques, les métriques de similarité et les modèles de transformation. Dans certains cas, il est souhaitable de capturer la variabilité temporelle des données d’imagerie médicale. Cette variabilité temporelle est importante dans des applications telles que l’imagerie cardiaque, les études respiratoires et l’IRM fonctionnelle, où les changements continus au sein d’un système biologique sont surveillés. Les données temporelles rajoutent une couche de complexité au recalage, car il faut un alignement dans les dimensions spatiale et temporelle à la fois. Une méthode efficace de recalage doit maintenir la cohérence temporelle, en compensant les mouvements (par exemple, les cycles cardiaques ou respiratoires) et autres variations temporelles.

La plupart des méthodes de recalage sont spécifiques à une dimensionnalité spatiale. Celles opérant sur des paires d'images se concentrent principalement sur les configurations de recalage 2D-2D (deux images 2D), 2D-3D (une image 2D et une image 3D) et 3D-3D (deux images 3D). Le recalage multimodal 2D-3D d'image échographique, consiste à aligner une image échographique 2D avec un volume 3D provenant d'une autre modalité, telle que l'IRM ou le scanner. Le développement de méthodes de recalage automatique pour cette configuration 2D-3D reste d'un grand intérêt, même avec la disponibilité de sondes échographiques 3D, car les dispositifs échographiques 2D sont accessibles dans des environnements à faibles ressources. Cependant, l'absence d'information échographique en dehors du plan d'imagerie est un défi pour le guidage, et il est difficile, voire impossible, d'effectuer un recalage déformable précis dans ce contexte. En revanche, le recalage multimodal 3D-3D d'image échographique consiste à aligner un volume échographique 3D avec un volume d'une autre modalité. Les problèmes de recalage 3D-3D sont généralement "mieux posés" par rapport au recalage 2D-3D. L'intérêt de la communauté pour cette configuration découle des nouvelles perspectives offertes par l'imagerie échographique 3D, qui enrichit l'information par rapport à l'échographie 2D. Ici, "échographie 3D" peut désigner des images échographiques 3D mécaniques [31, 32, 33, 34], où une sonde 2D avec suivi de position et balayage physique (2.11) ou motorisé génère un volume 3D par interpolation d'images 2D successives. Alternativement, "échographie 3D" peut également se référer aux images obtenues à l'aide de sondes à transducteurs arrangés en matrice [35].

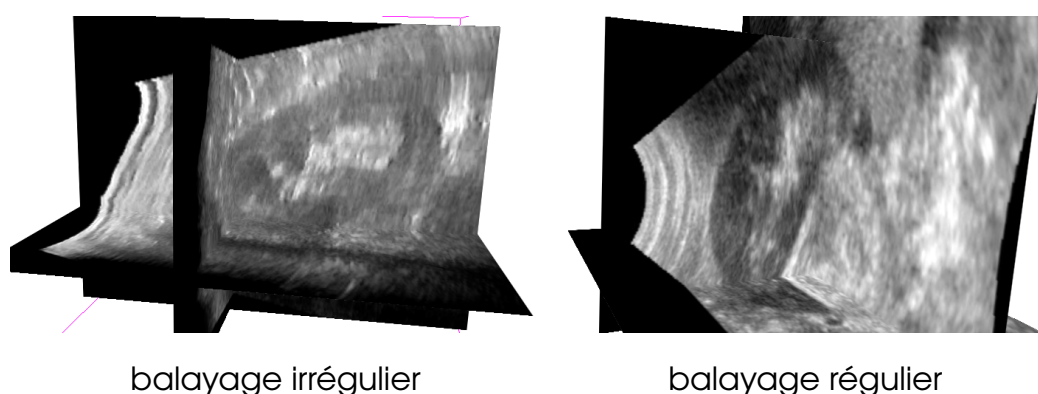


FIGURE 2.11 – Images échographiques 3D obtenues par balayage mécanique. Des discontinuités plus prononcées sont observées dans le cas du balayage irrégulier.

En contexte de recalage multimodal, l'image échographique est combinée avec d'autres modalités d'imagerie comme la tomodensitométrie, l'imagerie par résonance magnétique ou encore la tomographie par émission de positons (PET). Le choix de la modalité de combinaison dépend de l'usage clinique visé et

de la disponibilité de l'outil nécessaire. Le scanner ou CT est l'une des modalités couramment utilisées pour le recalage avec les images échographiques 2D ou 3D. L'avantage de cette modalité est sa capacité à fournir des images à haute résolution avec une excellente représentation anatomique des structures (voir Figure 2.12). Le recalage échographie-CT est utilisé, par exemple, pour des procédures cardiaques [36], spinales [32], rénales [33], hépatiques [37, 38, 39], ou prostatiques [40, 41, 42]. Dans ces cas, la combinaison d'images échographiques en temps réel avec les informations anatomiques précises du CT permet d'améliorer le guidage.

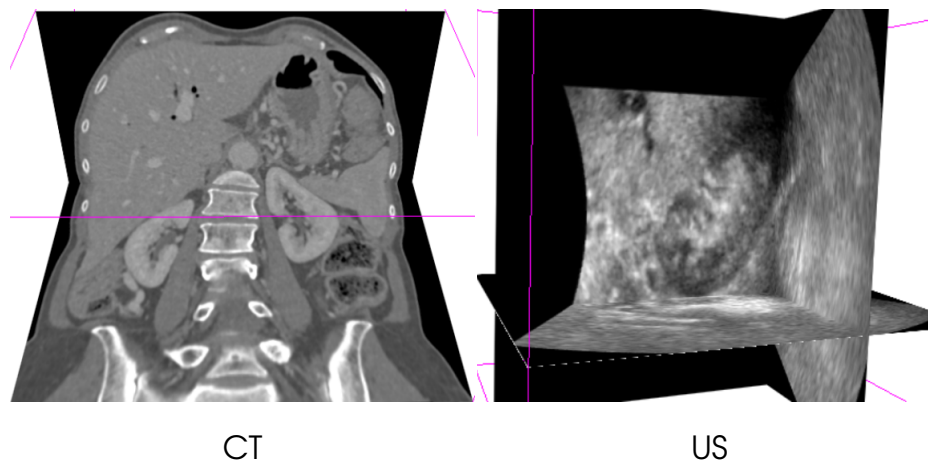


FIGURE 2.12 – Reins dans une image CT et dans une image US 3D

L'IRM est une modalité qui a l'avantage d'offrir un bon contraste des tissus mous. Le recalage échographie-IRM est particulièrement utilisé dans des procédures où il est important de visualiser précisément et préserver l'intégrité des tissus mous. Comme contribution pour la recherche en matière de recalage US-IRM, Baum et al. [43] ont publié un jeu de données d'images US-IRM appariées de la prostate en ouvrant le défi μ -RegPro (Figure 2.13) pour le recalage multimodal et la segmentation. Une difficulté principale dans ce contexte multimodal est la différence significative entre les caractéristiques des images échographiques (dépendantes de l'acoustique) et celles de l'IRM (dépendantes de la réponse des tissus à un champ magnétique).

Le recalage échographie-PET est moins fréquent, mais il présente des avantages dans des cas spécifiques, notamment pour le suivi du métabolisme des tissus. En effet, certaines lésions positives à l'imagerie PET peuvent avoir peu ou pas de corrélation avec l'échographie. Et puisqu'il est difficile de réaliser une biopsie guidée uniquement par PET, en raison de limitations techniques intrinsèques (manque de précision anatomique, absence de visualisation des structures environnantes, résolution spatiale limitée, accessibilité clinique limitée, ...), les informations de la TEP doivent être intégrées dans une procédure de biop-

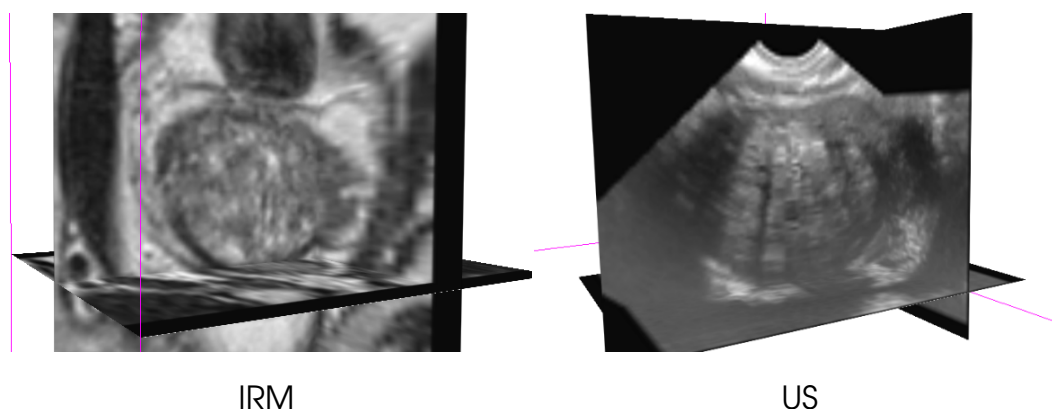


FIGURE 2.13 – Prostate dans une image d'IRM et dans une image US 3D [43].

sie guidée échographie [44]. Ainsi, les images PET fournissent des informations fonctionnelles qui peuvent être cruciales pour l'identification de tumeurs actives, tandis que l'échographie guide en temps réel les interventions (voir Figure 2.14). B. Fei et al. [45] montre par exemple la faisabilité du recalage échographie-PET en temps réel pour guider la biopsie chez les patients présentant un échec biochimique après traitement non chirurgical du cancer de la prostate.

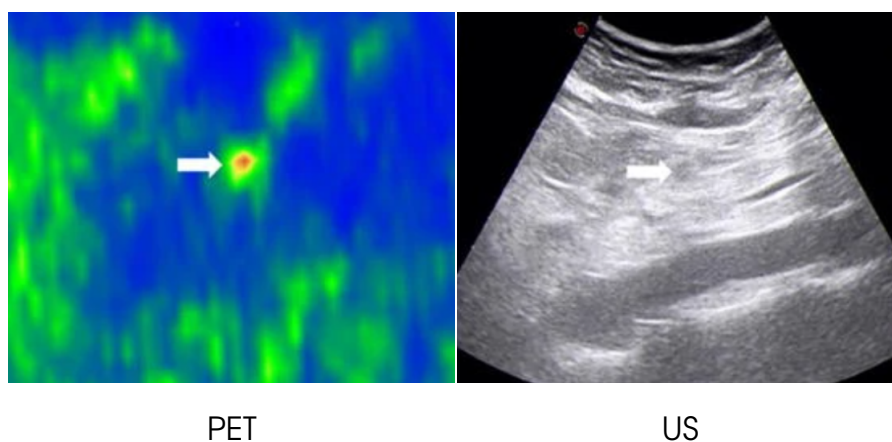


FIGURE 2.14 – Visualisation, pour un même patient, d'une lésion du côlon droit dans une image PET et dans une image US. La lésion est légèrement visible à l'échographie, mais clairement détectable à l'image PET [44].

Les techniques d'imagerie optique, comme la tomographie par cohérence optique (OCT – pour "Optical Coherence Tomography" en anglais), sont parfois intégrées avec l'échographie pour des applications spécifiques comme la médecine vasculaire [46, 47, 48]. Le recalage entre ces deux modalités repose sur la complémentarité entre la haute résolution des images optiques et les capacités de pénétration plus profonde de l'échographie.

Recalage d'images par paires.

Cette thèse traite du recalage des images par paires. Chaque paire est constituée d'une image fixe (ou de référence) et d'une image mobile (ou flottante). L'objectif du recalage dans ce contexte est d'aligner spatialement l'image mobile avec l'image fixe. Dans certains cas d'application, comme la création d'atlas, il peut être question d'aligner plusieurs images mobiles, chacune avec une image fixe. Dans certains cas encore, le but peut être de recaler plusieurs images ensemble sans définir explicitement une image de référence : recalage multi-images. Cependant, cette thèse porte uniquement sur le recalage d'images par paires, entre deux modalités, dont l'une est l'échographie tridimensionnelle.

Les applications du recalage multimodal d'images échographiques.

Le recalage multimodal permet d'enrichir les images peropératoires souvent pas bien contrastées (échographie, CT sans produit contraste), avec des images pré-opératoires mieux contrastées (CT avec produit de contraste). Cela permet une navigation précise des instruments à l'intérieur du corps [49]. La chirurgie robotique utilise des systèmes automatisés assistés par un chirurgien pour effectuer les interventions avec une précision accrue. Dans ce contexte, le recalage d'images permet d'améliorer la précision des robots chirurgicaux, en synchronisant des images préopératoires (comme l'IRM ou le CT) avec des images peropératoires en temps réel (comme l'échographie). Cela offre une visualisation détaillée et dynamique des organes et des tissus pendant l'intervention. Grâce au recalage, les robots peuvent naviguer avec une extrême précision, guidant les instruments chirurgicaux vers les zones ciblées tout en évitant les structures critiques. Ce qui réduit les complications et améliore les résultats pour le patient [50, 51, 52].

Dans la planification et la simulation d'interventions chirurgicales, le recalage d'images médicales permet de créer des modèles précis et personnalisés des organes du patient en fusionnant des images souvent issues de différentes modalités. Ces modèles réalistes sont utilisés pour simuler des interventions chirurgicales avant l'opération réelle, permettant aux chirurgiens de planifier, de répéter des gestes complexes dans un environnement virtuel, de définir des trajectoires, en tenant compte des spécificités anatomiques du patient [53]. Cela améliore la préparation et réduit les risques pendant l'intervention.

Dans le domaine des prothèses personnalisées, le recalage d'images permet de créer des modèles anatomiques précis pour concevoir des implants adaptés aux besoins spécifiques du patient. Par exemple, un scan CT de la surface de la peau peut être utilisé pour inférer la forme du crâne grâce à un processus de

recalage, et ainsi évaluer le résultat d'un traitement orthodontique [54].

Le recalage d'images est également utilisé pour superposer des données fonctionnelles (par exemple d'IRM fonctionnelle) sur des images structurales (CT ou IRM) [55]. L'imagerie fonctionnelle montre l'activité physiologique ou métabolique à l'intérieur de l'organe, tandis que l'imagerie structurale fournit des images anatomiques statiques, montrant la forme et l'emplacement des structures. Cette superposition permet d'identifier à la fois les régions actives et le type d'activité.

S'agissant du suivi, qu'il s'agisse de celui de l'évolution d'une maladie (croissance d'une tumeur, évolution d'une lésion), ou de l'évolution d'une procédure thérapeutique (radiothérapie, médicaments), le recalage permet d'aligner des images obtenues à différents moments afin de surveiller les changements sur la structure d'intérêt [56]. Par exemple, dans le cas du traitement par radiothérapie d'une tumeur cancéreuse, le recalage permet de surveiller les changements de taille de celle-ci. Le clinicien peut alors ajuster les doses de radiations en fonction de ces changements, ce qui limite l'exposition des tissus sains aux radiations.

2.2 Considérations et défis dans le recalage d'images médicales

Cette section explore les considérations et les défis qui se posent dans la mise en œuvre de techniques de recalage d'images médicales. Les questions abordées comprennent la modélisation de la transformation spatiale nécessaire pour aligner les images, la prise en compte des variations anatomiques et pathologiques propres à chaque individu, la modélisation de l'information contenue dans les images, et la validation du recalage. Il sera également question du choix entre recalage global et affinement précis, ainsi qu'entre approche automatique et approche interactive. Des références représentatives sont indiquées dans chaque cas. Un approfondissement sur les méthodes de pointe spécifiquement développées ou simplement applicables au recalage inter-modal d'images échographiques de rein est fait dans la section d'après.

2.2.1 Comment modéliser la transformation spatiale de recalage ?

En pratique, un alignement parfait des images n'est pas possible. Une méthode de recalage doit offrir une précision suffisante pour l'application clinique

visée. Les méthodes de recalage nécessitent alors une formulation d'hypothèse sur la transformation spatiale reliant les images : sélection de modèle de transformation spatiale.

Il y a généralement un compromis entre l'utilisation d'un modèle de transformation très flexible, qui pourrait théoriquement offrir une grande précision, et la stabilité du modèle de recalage. En effet, ce type de modèle de transformation peut rendre l'opération de recalage instable ou mal conditionnée en raison de l'insuffisance des informations visuelles contenues dans les images médicales.

La sélection du modèle de transformation spatiale implique quatre facteurs principaux : 1) la biomécanique de la structure anatomique d'intérêt, 2) le mouvement ou la déformation attendue de la structure dans les deux images, 3) les informations disponibles dans les images et 4) la précision de recalage ciblée. Toutefois, en contexte clinique, Quel que soit le choix de transformation fait, la validation du recalage suivant l'application visée, reste essentielle.

Divers modèles de transformation spatiale ont été envisagés dans la littérature sur le recalage d'images médicales, comme illustré dans la Figure 2.16. Une transformation rigide suppose que la structure anatomique subit uniquement des mouvements de translation et de rotation. Le recalage rigide d'images échographiques avec des images CT ou IRM est largement utilisé dans les procédures hépatiques, rénales ou prostatiques percutanées. Ces organes ne sont pas rigides à proprement dire [57, 58]. Mais l'hypothèse faite est que, les mouvements observés entre les modalités, souvent étendus aux transformations de similarité, sont dus à des déplacements globaux de la structure plutôt qu'à des changements internes dans sa forme. Il peut aussi être considéré que les transformations rigides (ou de similarité) sont suffisantes pour l'application visée. En contexte de procédures rénales, Leroy et al. [59], Delaunay et al. [60], ou Jackson et al. [61] ont fait des hypothèses de transformation rigide pour le recalage d'images échographiques et CT/IRM acquises sur fantôme. Quant à Xing et al. [62], Markova et al. [38], Cifor et al [63], Bing et al. [64], ou Wein et al. [65], ils ont fait une hypothèse de transformation rigide pour des procédures hépatiques, notamment focalisées sur les vaisseaux (voir Figure 2.15). Cette hypothèse de transformation pour le recalage de la prostate est également faite dans [66, 67].

Les transformations affines étendent les transformations rigides en incluant les changements d'échelle et les cisaillements, ce qui permet de modéliser des changements de forme plus complexes (12 degrés de liberté en dimension 3 contre 6 pour les rigides) tout en maintenant des hypothèses de linéarité. Cette approche est utile pour des organes qui peuvent changer de taille ou de forme, mais qui conservent une structure globale stable. C'est un affinement du

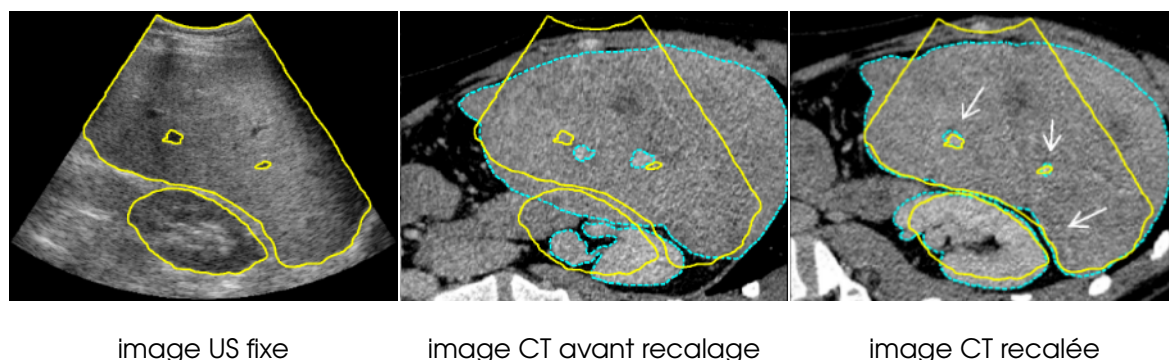


FIGURE 2.15 – Récadrage rigide des vaisseaux hépatiques dans les images échographiques et CT [63].

recadrage rigide dans le cas de déformations hypothétiquement linéaires. Les hypothèses qui peuvent soutenir le choix d'un recadrage affine sont :

- La structure anatomique peut subir des transformations globales, telles que des changements de taille ou des déformations linéaires, mais les relations spatiales locales restent relativement inchangées
- La déformation est homogène à travers la structure anatomique d'intérêt, ce qui signifie qu'il n'y a pas de variations locales importantes dans la manière dont la structure se déforme

Les méthodes de recadrage fondées sur des modèles paramétriques peuvent être utilisées pour ajuster les paramètres de la transformation affine. Ces méthodes sont souvent combinées avec des techniques de minimisation d'erreur utilisant l'information d'intensité ou de surface pour optimiser la correspondance entre les images. Pour le recadrage multimodal d'images échographiques de la région de l'abdomen, cette hypothèse a par exemple été faite dans [65, 68, 69, 70] pour le recadrage d'images du foie. Ces travaux ont montré en outre que l'hypothèse de transformation affine améliorerait l'alignement de l'organe dans les différentes modalités, par rapport à une hypothèse de transformation rigide. Hu et al. [71], et Song et al. [72] ont également utilisé l'hypothèse de transformation affine, cette fois, pour le recadrage de la prostate.

Les transformations non-rigides (ou déformables) permettent de modéliser des déformations complexes, y compris les contractions et les étirements locaux, qui sont courants dans les tissus mous. Ces transformations permettent des changements locaux de forme, modélisant ainsi les propriétés élastiques des tissus qui peuvent subir des étirements, des compressions, ou des déformations non linéaires. Les transformations non-rigides sont souvent utilisées dans des contextes cliniques où la précision locale est cruciale. Ces transformations sont particulièrement importantes pour le recadrage d'images échographiques, car les tissus mous se déforment en réponse à des forces internes comme la

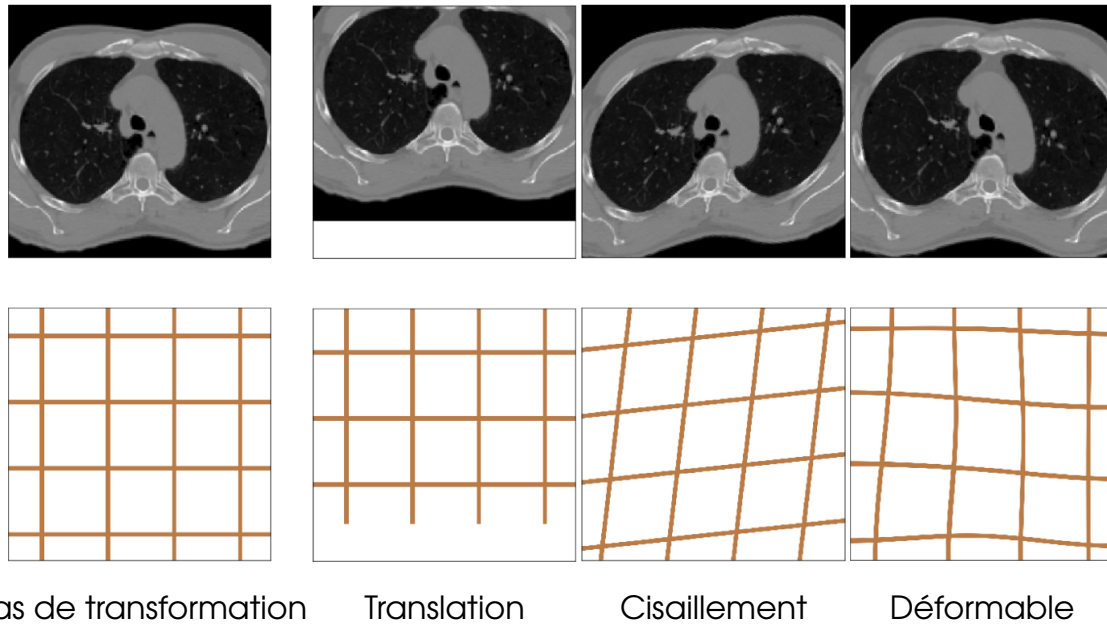


FIGURE 2.16 – Exemples de transformations rigides (translation), affine (cisaillement le long des axes), et déformable, appliquées à une coupe axiale de poumons dans une image CT [73].

respiration, mais aussi en réponse à des forces externes comme la pression de la sonde échographique. Le choix d'une hypothèse de transformation non-rigide peut être soutenu par le fait que :

- les propriétés mécaniques des tissus, telles que l'élasticité ou la viscosité, influencent les déformations observées et doivent être prises en compte dans le modèle de transformation,
- les déformations sont hétérogènes à travers la structure anatomique ; par exemple, certaines parties de l'organe peuvent être comprimées tandis que d'autres ne le sont pas.

Les transformations de recalage non-rigide incluent les champs de déplacement déformables, où chaque point de l'image peut être déplacé indépendamment pour obtenir un alignement optimal (voir Figure 2.16). Les méthodes utilisées pour modéliser ces transformations sont souvent guidées par des modèles physiques comme les modèles élastiques afin de garantir que les déformations sont réalistes [74, 5, 6]. Les approches d'apprentissage profond à base de réseaux de neurones convolutifs comme VoxelMorph [75], ou à base de Transformeur de Vision (ViT - "Vision Transformer" en anglais) comme TransMorph [76], ont également gagné en popularité pour le recalage non-rigide. Bien que les transformations non-rigides offrent une grande flexibilité, elles sont également plus complexes à optimiser et nécessitent plus de calculs que les transformations rigides ou affines. En contexte de recalage multimodal d'images échographiques, cette

hypothèse a été souvent utilisée sur la prostate [77]. Elle est également retrouvée dans des travaux portant sur le foie ou ses vascularisations [7, 78, 79], et un peu plus rarement dans des travaux portant sur le rein [80, 81].

2.2.2 Comment gérer les variabilités anatomique et pathologique

La variabilité intra/inter-patient fait référence aux différences observées entre images issues du même patient et celles observées entre images issues de patients différents. Ce critère est à considérer à la fois comme besoin pour les applications et comme enjeu pour la conception de méthodes de recalage multimodal d'images médicales (notamment échographiques). Il permet d'orienter la sensibilité ou la robustesse des méthodes.

Les différences intra-individus sont généralement créées par le changement de position du patient, l'activité physiologique (respiration, mouvements cardiaques, distension des organes), le processus d'acquisition échographique (profondeur d'imagerie, technique d'acquisition du praticien), ou le processus de traitement de l'image. La prise en compte de ces différences sert pour des besoins de suivi thérapeutiques, de surveillance, de guidage interventionnel, ou de gestion de déformations tissulaires [82].

Les différences inter-individus renvoient aux différences anatomiques et physiologiques entre différents patients. Elles incluent les variations de taille, de forme et de position des organes, ainsi que l'âge, le sexe et l'état pathologique du patient. Ces différences sont prises en compte lors de la création de modèles personnalisés à partir d'une référence (atlas) pour une planification d'intervention ou pour un traitement [83]. Elles sont également à considérer en cas de besoin d'apprentissage de caractéristiques communes sur cohortes ou de comparaison anatomique [84].

La présence de lésions importantes et les traitements chirurgicaux antérieurs peuvent également introduire une variabilité significative. En général, la plupart des approches du recalage d'images ne sont pas applicables lorsque la morphologie de la structure d'intérêt a changé de manière significative en raison d'une intervention chirurgicale. En pratique, dans la chirurgie percutanée, cette hypothèse est acceptable, car l'intervention chirurgicale se limite à l'insertion de petites aiguilles, ce qui ne modifie pas radicalement la forme de l'organe, contrairement à une résection chirurgicale.

Les techniques de "Machine Learning" sont de plus en plus utilisées pour traiter à la fois les variabilités intra- et inter-individus. Ces modèles peuvent être entraînés de manière spécifique pour apprendre des représentations invariantes

aux différences individuelles, tout en s'adaptant aux spécificités de chaque patient. C'est le cas dans [85] (voir Figure 2.17), où la variabilité intra-individu a été simulée par analyse des éléments finis (FEA - pour "Finite-element analysis" en anglais) [86]. Ces techniques nécessitent de disposer de jeu de données conséquent, pour leur entraînement.

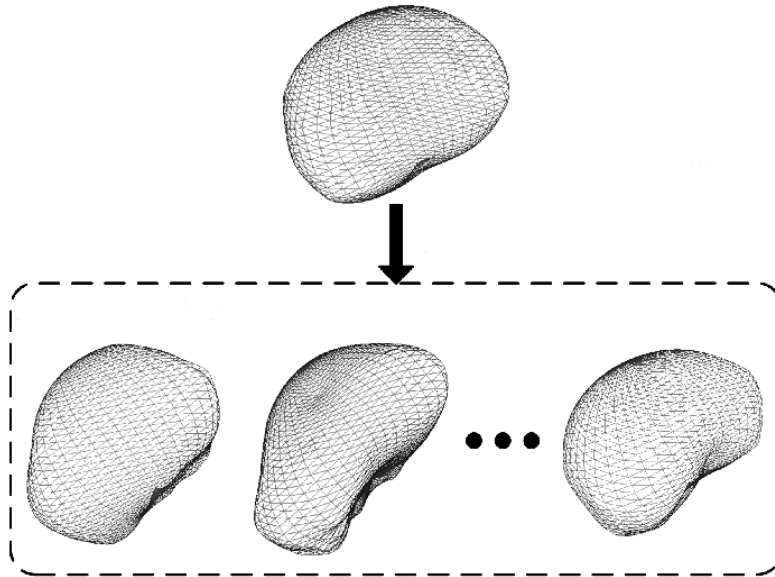


FIGURE 2.17 – Préparation des données d’entraînement : une série de formes de la prostate simulée sur la base d’une référence issue d’image d’IRM. De telles séries modélisent la variabilité intra-individu et sont utilisées comme données d’entraînement dans [85].

2.2.3 Comment obtenir des jeux de données pour le développement de modèles d’apprentissage ?

La comparaison objective des méthodes requiert l’usage de bases de données communes. De plus, le développement de modèles d’apprentissage nécessite des données d’entraînement. En effet, ces méthodes d’apprentissage modélisent le problème du recalage à partir d’exemples fournis. Et plus il y a d’exemples, plus il est possible de se rapprocher du « vrai » modèle. Pour le recalage en vue d’application en chirurgie percutanée, il est souhaitable que les données issues des différentes modalités soient appariées. Enfin, les méthodes actuellement performantes ont recours à des annotations vérité terrain. Ainsi, un jeu de données intéressant pour cette tâche doit être de taille importante (en termes de nombres d’individus), avoir des images appariées et posséder des annotations précises.

La création d'un jeu de données publiques constitué d'images médicales multimodales présente des défis éthiques et pratiques. Sur le plan pratique, il y a la conception du protocole d'acquisition, sa validation par une institution compétente et sa mise en œuvre, incluant les différents coûts. Sur le plan éthique, il y a le respect de la protection des données patient durant le cycle de vie des données et le contrôle de l'usage des données. Après l'acquisition, il y a la phase d'annotation précise. Une bonne pratique serait d'y associer aussi un protocole qui définit les annotations et les processus d'obtention, ce qui requiert une certaine expertise. Un autre défi de cette phase est l'exécution du processus d'annotation proprement dit, qui mobilise d'importantes ressources humaines et machines.

Créer un jeu de données remplissant les exigences des méthodes d'apprentissage profond pour le recalage multimodal d'images échographiques est donc particulièrement coûteux et long. Ce qui devrait expliquer le fait qu'il est difficile de trouver des jeux de données publiques pouvant servir à développer (entraînement et évaluation) des modèles d'apprentissage pour le recalage multimodal d'images échographiques.

Une revue de la littérature portant sur les jeux de données publiques utilisables dans le domaine du recalage multimodal d'images échographiques sera faite au Chapitre 3.

2.2.4 Comment modéliser l'information dans les images médicales ?

Au cours de la dernière décennie, diverses approches pour le recalage d'images médicales ont été proposées. Elles peuvent être catégorisées en fonction des informations dérivées des images, qui sont utilisées. Trois catégories principales ont émergé : 1) le recalage iconique (utilisant les valeurs d'intensité des voxels ou des pixels), 2) le recalage utilisant les surfaces (fondé sur les informations provenant des surfaces ou des contours de structures extraits des images), et 3) le recalage utilisant des points de repère (fondé sur l'utilisation de points identifiables et extraits des images). Chacune présente des forces et des défis uniques, en particulier pour le recalage inter-modal, qui est souvent plus compliqué à résoudre, en raison des différences entre les modalités d'imagerie. Le Tableau 2.1 résume les forces et les faiblesses de chaque approche, en soulignant leur pertinence pour différents scénarios en pratique médicale.

Recalage iconique.

Les méthodes iconiques s'appuient sur les valeurs d'intensité des pixels ou des

voxels des images pour réaliser l'alignement. Ces algorithmes transforment les images en maximisant ou en minimisant une métrique de similarité appropriée entre les valeurs d'intensité. Cette métrique peut être conçue manuellement ou formée par apprentissage profond (métrique apprise).

Les métriques manuelles sont fondées sur des hypothèses prédéfinies concernant la relation entre les valeurs d'intensité des deux images [87]. Les métriques courantes incluent la Somme Normalisée des Différences au Carré ("Normalized Sum of Squared Differences" - NSSD), la Corrélacion Croisée Normalisée ("Normalized Cross Correlation" - NCC) et l'Information Mutuelle ("Mutual Information" - MI). Bien que NSSD et NCC soient couramment utilisées pour le recalage intra-modal, elles sont inadaptées pour le recalage multimodal où les valeurs d'intensité des images ne peuvent pas être liées de manière linéaire. En revanche, l'Information Mutuelle (MI) est adaptée pour le recalage multimodal, car elle mesure la dépendance statistique ou l'information partagée entre deux images sans présumer de relation linéaire. Cela rend la MI adaptée aux images multimodales, où la même structure anatomique peut apparaître avec des valeurs d'intensité différentes dans les diverses modalités. Des variantes de la MI, telles que la MI normalisée (NMI) [88], la MI conditionnelle (cMI) [89], la MI régionale (RMI) [90] ont été proposées. Afin d'améliorer la précision et la robustesse des métriques de similarité, il y a aussi eu des approches fondées sur les divergences, développées avec un certain succès. Il s'agit notamment de la divergence de Kullback-Leibler [91] et de la divergence de Jensen-Shannon [92].

Les métriques de similarité formées par apprentissage profond (à partir de données d'entraînement) utilisent souvent des architectures comme les auto-encodeurs empilés [93, 94] ou les réseaux de neurones convolutifs (CNN) [95, 96], afin de capturer des relations complexes entre les images. Les chercheurs explorent de plus en plus la combinaison de métriques de similarité formées par apprentissage profond avec des métriques manuelles, pour créer des mesures hybrides qui améliorent les performances du recalage.

La flexibilité des méthodes iconiques de recalage permet d'appliquer des transformations rigides ou des transformations déformables aux images. Toutefois, les méthodes actuellement disponibles, se sont montrées plus efficaces lorsque les images partagent des distributions d'intensité similaires. En effet, bien que les méthodes iconiques puissent être appliquées au recalage inter-modal, elles rencontrent souvent des difficultés dues aux différences de contraste, de luminosité et de bruit entre les modalités d'imagerie. Une différence importante de domaine entre les images peut facilement entraîner des erreurs de correspondance et donc de recalage.

Recalage utilisant les surfaces.

Le recalage utilisant les surfaces aligne les surfaces anatomiques dérivées des images en se concentrant sur les propriétés géométriques ou de forme plutôt que sur les valeurs d'intensité ; ce qui le rend particulièrement adapté au contexte inter-modal. Ces méthodes commencent par créer un modèle de surface à partir des images, souvent en utilisant des réseaux de neurones convolutifs tels que nnU-Net [97], ou des réseaux intégrant des "Vision Transformer" (ViT), comme le U-Net Transformer [98]. Une fois les surfaces segmentées, elles peuvent être alignées spatialement à l'aide d'une méthode de recalage surface à surface. De nombreuses méthodes ont été proposées dans les communautés de la vision par ordinateur et des graphiques informatiques [99, 100]. Parmi les méthodes les plus couramment utilisées avec les données médicales, il y a la méthode itérative des plus proches voisins ("Iterative Closest Point" - ICP) et ses nombreuses variantes [101, 102], ainsi que la méthode de décalage cohérent de points ("Coherent Point Drift" - CPD) [103]. ICP et CPD sont des méthodes itératives qui optimisent les paramètres du recalage en minimisant une métrique de distance surface à surface. Il existe des variantes robustes de ces méthodes, qui peuvent gérer, dans une certaine mesure, les erreurs de segmentation. C'est le cas de la version tronquée de l'ICP [104].

Les méthodes de recalage utilisant les surfaces sont plus robustes aux variations d'intensité des pixels, grâce aux modèles de surface qui plongent les deux images dans un même domaine. Ces méthodes sont donc idéales pour les images à faible contraste ou pour des modalités présentant des caractéristiques très variées. De plus, le recalage utilisant les surfaces peut intégrer des transformations déformables pour s'adapter à des variations anatomiques importantes.

Une limitation de ces méthodes est qu'elles n'utilisent que les indices visuels correspondant aux frontières segmentées des surfaces, négligeant d'autres informations précieuses dans les données d'intensité.

Recalage utilisant des points de repère.

Le recalage utilisant des points de repère aligne les images en identifiant et en appariant des points spécifiques et facilement reconnaissables dans chaque image. Ces points de repère sont généralement des caractéristiques anatomiques ou des points distinctifs qui sont cohérents entre différentes images ; par exemple les coins de structures ou les jonctions de vaisseaux. Une fois les points de repère correspondants identifiés, l'algorithme du recalage calcule une transformation qui minimise la distance entre les points de repère appariés, permettant un alignement des images. Cette approche est efficace lorsqu'il existe des points de repère clairs et identifiables, ce qui la rend bien adaptée

Approche	Forces	Faiblesses
Iconique	- Utilise toutes les informations de pixel/voxel disponibles, augmentant la précision. - Pas d'extraction de surface ou de caractéristiques nécessaire.	- Sensible au bruit et aux artefacts qui affectent les valeurs d'intensité. - Risque de minimum local avec les approches d'optimisation.
Utilisant les surfaces	- Robustesse aux variations d'intensité dans le recalage inter-modal. - Basé sur des propriétés géométriques, efficace pour le recalage inter-modal.	- Extraction de surface difficile pour images à faible contraste. - Précision dépendante de la qualité de segmentation. - Perte d'information. - Risque de minimum local avec l'optimisation.
Utilisant des points de repère	- Simple et intuitif ; facile à comprendre. - Alignements clairs en utilisant des points anatomiques reconnaissables.	- Détection difficile dans des images bruitées. - Précision limitée par le nombre de points identifiables. - Erreurs de points peuvent nuire à la précision. - Perte d'information.

TABLE 2.1 – Forces et faiblesses des différentes approches de recalage d'images médicales.

au recalage d'images présentant des différences significatives d'intensité ou de modalité.

Dans des conditions idéales, lorsque les mêmes points de repère sont détectés dans les deux images, la transformation spatiale optimale peut être calculée efficacement. Pour les transformations rigides ou de similarité, l'algorithme d'orientation absolue de Horn est couramment utilisé [105]. Alternativement, si les positions transformées des points de repère peuvent être exprimées comme combinaisons linéaires des paramètres de recalage (comme dans les modèles

de translation, de transformation affine ou de déformation par B-splines), alors les paramètres optimaux peuvent être calculés en utilisant la méthode des moindres carrés linéaires. Cependant, le recalage utilisant des points de repère devient difficile dans des conditions non idéales, en particulier lorsque les points de repère sont difficiles à détecter ou inconsistants entre les images, comme cela se produit souvent lors du recalage inter-modal d'images échographiques. Des méthodes robustes d'alignement de points, telles que RANSAC [106], peuvent être appliquées dans de tels cas. Néanmoins, il peut être difficile d'obtenir un alignement de points robuste, lorsqu'il y a un faible nombre de points de repère disponibles, ou une proportion élevée de points mal appariés, ou encore lorsque la structure présente une forte symétrie.

Il est toutefois à noter que le recalage utilisant des points de repère n'est souvent pas adapté pour résoudre des transformations déformables complexes, puisqu'il implique une représentation éparse des structures, négligeant une quantité considérable d'informations d'intensité potentiellement utiles.

Aucune stratégie unique optimale - "No Free Lunch".

En résumé, chacune des trois principales approches pour le recalage d'images médicales (recalage iconique, recalage utilisant les surfaces et recalage utilisant des points de repère) présente des forces et des faiblesses, en particulier dans le contexte du recalage inter-modal. Le choix de l'approche du recalage appropriée dépend de facteurs tels que le contexte clinique, les types d'images à aligner, la disponibilité de caractéristiques anatomiques identifiables, ainsi que de considérations liées au coût computationnel et aux besoins en temps réel. En outre, certains chercheurs ont proposé des approches hybrides pour combiner les avantages des méthodes iconiques, utilisant les surfaces et utilisant les points de repère, tout en atténuant leurs défauts individuels. Celles de ces méthodes qui sont spécifiquement appliquées au recalage inter-modal des images échographiques, sont explorées dans la Section 2.3. Sont incluses, certaines approches utilisant des points de repère curvilignes pour intégrer des modèles de réseaux vasculaires dans le processus de recalage.

Le type de méthode de recalage dépend également de l'usage visé. Il s'agit de savoir si un alignement approximatif (recalage global - sans initialisation) est suffisant, ou s'il faut un recalage précis (affinement d'une estimation de recalage déjà existante). Ces concepts de recalage global et d'affinement du recalage sont examinés par la suite.

2.2.5 Recalage global ou affinement du recalage ?

2.2.5.1 Recalage global et principaux défis.

L'objectif du recalage global est de fournir une transformation qui aligne grossièrement (généralement à l'aide d'une transformation rigide) la structure d'intérêt, sans avoir besoin d'une estimation initiale comme point de départ. La précision n'est pas essentielle à ce stade. Ces approches peuvent être utilisées comme objectif principal de recalage, par exemple pour une vérification rapide ne nécessitant pas une grande précision. Les approches globales peuvent également servir d'étape intermédiaire pour initialiser une méthode de recalage plus précise (affinement du recalage), qui s'appuie sur une estimation initiale de recalage pour progresser.

Une méthode de recalage global fiable doit présenter un faible risque de produire un mauvais résultat d'alignement. Ce qui peut se produire pour l'une des deux raisons principales suivantes :

Premièrement, un mauvais recalage global peut résulter de l'utilisation d'une optimisation itérative sur une fonction de coût non convexe. Le processus d'optimisation peut alors se retrouver piégé dans un minimum local. Ceci est particulièrement problématique pour le recalage itératif iconique, où la fonction de coût est intrinsèquement non convexe en raison de la nature complexe des données d'imagerie et de la métrique de similarité elle-même (qu'il s'agisse de métriques manuelles ou de métriques de similarité apprises). Cette non-convexité introduit de multiples minima locaux qui peuvent piéger l'algorithme, l'empêchant d'atteindre l'optimum global. Ces "pseudo bassins de convergence", issus de la structure non linéaire des données d'imagerie, peuvent entraver le processus d'optimisation, augmentant le risque de se stabiliser dans des minima locaux qui ne capturent pas avec précision les véritables relations spatiales entre les paires d'images. Pour cette raison, le recalage iconique est rarement utilisé pour le recalage global.

Deuxièmement, un mauvais recalage global peut survenir en raison d'un problème mal conditionné, dans lequel la méthode de recalage dispose d'informations insuffisantes pour obtenir un bon alignement. En contexte de recalage inter-modal d'images échographiques, cela peut se produire, car la fenêtre d'imagerie échographique est souvent petite et centrée autour de la structure d'intérêt. Si la structure présente une forte symétrie naturelle, comme le rein (illustré dans la Figure 2.8), alors, sans une utilisation efficace des informations contextuelles des images, la méthode peut ne pas parvenir à réaliser un recalage correct. Comme le montre le Chapitre 4, ce problème se pose avec les méthodes de recalage utilisant les surfaces. Cela est dû à la perte d'informations sur les structures environnantes comme le foie, et sur les détails d'intensité à l'intérieur du rein. Ces informations pourraient pourtant aider à déterminer

l'orientation du rein.

2.2.5.2 Affinement du recalage et principaux défis.

L'objectif principal de l'affinement du recalage est de minimiser les divergences résiduelles du recalage global, afin d'atteindre un niveau de précision qui réponde aux exigences cliniques. Ce processus est généralement réalisé par optimisation itérative, en utilisant le résultat du recalage global comme estimation initiale. Contrairement au recalage global, l'affinement du recalage utilise, par définition, une estimation initiale comme point de départ, permettant ainsi l'application de méthodes d'optimisation itératives. Par conséquent, le recalage iconique est souvent privilégié pour l'affinement du recalage, car il permet d'exploiter toutes les informations disponibles dans les images pour atteindre la précision la plus élevée possible ; ce qui n'est pas le cas du recalage utilisant les surfaces ou les points de repère.

Les principaux défis de l'affinement du recalage incluent la garantie d'un alignement précis dans les régions de l'image qui sont soit très homogènes en intensité, soit obstruées par des artefacts (par exemple, les ombres acoustiques dans les images échographiques). Pour relever ce défi, des connaissances préalables sur la biomécanique spécifique des tissus peuvent être nécessaires pour contraindre le recalage déformable dans ces régions. Plusieurs études ont exploré cette direction, comme [107] et [108]. Cependant, l'acquisition et l'intégration efficace des informations biomécaniques restent un défi de recherche complexe et ouvert.

En conséquence, de nombreuses méthodes de recalage déformable s'appuient sur des modèles géométriques simples et génériques de mouvement des tissus, tels que les splines de plaque mince ("thin-plate splines") et les B-splines. Des exemples bien connus dans cette catégorie incluent VoxelMorph [75] et TransMorph [76].

2.2.6 Recalage automatique ou recalage interactif ?

Dans certains scénarios cliniques, il est admissible que l'initialisation du recalage implique une intervention humaine. Par exemple, pour l'identification d'un nombre limité de points de repère anatomiques homologues dans les deux images. Dans le cas d'images tridimensionnelles, un minimum de trois points non colinéaires est généralement requis afin de fournir des informations suffisantes pour résoudre le recalage global (six degrés de liberté, ou sept en cas de mise à l'échelle). Cependant, dans de nombreux contextes cliniques, le recalage interactif n'est pas souhaitable, car il ralentit le flux de l'activité clinique. Cela

est particulièrement vrai en chirurgie percutanée où le temps supplémentaire en salle d'opération (pour effectuer un recalage interactif) est à la fois coûteux et peut prolonger la durée de l'anesthésie générale. De plus, le chirurgien doit être le plus libre possible de toute charge cognitive supplémentaire, afin de se concentrer sur l'exécution de la procédure guidée par imagerie, plutôt que de détourner son attention pour réaliser un recalage interactif. Le recalage interactif comporte également des risques d'erreur humaine et de subjectivité. Il n'est pas réalisable pour les applications où le temps est critique, en particulier pour les besoins en temps réel. Pour ces raisons, le recalage automatique (qui ne nécessite pas d'interaction humaine) est devenu un objectif clé de la recherche. La combinaison du recalage global et de l'affinement du recalage offre une approche prometteuse pour atteindre cet objectif.

Dans la littérature, les tentatives pour résoudre de manière automatique le recalage complet (initialisation et affinement) sont plutôt rares. Par contre, il y a des tentatives de résolution du recalage global du foie [38] ou de la prostate [67], de manière automatique, en vue d'initialiser l'affinement. Mais ces travaux ne poussent pas l'analyse à l'évaluation des performances du système complet de recalage. Même si intuitivement, l'enchaînement entre les deux étapes se conçoit simplement, il ne faut pas négliger d'éventuelles difficultés pratiques.

Un défi important de l'automatisation des systèmes complets de recalage multimodal précis est la robustesse des modèles aux variabilités des données cliniques. Les images médicales, et tout particulièrement échographiques, varient en fonction des outils et protocoles d'acquisition, mais également des différences individuelles entre les patients et leurs conditions physiques. En conséquence, les modèles doivent être capables de généraliser sur de nouvelles données sans perdre en précision.

2.2.7 Comment valider la précision du recalage ?

L'évaluation de la précision d'une méthode du recalage d'images médicales à des fins de comparaison n'est pas un exercice simple. En effet, il n'est en général pas possible de disposer de transformation vérité terrain contre laquelle comparer une estimation de recalage. Ceci parce que l'information sur la transformation géométrique réelle qui aligne deux images acquises dans des contextes différents, est difficilement accessible. Parmi les approches les plus simples, des techniques artificielles ont donc été proposées pour évaluer les méthodes de recalage. Elles s'appuient principalement sur l'utilisation de points de repère anatomiques ou de segmentations produits manuellement [109]. La superposition de ces caractéristiques appariées dans l'image fixe et l'image

mobile après recalage, permet d'évaluer le recalage et de comparer différentes méthodes. Parmi ces métriques figurent :

- L'erreur sur la (les) cible(s) de recalage ("Target Registration Error" : TRE). Elle utilise des points de repère anatomiques. Elle évalue la distance euclidienne entre la position d'un point cible dans l'image de référence et sa position dans l'image mobile après recalage.
- Le score de Dice 2.1. Il évalue la superposition entre le masque de segmentation de la structure d'intérêt dans l'image fixe et son masque dans l'image mobile après recalage.

$$\text{Score de Dice} = \frac{2 \cdot |S_t \cap S_m|}{|S_t| + |S_m|} \quad (2.1)$$

où : S_t et S_m représentent les segmentations de la structure d'intérêt dans les images cible et mobile respectivement.

- La distance moyenne du plus proche voisin ("Nearest Neighbor distance"). Il évalue la distance moyenne entre les surfaces segmentées de la structure d'intérêt dans l'image fixe et dans l'image mobile après recalage.
- La distance de Hausdorff 2.2. Elle évalue le pire cas d'éloignement entre deux ensembles. Plus spécifiquement, si elle est faible, alors tous les points du premier ensemble sont proches du second ensemble, et réciproquement. Si elle est élevée, alors il existe au moins un point de l'un des ensembles qui est éloigné de l'autre ensemble.

$$HD(S_t, S_m) = \max \left\{ \sup_{i \in S_t} \inf_{j \in S_m} d(i, j), \sup_{j \in S_m} \inf_{i \in S_t} d(i, j) \right\} \quad (2.2)$$

Ces techniques artificielles d'évaluation du recalage posent souvent quelques difficultés en pratique. Premièrement, les jeux de données ne sont pas nécessairement fournis avec des points de repère ou des segmentations manuelles. Deuxièmement, quand bien même ces annotations manuelles sont disponibles, leurs précisions ne sont quasiment jamais garanties, puisque les annotations sont sujettes aux biais humains. Troisièmement, l'alignement de points de repère, généralement épars, renvoie surtout une information globale, pendant que la superposition des segmentations renvoie une information plus locale. Ainsi, les deux types de caractéristiques ne renvoient pas la même information. Ce qui entraîne parfois des comparaisons difficiles. Par exemple, lorsque les points de repère donnent avantage à une méthode par rapport à une autre, tandis que les segmentations donnent avantage à la seconde par rapport à la première. Il est alors intéressant de considérer les deux types de caractéristiques d'évaluation et de comparer les méthodes de recalage en fonction de l'usage que l'on souhaite faire de ce recalage.

2.3 Méthodes de pointe pour le recalage multimodal d'images échographiques

Cette section est focalisée sur les méthodes de pointe publiées pour le recalage multimodal d'images échographiques. En raison de la grande quantité de travaux dans ce domaine, le champ de cette revue est limité à l'échographie abdominale, et inclus tous les travaux qui ont abordés le recalage des reins.

2.3.1 Méthodes traditionnelles de recalage multimodal d'images échographiques

Il s'agit ici de méthodes automatiques de recalage n'utilisant pas de modèles d'apprentissage profond.

2.3.1.1 Les méthodes globales

Rappelons que dans un contexte de recalage global, le but n'est pas nécessairement d'obtenir un alignement précis de la structure d'intérêt dans les deux images. Il ne s'agit souvent que d'une étape d'initialisation qui sera ensuite affiné par un second recalage, ou d'un cas d'usage où un recalage imprécis, mais rapide, suffit. L'hypothèse de transformation ici est rigide ou une similarité. Dans la littérature, comparément aux méthodes d'affinement, il n'existe que peu de tentatives pour résoudre ce problème complexe de manière automatique. Les méthodes existantes s'appuient sur des caractéristiques extraites de la géométrie de l'organe, de son contexte anatomique ou de ses éléments vasculaires.

Utilisation des caractéristiques de géométrie.

Ces méthodes utilisent l'information de surface pour créer un pont entre les modalités d'imagerie et réduire leur écart de domaine. Cette information de surface est déterminée par segmentation, puis des caractéristiques permettant l'alignement des surfaces sont extraites.

Rusu et al. [110] ont proposé la méthode FPFH ("Fast Point Feature Histograms"). C'est un algorithme qui extrait des descripteurs géométriques locaux des points d'un nuage. Pour chaque point du nuage, il calcule une signature fondée sur ses coordonnées et sur l'orientation des vecteurs normaux dans son voisinage local. Cela capte la géométrie locale autour de chaque point. Les informations géométriques extraites sont représentées sous forme d'histogrammes. Ces histogrammes permettent de comparer et apparier les points entre deux nuages.

Une estimation de la transformation de recalage peut ensuite être réalisée par un algorithme de type RANSAC ("RANDOM Sample Consensus" en anglais) [106] ou TEASER++ [111]. Cette méthode peut être adaptée pour les images échographiques par la modélisation des surfaces en nuages de points 3D (Figure 2.18).

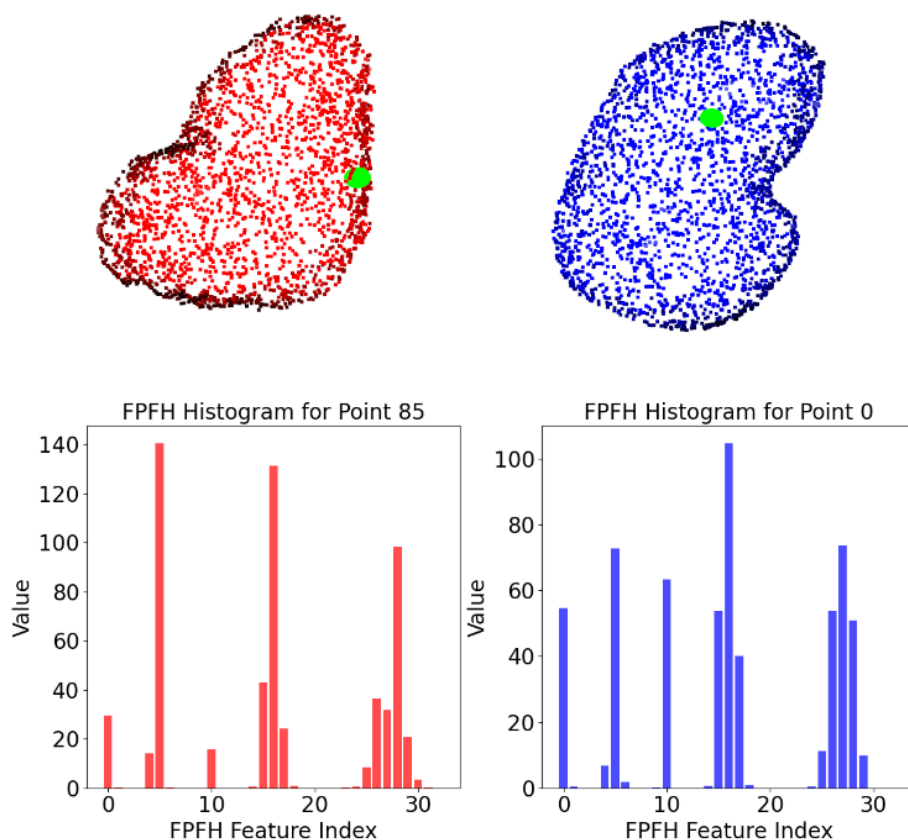


FIGURE 2.18 – Caractéristiques FPFH. La première ligne représente deux nuages de points issus de masques de segmentation du rein dans une image CT (rouge) et échographique 3D (bleu). La seconde ligne représente les histogrammes de caractéristiques FPFH du point vert dans chaque modalité.

Foruzan et al. [112] et Rehman et al. [113] utilisent les axes principaux des géométries. En effet, une analyse en composantes principales ACP des points d'un nuage ou d'un maillage résume assez bien la répartition spatiale du volume. Les auteurs alignent donc les axes principaux pour obtenir un recalage global. La précision de cette approche dépend de la qualité de la segmentation (voir Figure 2.19). Le Chapitre 4 de cette thèse illustre la sensibilité de cette approche aux ambiguïtés de symétrie.

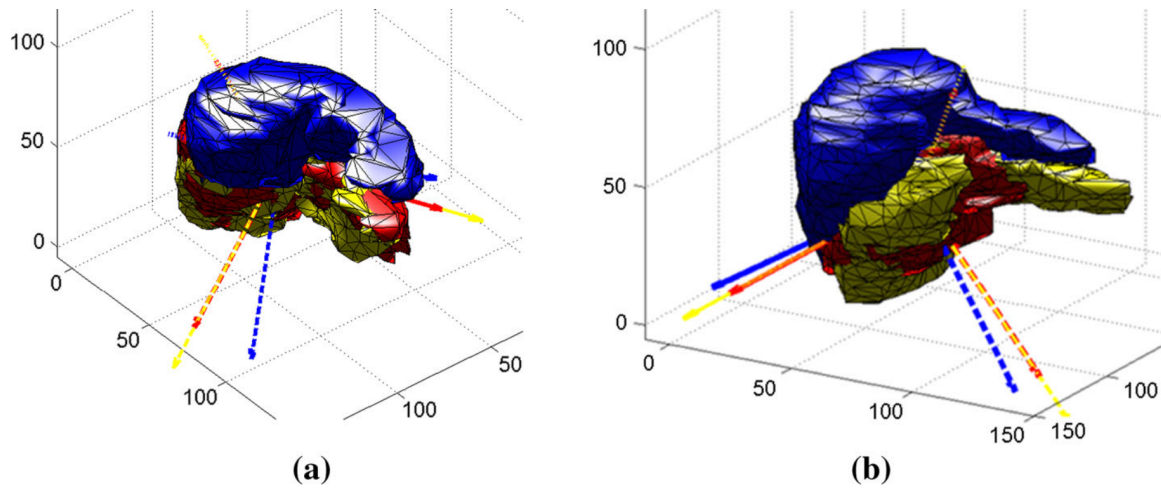


FIGURE 2.19 – Deux cas de recalage global de volumes de foie avec la méthode ACP. Le volume fixe et ses axes principaux sont en rouge. Le volume mobile et ses axes principaux avant recalage sont en bleu, et après recalage sont en jaune. (a) représente un cas où les masques de segmentation du foie sont complets, (b) représente un cas où les masques sont incomplets [112].

Utilisation des éléments vasculaires.

Les approches retrouvées dans la littérature dans ce contexte, n'optimisent pas la transformation de recalage. Elles estiment la transformation par simulations. Cifor et al. [63] estiment en deux temps, la transformation rigide qui aligne un volume CT "source" avec un balayage US "cible" formé d'un ensemble d'images (trames) 2D. Dans un premier temps, une image de l'ensemble US est prise comme référence et recalée manuellement au volume CT. Les recalages des autres images US sont ensuite estimés par recuit simulé (optimisation itérative par exploration aléatoire, acceptant à certaines itérations des solutions sous-optimales pour éviter les optima locaux). Ramalhinho et al. [114] ont utilisé le modèle vasculaire segmenté dans le CT (avec possibilité d'étiquettes multiples) pour simuler dans une étape préopératoire les différentes positions possibles de la sonde échographique laparoscopique 2D avec leur contenu vasculaire respectif (Figure 2.20). Les hypothèses de recalage global sont obtenues en trouvant les positions qui représentent le mieux le contenu vasculaire d'une nouvelle image échographique segmentée donnée. Le problème étant mal posé et ne garantissant pas l'unicité de la solution, un modèle de Markov caché discret est construit pour estimer la séquence la plus probable d'hypothèses de recalage global.

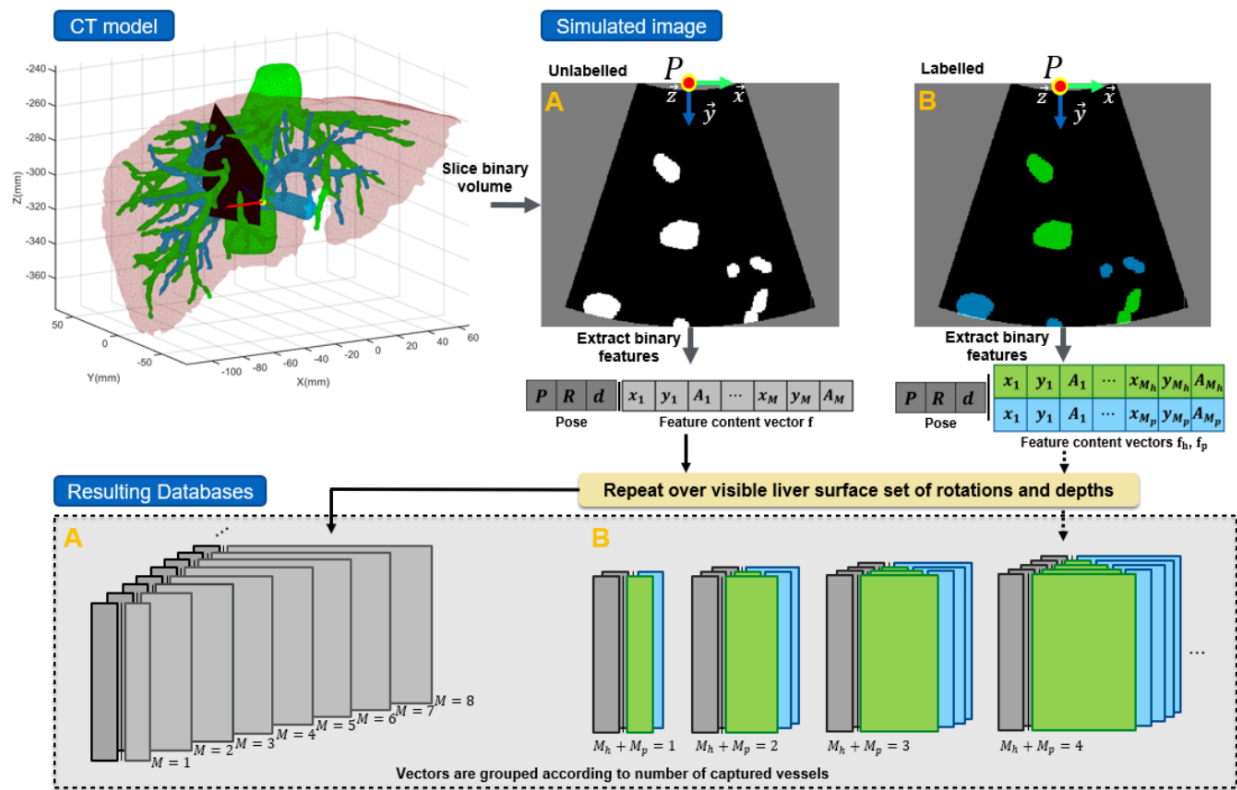


FIGURE 2.20 – Génération d'un jeu de données simulant des positions de sonde échographique et celles des structures vasculaires du foie, telle que décrit dans [114]. Pour chaque triplet (P =point de la surface hépatique, R =rotation, d =translation dans la direction de la profondeur), une image 2D ("Simulated image") contenant M sections de vaisseaux est simulée à partir du modèle CT ("CT model"). L'image A illustre le cas où les M sections de vaisseaux ont la même étiquette ("Unlabelled"). L'image B illustre le cas où les $M_h + M_p$ sections de vaisseaux sont étiquetées comme veine hépatique (h, en vert) ou veine porte (p, en bleu), M_h et M_p étant le nombre de sections de chaque type, respectivement : ("Labelled"). Le vecteur de caractéristiques $f = (x_1, y_1, A_1), \dots, (x_M, y_M, A_M)$ représente pour chaque section de vaisseau, la position 2D de son centroïde (x, y) et sa surface A .

2.3.1.2 Les méthodes d'affinement du recalage

Les méthodes d'affinement ont pour objectif d'estimer des transformations pour un alignement précis de la structure d'intérêt dans les deux modalités. Elles sont par exemple applicables en contexte de guidage d'aiguille en biopsie. Les approches développées sont nécessairement automatiques et fonctionnent généralement bien si une transformation initiale (globale) leur est fournie a priori, afin d'éviter leur convergence vers un optimum local plutôt que global. Il s'agit donc très souvent d'amélioration d'un recalage global (initial). Les hypothèses de transformation ici vont de rigide à déformable (transformation représentée en champ de déformation). Diverses méthodes d'affinement du recalage ont été

proposées, certaines utilisant directement et uniquement les images (méthodes iconiques), d'autres utilisant les surfaces, et enfin d'autres combinant les deux types d'information (méthodes hybrides).

Les méthodes utilisant les surfaces.

Les méthodes fondées sur l'utilisation des surfaces optimisent une fonction calculée à partir d'informations issues de segmentations de l'organe dans les images. L'utilisation des masques de segmentation ou de nuages de points permet de plonger les deux images dans un domaine unique de modalité. Ce qui facilite l'utilisation de métriques de similarité traditionnelles comme l'erreur moyenne au carré (MSE : "Mean Squared Error"). Toutefois, ce changement de domaine conduit à une perte de l'information donnée par les images. Parmi les plus populaires, il y a : ICP ("Iterative Closest Point") [115], CPD ("Coherent Point Drift") [116], BCPD ("Bayesian Coherent Point Drift") [117].

Le classique ICP, proposé par Arun et al. [115], est l'une des méthodes les plus utilisées pour le recalage de nuages de points. L'algorithme fonctionne en itérant entre l'appariement des points les plus proches et la minimisation de la distance entre ces points appariés. Il en existe des versions non-rigides [118, 119]. L'algorithme ICP a par exemple été utilisé dans [120]. Les auteurs présentaient les principes de la ponction rénale percutanée assistée par ordinateur, grâce à une planification 3D préopératoire précise, à partir d'images CT. Les segmentations du rein dans les images et l'initialisation du recalage ont été effectuées manuellement, tandis que l'algorithme ICP a été utilisé pour affiner l'alignement. Comme autres exemples, Song et al. [121], et Xing et al. [62], ont utilisé l'algorithme ICP pour affiner le recalage des lignes centrales de vaisseaux hépatiques issues d'images échographiques et CT/IRM. Dans [62], les lignes centrales ont été extraites par segmentation automatique et imparfaite de vascularisations hépatiques (Figure 2.21).

La méthode CPD ("Coherent Point Drift") introduite dans [116], ainsi que sa formulation bayésienne BCPD ("Bayesian Coherent Point Drift") [117], proposent un modèle probabiliste pour le recalage des nuages de points. Ces méthodes formulent l'alignement de deux nuages de points, disons M ("moving") and F ("fixed"), comme un problème d'estimation de la densité de probabilité d'un modèle hypothétique de mélange gaussien dont les centroïdes sont les points de l'ensemble M, tandis que l'ensemble F représente une observation de cette distribution de mélange. L'ajustement des valeurs de M est fait par maximum de vraisemblance, tout en forçant ces derniers à se déplacer de manière cohérente en tant que groupe. Ces méthodes se distinguent par leur capacité à gérer des déformations non rigides, ce qui le rend adaptées au recalage déformable.

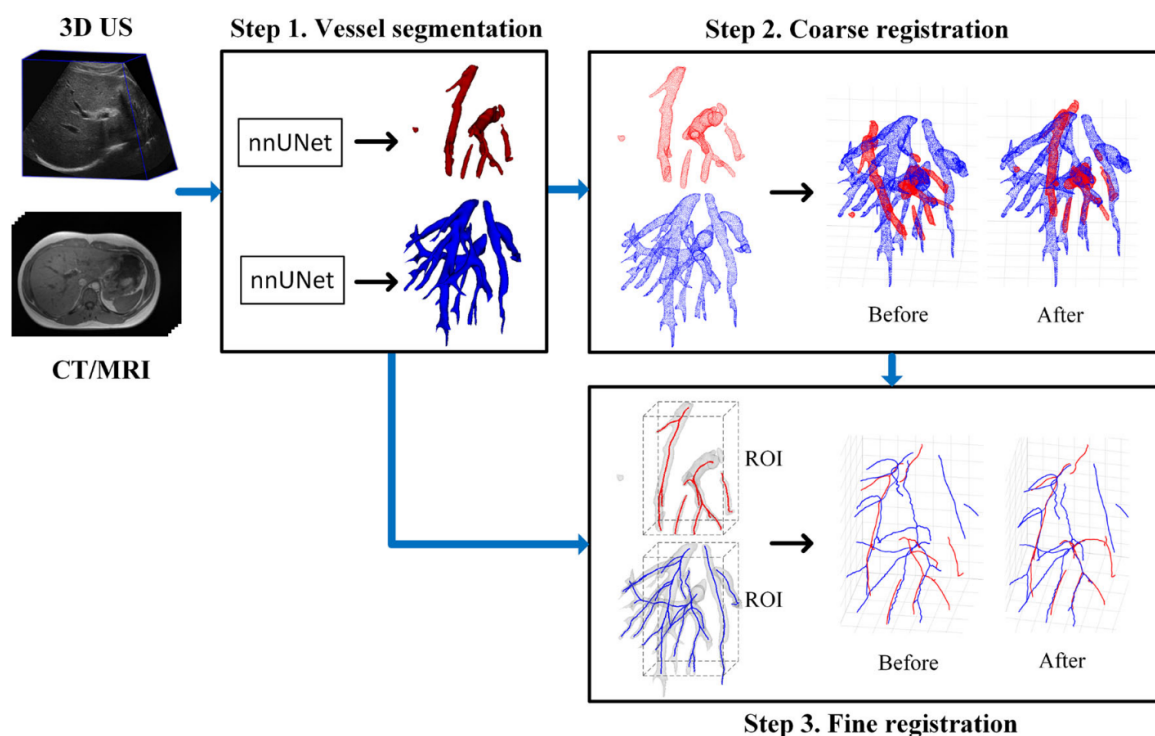


FIGURE 2.21 – Pipeline de segmentation et recalage de lignes de vaisseaux hépatiques. Module 1 ("Step 1") : segmentation automatique et imparfaite des vaisseaux par la méthode de segmentation nnU-Net [97]. Module 2 ("Step 2") : recalage global par la méthode BCPD. Module 3 ("Step 3") : extraction des lignes centrales des vaisseaux et affinement du recalage par la méthode ICP [62].

CPD a par exemple été utilisé dans [78] pour le recalage multimodal de lignes centrales du foie obtenues après segmentation automatique dans des volumes échographiques et d'IRM ou CT.

Il est toutefois à noter que, les méthodes ICP, CPD and BCPD sont sensibles aux minima locaux et nécessitent une bonne initialisation pour converger vers une solution correcte. C'est d'ailleurs à cause de cette dépendance à l'initialisation qu'elles n'apparaissent pas parmi les méthodes de recalage global de cette analyse, quoiqu'on les retrouve dans quelques rares travaux comme méthode globale, comme c'est le cas dans [62].

Les méthodes iconiques.

Les méthodes iconiques utilisent, dans leur processus d'optimisation, des mesures de similarité entre les valeurs d'intensité des deux images à aligner. Ces méthodes comparent directement les valeurs d'intensité des pixels ou des voxels entre les deux modalités. En contexte de recalage multimodal d'images échographiques, la différence de domaine importante avec les autres modalités (CT ou IRM), rend difficile cette comparaison. Ce qui conduit à l'échec des mesures de similarité classiques comme l'information mutuelle ou le rapport de

corrélation ("correlation ratio") [65]. La littérature propose différents travaux ne faisant pas intervenir de modèles d'apprentissage profond, les méthodes se focalisant en général sur deux objectifs principaux : réduire adéquatement le gap entre les valeurs ou les distributions d'intensité des images et construire une mesure de similarité robuste à cette différence.

Quelques contributions majeures sont à ce jour implémentées dans le logiciel ImFusion Suite (ImFusion GmbH, Munich, Allemagne). Comme métrique de similarité utilisée à cet effet, il y a par exemple la "Linear Correlation of Linear Combination" (LC^2), introduite par [65] dans la cadre du recalage d'images US et CT de foie. Les auteurs modélisent l'intensité d'un voxel dans l'image US comme une combinaison linéaire de deux signaux estimés à partir de l'image CT : l'échogénicité et la réflexion. Ils définissent alors une nouvelle métrique de similarité pour le recalage multimodal, (LC^2), qui évalue la corrélation entre l'image échographique et la combinaison linéaire estimée. La métrique (LC^2) a aussi été utilisée dans [68] pour l'affinement du recalage du foie dans les images US et d'IRM. Un autre exemple est la métrique "Local Normalized Cross-Correlation" (LNCC) introduite par [122] en contexte de geo-recalage multimodal afin de résoudre les défis posés par l'inversion partielle ou totale de contraste entre les modalités, et les ombres acoustiques. Ces difficultés sont similaires à celles rencontrées en contexte de recalage multimodal d'image échographique. De plus, une version précédente, la "Local Normalized Correlation" (LNC), avait été utilisée avec succès dans [123], pour le recalage d'images portales ([124]) et CT de différentes structures.

Les méthodes hybrides.

Ces méthodes combinent des informations géométriques et d'intensité.

Dans ce registre, plusieurs contributions pour le recalage du foie dans les images "échographiques et CT ou IRM du foie, ont été retrouvées. Par exemple, Banerjee et al. [69] ont employé une stratégie d'appariement de blocs avec une reformulation à eux de la métrique LC^2 qu'ils ont appelée "multiple correlation coefficient" (MCC). Cet appariement servait à établir les correspondances entre des points caractéristiques dans les deux modalités, sélectionnés à l'intersection du masque de segmentation du foie dans le CT avec l'image US entière grossièrement recalée, en raison de leur forte variance locale, et uniformément répartis à l'aide d'une structure en octree (voir Figure 2.22). Après élimination de paires aberrantes, une transformation affine de recalage est estimée à l'aide des paires de points retenues.

Comme autre exemple, Lange et al. [125] ont formulé le recalage d'images US et CT du foie comme un problème de minimisation de la dissimilarité des

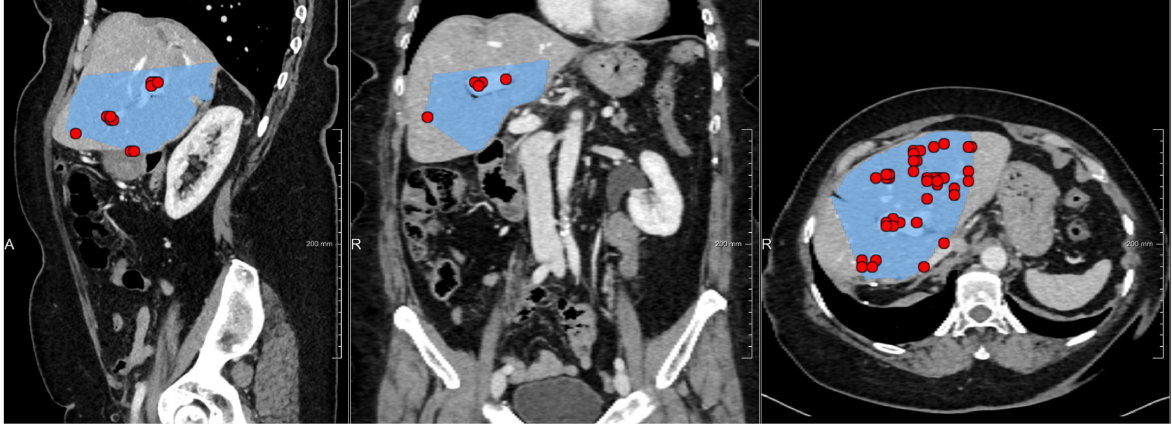


FIGURE 2.22 – Vues suivant 3 directions de la sélection des points caractéristiques (rouges). La zone bleue, intersection entre la région du foie dans l'image CT et l'image échographique, est la zone d'intérêt dans laquelle la sélection de points est réalisée [69].

distributions d'intensité, sous contrainte de superposition des points de bifurcations de vaisseaux. Ce qui a conduit à une amélioration des performances de recalage en moyenne, en comparaison à l'usage de points de repère seuls. Leur formulation du problème de recalage a été énoncée comme suit :

estimer la déformation $y = (y_1, y_2, y_3) : \Omega \rightarrow \mathbb{R}^3$ solution du problème d'optimisation sous contrainte

$$\min_{\mathbf{y}} \quad J(\mathbf{y}) = D(R, T(\mathbf{y})) + \alpha S(\mathbf{y} - \mathbf{y}_0) \quad (2.3)$$

$$\text{tel que :} \quad C_j(\mathbf{y}) = y(r_j) - t_j = 0, \quad j = 1, \dots, N. \quad (2.4)$$

où les $r_j \in \mathbb{R}^3$ sont les N points de repère dans l'image de référence $R : \Omega_R \rightarrow \mathbb{R}$, les $t_j \in \mathbb{R}^3$ sont les N points correspondants dans l'image déformée $T : \Omega_T \rightarrow \mathbb{R}$, avec $\Omega_R, \Omega_T \subset \mathbb{R}^3$, D mesure la dissimilarité entre l'image de référence R et l'image déformée notée $T(\mathbf{y})$, α est un terme de régularisation pour contrôler le lissage du champ de déformation.

2.3.2 Méthodes d'apprentissage profond appliquées au recalage multimodal

Cette section traite d'approches issues de la littérature, pour le recalage multimodale d'images échographiques intégrant partiellement ou totalement des modèles de réseaux de neurones profonds. Ces modèles peuvent être utilisés pour apprendre, à partir des images ou de leurs segmentations, des caractéristiques de la structure d'intérêt, communes aux différentes modalités. Ces modèles peuvent également être utilisés pour apprendre des transformations de

recalage entre les images. Enfin, ils peuvent apprendre à prédire des mesures de similarité multimodales. L'analyse s'intéresse à quelques contributions.

2.3.2.1 Les méthodes globales

Les méthodes globales existantes apprennent, soit des descripteurs géométriques de la structure d'intérêt, notamment à partir de segmentations, soit des descripteurs iconiques directement à partir des images.

Apprentissage de descripteurs géométriques.

Diverses stratégies d'apprentissage de descripteurs géométriques éparés ont été proposées. Elles nécessitent que les segmentations de l'organe dans les deux modalités soient fournies. Les contributions portant sur la région de l'abdomen, traitent davantage du recalage de modèles de foie issus d'images CT ou IRM, et de vidéos laparoscopiques. Le principal défi ici est le champ de vision limité du laparoscope, comme c'est le cas pour l'échographe. Il est donc intéressant de les mentionner.

Le modèle décrit dans [126] possède deux principaux modules. Le premier, appelé "Neighborhood Feature Fusion Module" (NFFM) commence par plonger les points des deux nuages (source et mobile) dans des espaces de caractéristiques abstraites. Les valeurs des caractéristiques sont ensuite agrégées par Max-polling sur le voisinage de chaque point. Les dépendances intra-nuage et inter-nuages entre points sont capturées grâce à des mécanismes d'auto-attention et d'attention croisée respectivement, permettant d'obtenir à la sortie de ce module, des caractéristiques de voisinage fusionnées. Le second module génère des correspondances entre points clés. En effet, il projette les caractéristiques issues du module NFFM dans un espace plus petit de "caractéristiques clés". Les caractéristiques clés sont utilisées pour réduire les deux nuages de points initiaux en deux ensembles de points clés homologues. Les caractéristiques clés sont aussi utilisées pour calculer la matrice de poids qui établit les correspondances entre les points clés (voir Figure 2.23). Le modèle est entraînable de bout en bout. Le modèle LiverMatch proposé un peu plus tôt par Yang et al. [128], repose, à quelques exceptions près sur le même principe que [126] : sous-échantillonner en points clés les nuages de points d'entrée en les plongeant dans des espaces de caractéristiques abstraites combinées afin de prendre en compte les dépendances intra-nuage et inter-nuages des points, les correspondances sont établies par une matrice d'appariement construite à partir de caractéristiques abstraites. Les deux principales différences étant le sur-échantillonnage des caractéristiques combinées avant le calcul de la matrice d'appariement, et le calcul d'un score de visibilité des points (voir Figure

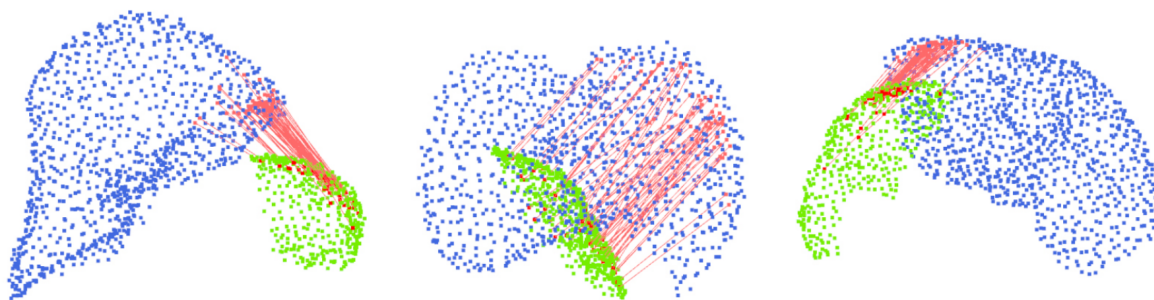


FIGURE 2.23 – Correspondances de points clés dans le jeu de données 3Dircadb [127]. Les points bleus représentent le nuage de points source, et les points verts représentent le nuage de points cible [126].

2.24).

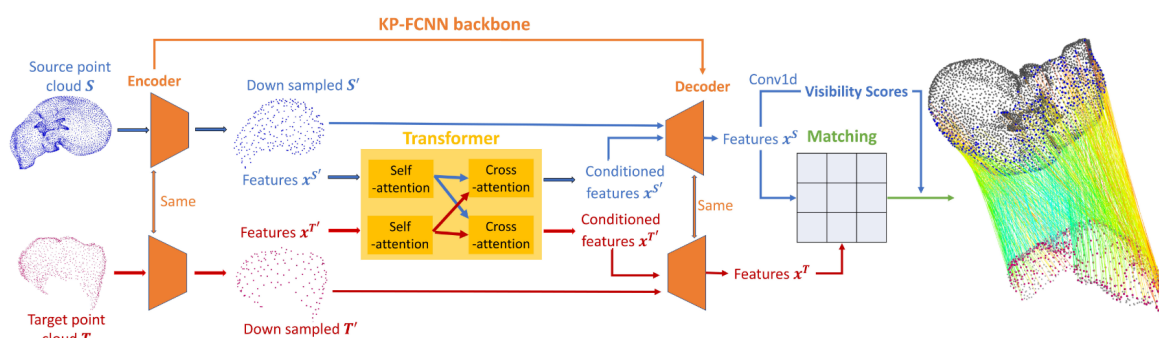


FIGURE 2.24 – Le modèle LiverMatch : 1) Encodeur - sous-échantillonnage des nuages de points d'entrée et extraction des caractéristiques associées ; 2) Transformer - mise à jour des caractéristiques pour obtenir des caractéristiques avec des informations intra et inter nuage ; 3) Décodeur - sur-échantillonnage des caractéristiques pour obtenir des caractéristiques par point ; 4) Appariement - calcule une matrice de confiance pour sélectionner les correspondances. Une convolution 1D supplémentaire décode les scores de visibilité [128].

A côté des approches précédentes, il existe des modèles d'apprentissage profond pour extraire des descripteurs géométriques. Ces modèles généralisent relativement bien sur une grande variété de formes et de capteurs. Pour cela, ils utilisent une représentation canonique des descripteurs de points, dans des référentiels locaux (LRF : "Local Reference Frame"). Et grâce à cette représentation canonique, ces descripteurs sont invariants aux transformations rigides du nuage. Ces modèles seraient donc aussi applicables pour le recalage d'organes abdominaux (rein, foie, prostate).

Dans cette catégorie, il y a le modèle DIP [129], où un réseau de type Point-Net est formé avec une stratégie siamoise (une copie du réseau pour chaque nuage) afin d'encoder, pour chaque paire de patches canonisés, une paire de descripteurs dont la similarité indique si les représentants de ces patches sont homologues ou pas. Ce qui permet ensuite d'établir les correspondances entre les

points des deux nuages. Les patches sont canonisés par un réseau intermédiaire de transformation également de type PointNet (TNet), qui produit en quelque sorte les LRFs associés. Le réseau intermédiaire produit en effet la transformation affine servant à recalculer les coordonnées des points d'un patch dans le LRF associé. Il est entraîné (optimisé) avec comme fonction de perte, la distance de Chamfer (distance moyenne symétrique entre chaque point d'un nuage et les points les plus proches de l'autre nuage) [130]. La Figure 2.25 illustre la stratégie d'entraînement du modèle DIP. Le modèle GeDi [131] quant à lui est une amélioration de DIP. GeDi utilise PointNet++ à la place de PointNet comme encodeur des descripteurs afin de rendre les représentations des patches plus informatives grâce aux champs réceptifs en cascade. De plus, pour atténuer le problème d'un LRF mal estimé avec TNet, un réseau de transformation quaternion (QNet) est formé. QNet étant intégré comme sous réseau de l'encodeur, il n'est pas nécessaire d'avoir un terme de perte dédié.

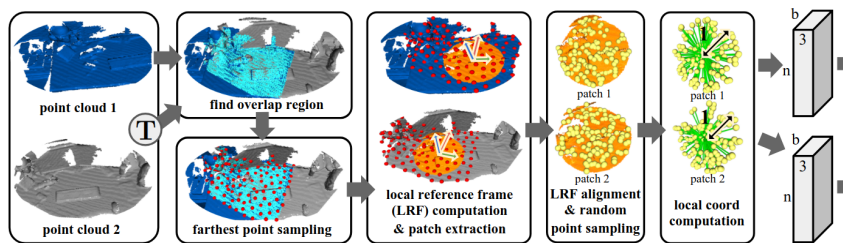


FIGURE 2.25 – Pipeline d'entraînement de DIP. Deux nuages de points qui se chevauchent sont alignés en utilisant la transformation de vérité terrain. Un ensemble de b points (en rouge) appartenant à la région de chevauchement (en cyan) est échantillonné en utilisant la méthode de "Farthest Point Sampling". Une approche siamoise est utilisée pour entraîner simultanément deux réseaux neuronaux profonds avec des paramètres partagés. Pour chaque branche, les opérations suivantes sont effectuées : (i) pour chaque point, un patch (en orange) avec un rayon τ_r est extrait et le référentiel local correspondant (Local Reference Frame, LRF) est calculé en utilisant les points du patch ; (ii) ce patch est rigidement transformé en utilisant le LRF et n points sont échantillonnés aléatoirement à partir du patch (points jaunes) ; (iii) les coordonnées de ces n points sont exprimées par rapport au centre du patch et normalisées pour avoir un rayon unitaire ; (iv) ces n points sont donnés en entrée au réseau profond pour apprendre le descripteur [129]. La perte finale est calculée comme une combinaison linéaire de la perte de Chamfer appliquée à la sortie de TNet et de la perte contrastive la plus difficile appliquée à la sortie du réseau.

Apprentissage de descripteurs iconiques

Ces méthodes d'apprentissage de descripteurs pour le recalage global prennent en entrée les images elles-mêmes. Elles n'utilisent pas les segmentations ou les nuages de points. L'intérêt pour ces méthodes est qu'elles écartent

l'étape intermédiaire de segmentation qui, certes, réduit le gap entre les deux distributions d'intensité, mais théoriquement entraîne une perte de l'information apportée par les images. De plus, l'étape de segmentation est coûteuse et son imperfection apporte de nouveaux biais dans la forme de la structure ou de l'organe. Le principe reste de détecter des points clés dans les images et de leur associer des descripteurs permettant de les apparier d'une modalité à l'autre. On y retrouve des méthodes de détection implicite de points clés, et d'autres de détection explicite.

Parmi les méthodes de détection implicite, il y a le modèle de Markova et al. [38]. Durant l'entraînement, il prend en entrée les images d'échographie (2D) et d'IRM (3D) à travers deux réseaux convolutifs distincts. Les réseaux produisent des descripteurs locaux sur une grille uniforme d'une résolution égale au huitième de celle des images d'origine, sur laquelle des points sont uniformément repartis. Pour chaque point sur la grille échographique, la transformation de vérité terrain (obtenue par recalage non-rigide) est appliquée pour obtenir la position correspondante dans l'image IRM. Ces correspondances sont ensuite utilisées pour définir des affectations souples entre points des grilles uniformes US et IRM, en prenant en compte les correspondances exactes, mais aussi les légères imprécisions afin de ne pas sur-pénaliser les correspondances qui sont légèrement décalées par rapport à la vérité terrain. Les descripteurs des points sont comparés par leur produit scalaire, et les résultats sont placés dans une matrice de scores de similarité. Un "Dual-Softmax" est appliqué pour transformer ces scores en probabilités de correspondance. Le réseau est entraîné de bout en bout en minimisant une fonction de perte construite sur la log-vraisemblance négative des appariements de vérité terrain, ce qui permet d'améliorer la cohérence des descripteurs entre points homologues à travers les modalités et d'entraîner les réseaux à générer des descripteurs similaires pour les points homologues. En inférence, les points sur la grille de chaque modalité sont appariés en choisissant les correspondances avec les scores les plus élevés. L'algorithme RANSAC est ensuite utilisé pour estimer la transformation de recalage.

Le travail beaucoup plus récent de [60] reprend l'idée de [38], rajoute un module à base de Vision Transformer à la suite des CNNs produisant les descripteurs locaux sur grille uniforme. Ce qui permet d'obtenir des descripteurs multimodaux et multi-dimensions. De plus, RANSAC est remplacé par un estimateur différentiable : "Differentiable Weighted Procrustes estimator" (DWP). Une technique utilisée pour trouver une transformation rigide optimale entre deux ensembles de points tout en tenant compte des poids individuels. La méthode a été appliquée au recalage global du rein dans les images US 2D et TDM sur fantôme (voir Figure 2.26).

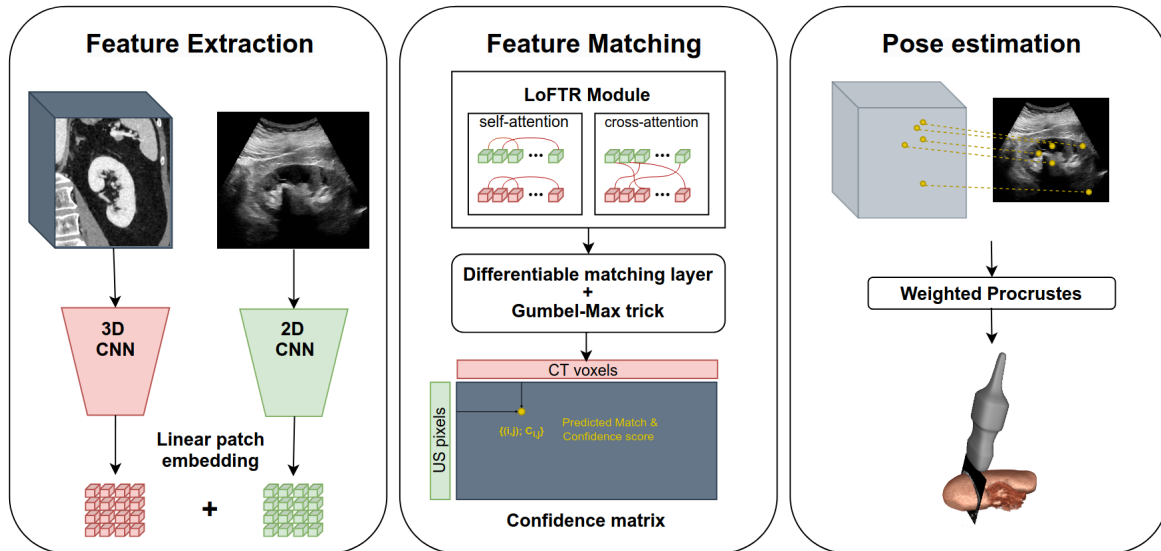


FIGURE 2.26 – Deux CNNs distincts extraient des images, des cartes de caractéristiques (descripteurs) converties en encodages linéaires de patches pour être traitées par le module LoFTR [132]. Ensuite, une couche de correspondance différentiable est utilisée pour appairer les caractéristiques transformées, ce qui donne une matrice de confiance. Enfin, les correspondances prédites et les scores de confiance sont utilisés pour prédire une transformation rigide. Les correspondances prédites et les scores de confiance sont utilisés pour prédire une transformation rigide avec la méthode Differentiable Weighted Procrustes (DWP) [60].

Parmi les méthodes de détection explicites de points clés pour le recalage global multimodal d'images échographiques, il y a la contribution de Fang et al. [67]. Ici, deux CNNs régresseurs sont utilisés pour localiser trois points de repère anatomiques dans les images US et IRM de la prostate, en estimant leurs coordonnées spatiales (voir Figure 2.27). Ces points sont utilisés pour estimer la transformation de recalage. L'une des premières investigations sur l'utilisabilité des CNNs régresseurs pour localiser plusieurs points de repère à la fois a été faite par [133].

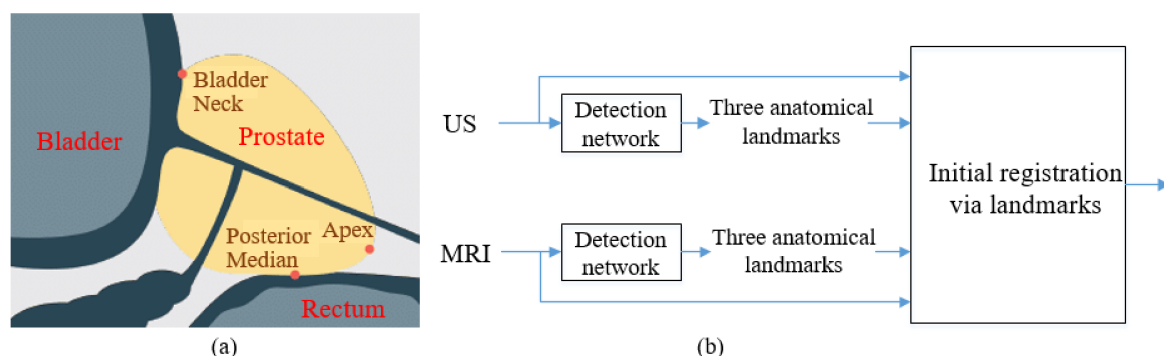


FIGURE 2.27 – (a) Les trois points de repère anatomiques de la prostate utilisés : l’apex, le col de la vessie, le médian postérieur. (b) Le processus pour détecter les points de repère anatomiques dans les deux modalités et pour calculer une transformation rigide approximative à l’aide de l’ajustement par moindres carrés [67].

2.3.2.2 Les méthodes d’affinement du recalage

Il s’agit de produire un recalage précis. La grande majorité des méthodes proposées dans la littérature sont focalisées sur ce type. Elles utilisent des informations de surface ou d’intensité. L’hypothèse est ici faite qu’un recalage initial est fourni a priori. Cette analyse se focalise donc sur ce qui se passe à partir de quand l’initialisation est fournie, comme, la création de données d’entraînement, les spécificités d’architectures, ou les particularités de fonctions de perte, ceci uniquement pour l’étape d’affinement du recalage.

Les méthodes utilisant les surfaces

Ces approches utilisent essentiellement les nuages de points de la structure ou de l’organe. Cette représentation géométrique de la structure peut être plus ou moins complète suivant la qualité de la segmentation. Le choix est fait ici de séparer l’étape de production des géométries d’avec celles du recalage proprement dit. Ainsi, en plus de ne s’intéresser qu’aux méthodes appliquées aux organes abdominaux, cette sous-section analyse plus spécifiquement ce qui se passe une fois les géométries fournies et initialement recalées.

Fu et al. [134] ont proposé un modèle d’apprentissage profond pour le recalage de nuages de points prenant en compte les contraintes biomécaniques de l’organe. Un réseau nommé ProRegNet est entraîné en utilisant des données simulées par éléments finis (FE) pour modéliser les déformations internes de la prostate entre les images IRM et échographiques trans-rectales, à partir de nuages de points volumétriques (Figure 2.28). Les modèles FE sont construits avec le logiciel ANSYS, avec comme conditions limites les correspondances de surface entre les deux modalités établies avec un algorithme de type ICP. La prostate est modélisée comme un matériau hyperélastique isotrope avec

un module de Young de 5 kPa et un coefficient de Poisson de 0,49. Une fois la simulation FE terminée, les déplacements nodaux obtenus sont étendus par interpolation aux champs de déplacement volumétriques ("deformation vector field" : DVF). Les DVF sont utilisés comme vérité terrain pour l'entraînement du réseau de prédiction des vecteurs de déplacement entre les points IRM et TRUS. L'architecture de ProRegNet est illustrée par la Figure 2.29. La fonction de

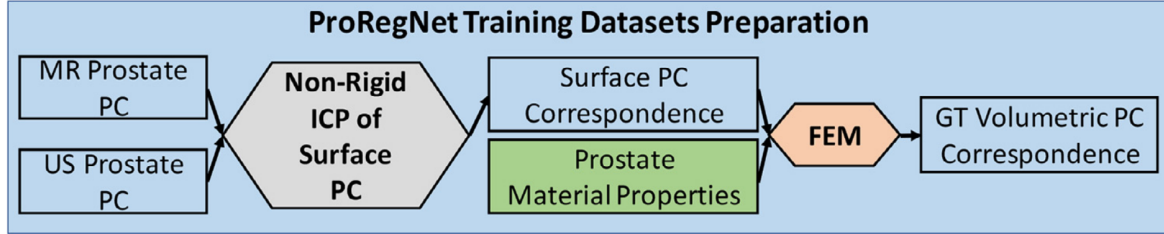


FIGURE 2.28 – Diagramme de la préparation des données d'entraînement du modèle ProRegNet [134].

perte combine l'erreur quadratique moyenne entre les vecteurs de déformation prédits par le réseau et la vérité terrain ("Intraprostatic PC Loss") et la distance de Chamfer ("Prostate Surface PC Loss") évaluant la correspondance entre les points de surface des nuages de points IRM et d'échographique transrectale (TRUS). Cette combinaison capture à la fois la précision des déplacements et la correspondance entre les points de surface.

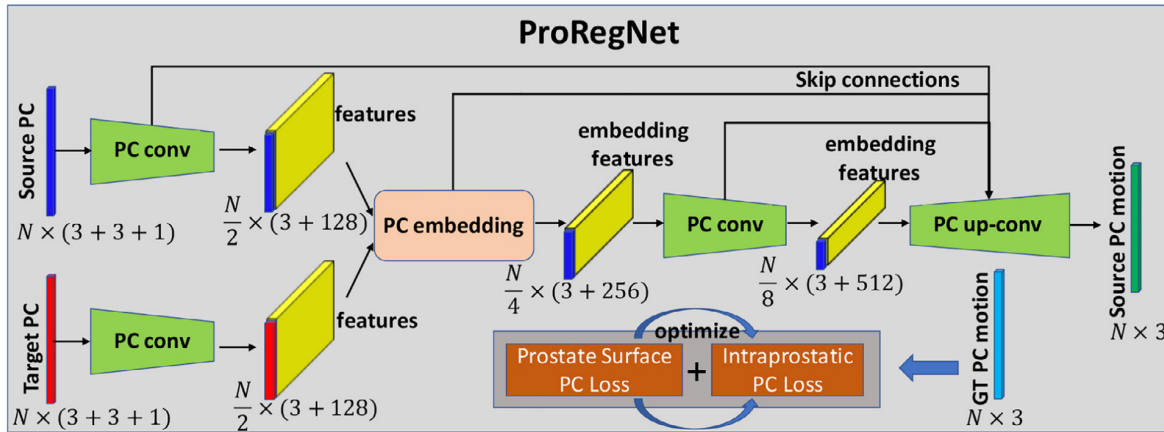


FIGURE 2.29 – Architecture de ProRegNet [134].

Baum et al. [135] ont proposé "Free Point Transformer" (FPT), une architecture de réseau de neurones profonds pour le recalage non-rigide de nuages de points. FPT ne suppose pas de contrainte explicite due à la proximité des points. Le modèle proposé, comportant deux modules, est entraîné par une approche siamoise (voir Figure 2.30) pour le recalage de la prostate. Le premier module est chargé d'extraire, via un réseau de type PointNet partagé, des caractéristiques globales de nuages de points source $\{p_s\}$ et cible $\{p_t\}$. Ces caractéristiques

globales sont associées par concaténation à chaque point d'un nuage du même domaine que la source $\{\tilde{p}_s\}$, qui sera utilisé dans le calcul de la perte, et l'ensemble est passé au second module constitué d'un MLP partagé suivant les $\{\tilde{N}_s\}$ différents points de $\{\tilde{p}_s\}$. Ce MLP prédit un vecteur de déplacement pour chaque point de $\{\tilde{p}_s\}$ afin que l'ensemble soit aligné avec la cible. Cet alignement est évalué par une fonction perte entre le transformé $\{\tilde{p}'_s\}$ et un nuage du même domaine que la cible $\{\tilde{p}_t\}$. L'intérêt d'utiliser lors du calcul de la perte des nuages des domaines source et cible, mais non nécessairement égaux aux nuages source et cible, est que l'on peut régler le niveau de supervision lors de l'entraînement.

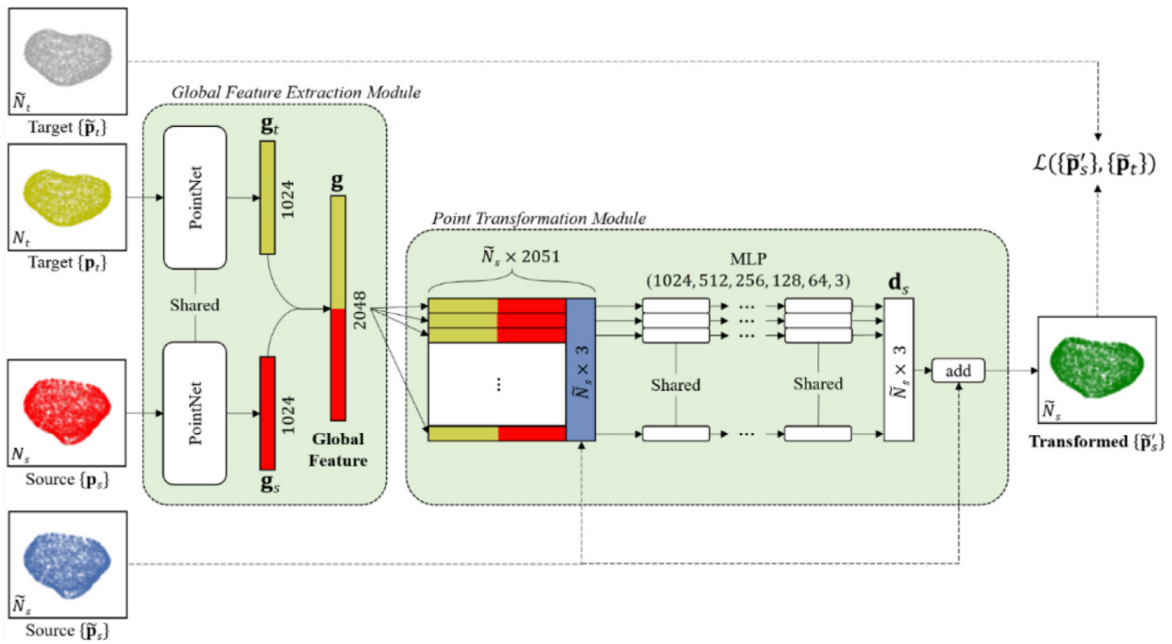


FIGURE 2.30 – Architecture de "Free Point Transformer" (FPT) [135].

Les méthodes iconiques et mixtes.

Dans cette sous-section, sont analysées les approches utilisant directement les images : méthodes iconiques, ou combinant les images avec des informations supplémentaires comme les segmentations. Les premiers modèles d'apprentissage profond pour le recalage iconique multimodal tentaient d'estimer un champ de déformation dense ("Dense Deformation Field" : DDF) représentant le déplacement de chaque voxel entre l'image source et l'image cible, en optimisant une fonction de perte calculée sur les masques de segmentation de l'organe. Il s'agit de méthodes conduites par les labels. La formation de tels modèles requiert comme données, les images et les masques de segmentation de la structure d'intérêt dans les deux modalités. Toutefois, les masques de segmentation ne sont pas utilisés à l'inférence. Cette configuration a été utilisée dans

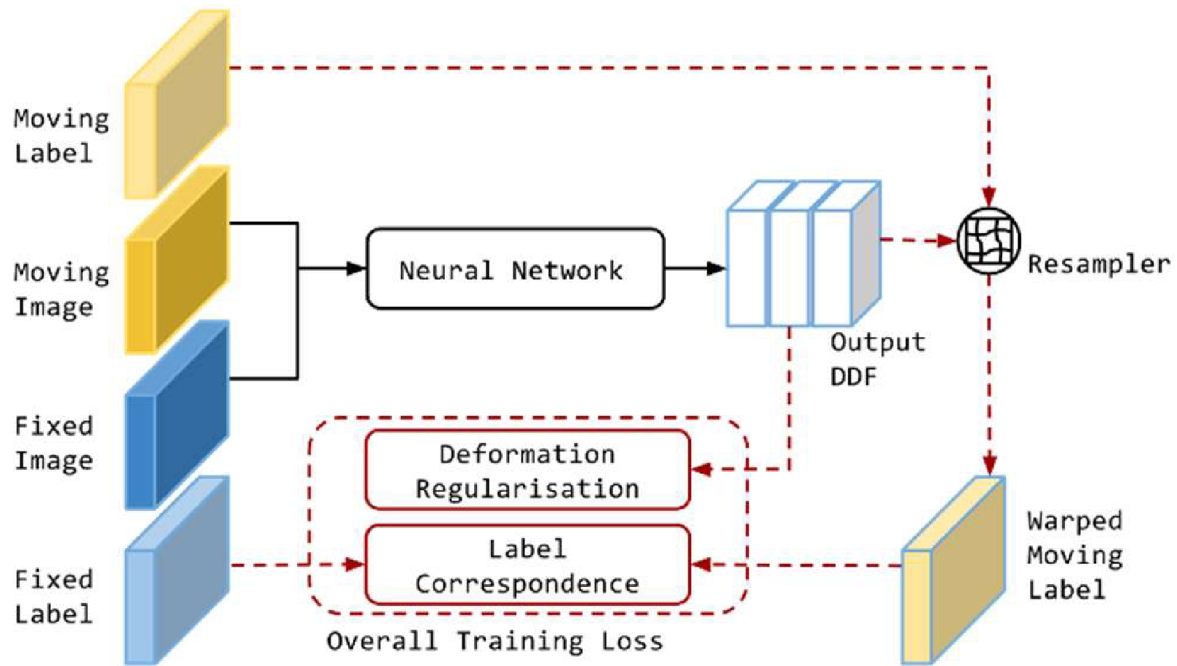


FIGURE 2.31 – Modèle général guidé par les labels [136].

[136] et [71] pour le recalage de la prostate dans les images échographiques transrectales et d'IRM. La description générale du réseau est représentée à la Figure 2.31. Il prend en entrée les images source ("Moving Image") et cible ("Fixed Image"), produit un champ de déformation ("Output DDF") qui est appliqué au masque vérité terrain de l'image source ("Moving Label"), et l'erreur est évaluée entre le masque vérité terrain de l'image cible et le masque déformée précédemment. La fonction de perte dans [136] était l'entropie croisée combinée à un terme contrôlant la régularité du DDF. Et le module dénommé "Neural Network" était la combinaison d'un sous-réseau prédisant une transformation affine globale avec un sous-réseau local, entraîné conjointement, prédisant un DDF local (voir Figure 2.32). C'est la composition des deux prédictions qui permettait d'estimer le DDF final. Ce qui conduisait à une difficulté de propagation du gradient. Pour résoudre cette difficulté, Hu et al. [71] ont remplacé cette combinaison par un seul réseau plus simple qui prédit directement le DDF. En outre, Hu et al. [71] ont proposé deux alternatives pour remplacer le terme d'entropie croisée : l'entropie croisée multi-échelle, et la perte de Dice multi-échelle.

A la suite de ces tentatives, il y a eu le premier modèle d'apprentissage profond pour le recalage multimodal déformable d'images médicales qui permet d'utiliser efficacement des jeux de données non annotés : VoxelMorph [75]. Il optimise directement une fonction de perte utilisant la similarité des images. L'architecture de VoxelMorph, à base de réseaux convolutifs, est représentée par la Figure 2.33. Elle comporte plusieurs modules avec des fonctions spécifiques :

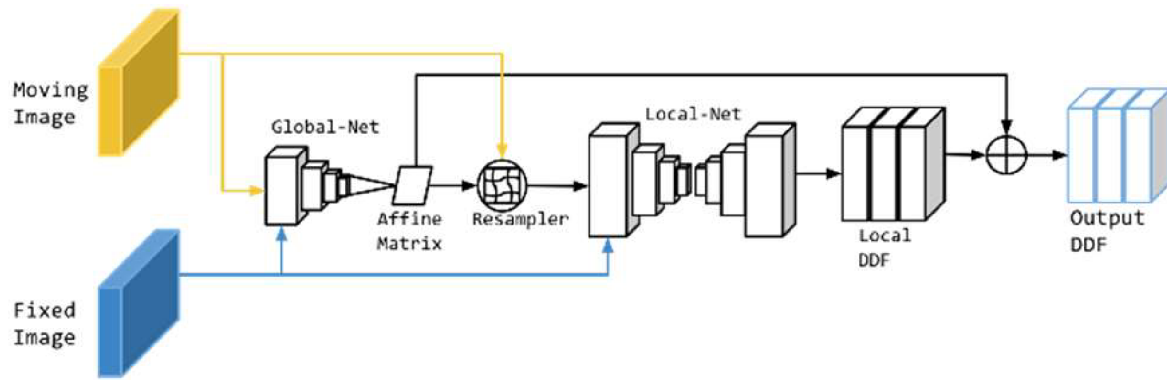
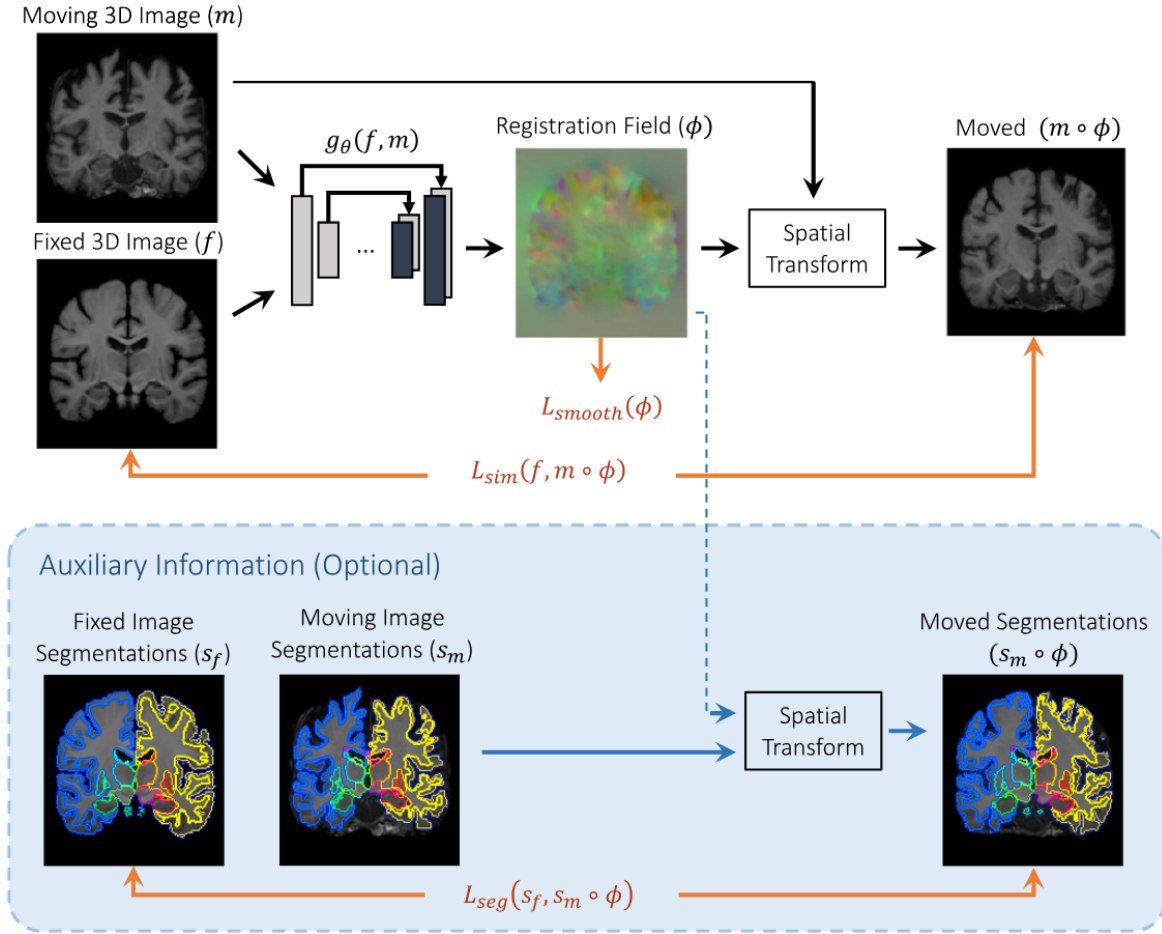


FIGURE 2.32 – Architecture du réseau combinant un sous-réseau prédisant une transformation affine globale et un sous-réseau local prédisant un DDF local [136].

1. Un CNN ($g_\theta(f, m)$) "Encoder-Decoder" pour extraire les caractéristiques abstraites des images mobile (m) et fixe (f).
2. Un champ de déformation (ϕ) qui est en fait une fonction paramétrique entraînable prenant une valeur spécifique (le champ de déformation à appliquer sur l'image mobile) suivant la paire d'images fournie en entrée.
3. Une fonction différentiable de transformation spatiale (Spatial Transform [137]), qui est un sous-réseau de neurones permettant la manipulation spatiale (appliquer directement des transformations spatiales) des données au sein du réseau sans supervision supplémentaire ou modification du processus d'optimisation. Elle est appliquée à l'image mobile pour la déformer ($m \circ \phi$) de sorte qu'elle soit similaire à l'image fixe.
4. Une fonction de perte fondée sur la similarité entre les images fixe et déformée ($L_{sim}(f, m \circ \phi)$). Elle doit être robuste à la différence de domaine entre les deux images. De plus, elle est associée à une fonction de régularisation ($L_{smooth}(\phi)$) qui impose une contrainte sur le champ de déformation estimé pour qu'il soit lisse et réaliste, en évitant des déformations irréalistes entre l'image source et l'image cible.
5. Enfin, la possibilité d'utiliser, en plus des images, des segmentations (lorsqu'elles sont disponibles), dans le calcul de la fonction de perte totale, en rajoutant un terme ($L_{seg}(s_f, s_m \circ \phi)$) calculant la perte entre le masque fixe (s_f) et le masque déformé ($s_m \circ \phi$).

Bien que la littérature ne fournisse pas à ce jour, de comparaison entre les méthodes utilisant les surfaces et VoxelMorph, ce dernier représente une avancée majeure vers le recalage multimodal iconique grâce à l'apprentissage profond.

Le succès des modèles à base de "Vision Transformer" a conduit à une adaptation de VoxelMorph fondée sur ces modèles : TransMorph [76]. La différence

**FIGURE 2.33** – Architecture de VoxelMorph [75].

fondamentale entre les modèles de vision utilisant les CNNs (Convolutional Neural Networks) et ceux utilisant les Vision Transformers (ViTs) réside dans la manière dont ils capturent les relations spatiales dans les images. Les CNNs utilisent des convolutions locales pour extraire des caractéristiques à partir de petites régions de l'image, ce qui leur confère une capacité à capturer des dépendances locales et une invariance spatiale, mais avec des limitations à long terme (en profondeur) pour les relations globales. En revanche, les ViTs utilisent des mécanismes d'auto-attention pour capturer des relations à longue portée dans l'image dès le début, en traitant l'image comme une série de patches, ce qui leur permet de modéliser des dépendances globales de manière plus explicite, mais avec des besoins en données et en calcul souvent plus importants.

Le diagramme décrivant l'architecture TransMorph (voir Figure 2.34) est assez similaire à celui de VoxelMorph. La principale différence est que le réseau CNN "Encoder-Decoder" de VoxelMorph a été remplacé par un ViT "Transformer Encoder (encodage à base de blocs Swin-Transformer [138]) - CNN Decoder (décodage à base de CNN)". L'implémentation de TransMorph propose une

option dédiée au recalage affine qui a été utilisée dans ce travail. Pareillement

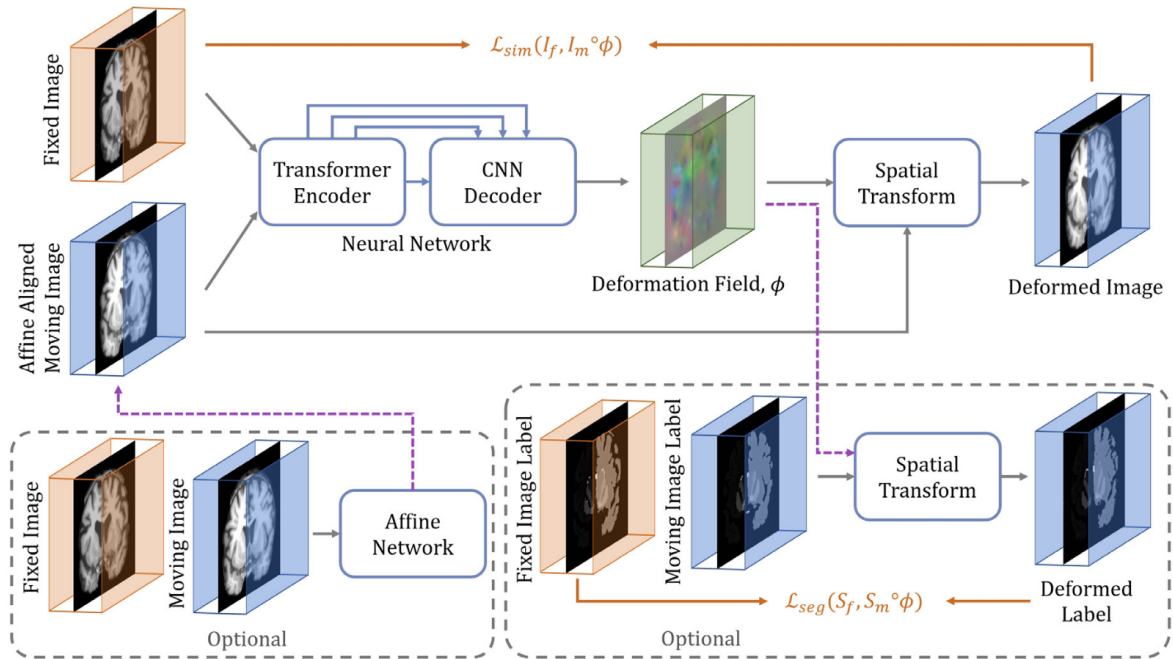


FIGURE 2.34 – Architecture de TransMorph [76].

à VoxelMorph, la littérature ne fournit pas, à ce jour, de comparaison entre les méthodes utilisant les surfaces et TransMorph. Dans le Chapitre 3 de cette thèse, ce dernier est expérimenté parmi les méthodes de référence pour l'affinement du recalage multimodal d'images échographiques 3D du rein.

2.3.2.3 Les métriques de similarité apprises

Il a déjà été indiqué que les approches de recalage multimodal s'activent principalement sur deux axes de recherche : réduire de quelque façon que ce soit le gap entre les différents domaines dont sont issues les images, et développer des mesures de similarité robustes à cette importante différence entre les distributions d'intensité. Des approches utilisant l'apprentissage profond ont également été proposées pour construire des mesures de similarité.

A cet effet, Haskins et al. [95] ont proposé d'entraîner un CNN pour régresser une métrique de similarité pour le recalage rigide d'images échographiques transrectales et IRM de la prostate (voir Figure 2.35). Ce réseau prend en entrée une paire d'images et prédit la distance euclidienne moyenne entre les points de la surface, qui est utilisée comme mesure de similarité. Les données d'entraînement sont issues de recalages rigides manuellement réalisés par un expert. Une fois la métrique formée, sachant que ni sa convexité ni sa régularité ne sont garanties, elle doit être utilisée de manière robuste pour le recalage. Les auteurs estiment une transformation initiale en utilisant "l'évolution différentielle"

(algorithme itératif d'optimisation qui, à chaque itération, crée de nouvelles solutions candidates par mutation et recombinaison des anciennes) [139]. Le recalage initial est ensuite passé à une méthode de type Newton, par exemple l'algorithme BFGS (Broyden–Fletcher–Goldfarb–Shanno) [140], pour affiner la transformation.

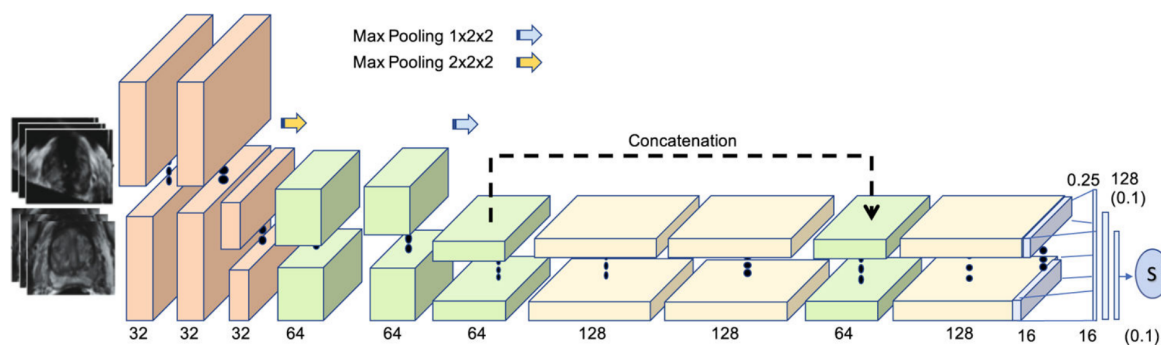


FIGURE 2.35 – Architecture du CNN conçu et utilisé pour apprendre la métrique de similarité [95].

Une contribution un peu plus récente est celle de Ronchetti et al. [141]. Les auteurs proposent d'entraîner un réseau pour approximer la métrique "Linear Correlation of Linear Combination" (LC^2) introduite dans [65]. L'idée d'approximer la métrique LC^2 en particulier vient de ce qu'elle permet déjà d'obtenir de bonnes performances en termes de recalage iconique multimodal, notamment en présence d'images échographiques. Les auteurs souhaitent réduire le temps de calcul de la métrique LC^2 en confiant la tâche à un réseau de neurones profond. L'approche a justement permis d'accélérer le calcul de la métrique LC^2 , jusqu'à la rendre utilisable en contexte clinique temps réel, tout en conservant une bonne précision. Le modèle neuronal utilisé dans cette approche est un CNN régresseur. Les données d'entraînement sont des paires de patchs échantillonnées aléatoirement et dont les valeurs de LC^2 vérité terrain sont automatiquement calculées. La similarité totale entre les images entières est alors vue comme une somme pondérée des similarités locales à l'échelle du patch. Le réseau optimise l'erreur quadratique.

2.3.3 Conclusion

L'état de l'art des approches de recalage multimodal d'images échographiques de la région de l'abdomen montre une activité importante autour du foie et de la prostate, par rapport au rein. Avec un passage progressif vers des méthodes complètement automatiques, notamment grâce aux modèles d'apprentissage profond. Néanmoins, la question du recalage automatique

du rein reste un problème non résolu, autant que celui du foie et de la prostate. La localisation et l'orientation du rein dans une image échographique est particulièrement variable. Non seulement l'organe lui-même est sujet aux mouvements de l'abdomen (la respiration par exemple), mais aussi sa position dans le volume va dépendre du praticien qui réalise l'acquisition. De plus, il n'est souvent pas capturé de manière exhaustive à cause d'occultations par les côtes. Enfin, en l'absence d'information environnante (structures aux alentours), ce qui est souvent le cas, il peut être très difficile de distinguer les pôles supérieur et inférieur du rein, en raison de la symétrie de l'organe. Toutes ces difficultés font du problème du recalage automatique multimodal du rein dans les images échographiques, un problème de recherche ouvert. La compréhension automatique de sa position initiale dans les images échographiques étant une priorité. De plus, les méthodes d'apprentissage profond requièrent des données importantes et correctement annotées pour leur entraînement. Ceci motive les contributions faites dans cette thèse et développées dans les chapitres suivants.

3. LE JEU DE DONNÉES TRUSTED

Résumé du chapitre

Ce chapitre présente le jeu de données TRUSTED (Tridimensional Renal Ultra Sound Tomodensitometrie Dataset), développé, analysé et proposé en libre accès pour stimuler la recherche sur les méthodes automatiques de segmentation et de recalage multimodal ou inter-modal des reins. Ce jeu de données unique comprend des images échographiques tridimensionnelles (3D) transabdominales et scanner appariées, issues de 48 patients. Les images sont accompagnées de doubles annotations indépendantes et précises, réalisées par deux radiologistes expérimentés. Ces annotations incluent des segmentations et des points de repère anatomiques du rein.

Les vérités terrain ont été estimées à partir de ces doubles annotations, indiquant, pour les segmentations, des concordances inter-annotateurs moyennes supérieures à 93 % en termes de score de Dice.

Afin de démontrer l'utilité du jeu de données pour les méthodes de segmentation, quatre modèles de segmentation de pointe, à base de réseaux de neurones profonds, ont été entraînés et évalués : 3D U-Net, VNet, nnU-Net et le "Vision Transformer" CoTr. Les performances des modèles ont montré des scores de Dice allant de 79,63 % à 90,09 % pour les images CT et de 70,51 % à 80,70 % pour les images US.

Pour évaluer l'utilité du jeu de données pour les méthodes de recalage inter-modal, trois méthodes d'affinement du recalage ont été testées en combinaison avec une initialisation manuelle, dont deux utilisant les surfaces issues de segmentation automatique du rein : "Iterative Closest Point" - ICP et "Bayesian Coherent Point Drift" - BCPD, une méthode iconique : ImFusion Suite, et le "Vision Transformer" TransMorph. Les méthodes ont été comparées en utilisant deux modèles de transformation spatiale : la transformation rigide et la transformation affine. Parmi elles, la méthode BCPD a obtenu les meilleurs résultats avec une erreur sur cible de recalage ("Target Registration Error" - TRE) moyenne de 4,47 mm et un score de Dice de 84,10 %.

Ce jeu de données, disponible en libre accès, représente une ressource précieuse pour le développement et la validation de méthodes de segmentation et de recalage inter-modal, et devrait permettre de débloquer et de stimuler des progrès importants dans ces domaines. En utilisant le jeu de données TRUSTED,

des méthodes de recalage inter-modal déjà disponibles pourront être testées ou améliorées, mais également de nouvelles méthodes pourront être développées avec l'objectif de parvenir à un déploiement en contexte clinique.

Le chapitre a cinq sections principales. La première section présente un état de l'art en matière de jeux de données publiques d'images échographiques ou pouvant être utilisés par des méthodes de recalage multimodal. La deuxième section décrit les protocoles d'acquisition et d'annotation du jeu de données. La troisième évalue la qualité des annotations. La quatrième section décrit les expérimentations validant l'utilité du jeu de données pour les modèles de segmentations automatiques. La cinquième décrit les expérimentations validant son utilité pour les méthodes de recalage inter-modal (ou multimodal).

3.1 Introduction

Les méthodes de pointe pour le recalage inter-modal (ou multimodal) d'images médicales ainsi que pour la segmentation utilisent désormais des modèles de réseaux de neurones profonds (DNNs - pour "Deep Neural Networks" en anglais) [142, 143]. La formation et l'évaluation des plus performants de ces modèles nécessitent des données suffisantes et représentatives qui doivent être soigneusement préparées, annotées et évaluées. De plus, une comparaison objective de méthodes, d'architectures ou de modèles nécessite l'utilisation d'une base de données commune. La disponibilité limitée de données publiques ainsi qu'une évaluation insuffisante de celles déjà disponibles constituent un important goulot d'étranglement pour convertir les systèmes développés par les chercheurs en applications cliniques [144, 145]. Un autre besoin de la recherche dans le domaine du recalage inter-modal d'images (IMIR - pour "Inter-Modal Image Registration" en anglais) est celui de données appariées, c'est-à-dire, pour un même patient, des données issues des deux modalités d'imagerie.

À ce jour, il existe très peu de jeux de données publiques d'images échographiques 3D de la région abdominale humaine, avec des images de tomodensitométrie (TDM, ou CT - pour "Computed Tomography" en anglais) ou d'imagerie par résonance magnétique (IRM) appariées, pouvant être utilisés pour entraîner et évaluer des méthodes d'IMIR à base de DNNs. Deux jeux de données publiques de ce type ont été retrouvés, et il s'agit d'images échographiques transrectales (TRUS - pour "Transrectal Ultrasound" en anglais) de la prostate, appariées avec des images d'IRM. Les échographies transrectales diffèrent des échographies transabdominales (généralement utilisées pour les procédures percutanées dans la région de l'abdomen) par la voie d'accès de la sonde.

Les premières, accèdent aux structures internes via le rectum, traversent moins de tissus et fournissent de meilleures images. En revanche, les échographies transabdominales accèdent aux structures internes via la paroi abdominale. Ce qui atténue les ondes ultrasonores à travers les tissus et les os des côtes, affectant la qualité des images. Toutefois, les échographies transabdominales offrent une meilleure vue d'ensemble des organes, ce qui, parfois, peut servir comme information de contexte.

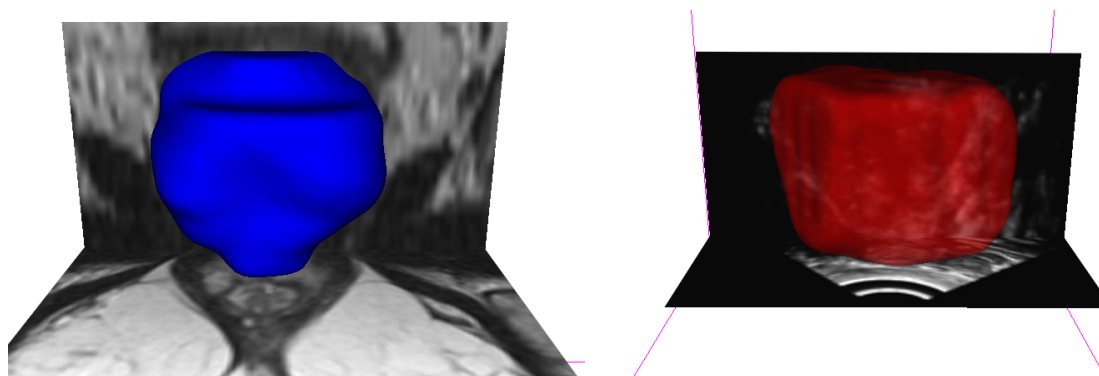


FIGURE 3.1 – Paire d'images IRM et US 3D d'un individu du jeu de données Mu-RegPro [43], ainsi que les annotations de segmentation de la prostate en vert dans l'IRM et en rouge dans l'image US 3D.

Le plus récent d'entre les deux jeux de données mentionnés ci-dessus est l'ensemble de données associé au défi Mu-RegPro de la conférence internationale MICCAI 2023 [43]. Il propose des images appariées de 108 patients, avec des segmentations manuelles de l'organe (voir Figure 3.1). Le fait qu'il s'agisse d'un défi MICCAI 2023 illustre l'intérêt de la communauté pour les jeux de données d'images inter-modales incluant l'échographie 3D. Un jeu de données antérieur constitué d'images issues de 3 patients a été publié par Fedorov et al. [146], avec des segmentations manuelles et des points de repère anatomiques (voir Figure 3.2). Le nombre assez réduit de cas ici rend son utilisation de manière isolée à des fins d'entraînement très limitée.

Des jeux de données publiques d'images unimodales d'échographie abdominale ont été publiés pour la recherche sur le recalage ou la segmentation. Par exemple, DeLuca et al. [147] ont publié comme jeu de données associé au défi CLUST de MICCAI 2014, des séquences 2D et US 3D du foie de 54 volontaires humains, avec des points de repère anatomiques intra-hépatiques pour le recalage US unimodal intra-patient. Riscinti et al. [148] ont publié le "POCUS Atlas", un atlas participatif et en libre accès d'images et de vidéos d'échographiques 2D couvrant différentes régions ou structures anatomiques (musculo-squelettique, gastro-intestinale, abdominale, oculaire et pulmonaire). Il y a différentes formes

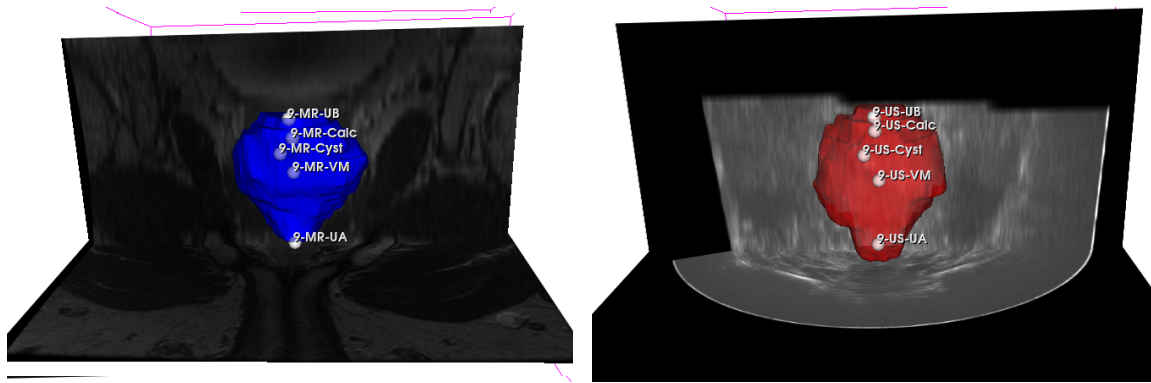


FIGURE 3.2 – Paire d’images IRM et US 3D d’un individu du jeu de données de [146], ainsi que les annotations de segmentation de la prostate en vert dans l’IRM et en rouge dans l’image US 3D, ainsi que des points de repère anatomiques.

d’annotations : segmentations sémantiques, étiquettes d’image, points de repère anatomiques. Le jeu de données a été conçu principalement à des fins éducatives. Un contrôle de la qualité des annotations n’a pas été réalisé. Xu et al. [149] ont récemment publié un ensemble de données d’images US 2D unimodales de 735 images de foie, avec des segmentations sémantiques de la région du foie et de lésions hépatiques.

Comme jeu de données publiques multimodales et appariées dédiées au recalage et à la segmentation d’organes dans des images CT ou d’IRM, il y a la contribution de Siebert et al. [150]. Elle présentait 16 images CT et IRM abdominales appariées provenant du dépôt TCIA [151]. Dans les images, le foie, la rate et les reins ont été segmentés manuellement. En revanche, divers jeux de données non appariés d’images CT et d’IRM abdominales ont été publiés. Par exemple, Amos et al. [152] ont publié le jeu de données AMOS, contenant 500 images CT et 100 images d’IRM (multi-centres, multi-phases et multi-pathologies). 15 organes abdominaux ont été segmentés dans chaque image (rate, reins, vésicule biliaire, œsophage, foie, estomac, aorte, veine cave inférieure, pancréas, glandes surrénales, duodénum, vessie et prostate ou utérus). Kavur et al. [153] fournissent le jeu de données CHAOS Challenge contenant 40 images CT et 120 images IRM, avec des segmentations manuelles du foie, des reins et de la rate. Jimenez-del-Toro et al. [154] présentent le défi ISBI 2014 Visceral, pour la segmentation automatique de structures abdominales (14 structures) et la localisation de points de repère (jusqu’à 53 points de repère sur différentes structures). Le défi s’appuie sur un jeu de données de 30 images CT et 30 images d’IRM non appariées. La présentation de ce défi a été suivi par une étude évaluant de manière détaillée des méthodes concurrentes pour la segmentation et la localisation de point de repère [155].

Dans le domaine plus large de la recherche en traitement d'images médicales, les travaux portant sur des jeux de données bien préparées, combinés à une évaluation de méthodes concurrentes, ont une importance croissante. En effet, ils jouent souvent un rôle moteur dans l'amélioration de l'état de l'art et l'ouverture de nouvelles voies de recherche. Depuis 2020, plusieurs travaux de ce type ont été publiés dans des revues majeures. Par exemple, dans le journal "Medical Image Analysis ont été publiés : CADIS (jeu de données sur la cataracte) [156], Padchest (image de la poitrine) [157], WORD (organes abdominaux en général) [158], LiTS (foie et tumeurs hépatiques) [159], MIDOG (images histopathologiques) [160]. Il faut ajouter à celles-ci, les publications portant sur des défis ouverts tels que : CHAOS [153], SegPC-2021 [161], DigestPath [162], pour la segmentation ou la détection de pathologie ; ou encore LiTS, MIDOG cités ci-dessus.

Ainsi, le travail présenté dans ce chapitre vise à repousser cette limitation importante qu'est l'absence de jeux de données publiques d'images multimodales appariées de la région abdominale, incluant des images échographiques transabdominales.

3.2 Le jeu de données TRUSTED

3.2.1 Acquisition des images

Les images échographiques 3D et les scans CT de ce jeu de données ont été collectés de manière prospective dans le cadre d'une étude monocentrique appelée "Tridimensional Renal Ultra Sound Tomodensitometrie Database" (TRUSTED) avec pour identifiant national ID-RCB : 2020-A01029-30/SI :20.05.01.539110. L'étude a été approuvée par le Comité de Protection des Personnes Sud-Méditerranée II du gouvernement français. Elle a porté sur 48 hommes et femmes âgés de 18 à 80 ans qui devaient subir un scanner abdominal ou abdomino-pelvien standard dans le cadre de leurs soins de routine au Nouvel Hôpital Civil (NHC) de Strasbourg, en France. Chaque participant s'est inscrit à l'étude avec un consentement éclairé écrit.

Les images tomodensitométriques ont été acquises par un appareil de type Siemens Somatom Force 384, en trois phases (ou temps) : un temps sans produit de contraste, une phase avec injection de contraste intra-veineux et une phase excrétrice tardive. Les dimensions des images CT sont de $512 \times 512 \times (586 \pm 219)$ voxels et $(0.81 \pm 0.08) \text{ mm} \times (0.81 \pm 0.08) \text{ mm} \times (1.05 \pm 0.28) \text{ mm}$ d'espacement entre les voxels. Immédiatement après l'acquisition CT, l'acquisition échographique trans-abdominale était réalisée, le patient étant dans la même position (tant que

possible) en décubitus dorsal et sur la table d'acquisition du scanner. Pour cela, deux sondes échographiques étaient utilisées consécutivement : une sonde transabdominale convexe 2D avec tracking électromagnétique et une sonde transabdominale 3D Seimens Acuson S3000 7CF1 HD mécanique. La sonde échographique 3D acquérait les volumes en 2 secondes environ avec une résolution de $760 \times 540 \times (739 \pm 55, 15)$ voxels avec un espacement isotrope des voxels de 0,3 mm. Les reins gauche et droit du patient ont été imagés avec la sonde US 3D transabdominale, avec 4 ± 2 acquisitions d'images par patient. Toutes les images ont été anonymisées en supprimant tout texte sensible du contenu des images et des métadonnées DICOM.

3.2.2 Annotation

Avant annotation, chaque image rénale échographique 3D a d'abord été inspectée par un radiographe expérimenté (16 ans). Pour chaque patient, les volumes gauche et droit où le rein était entièrement contenu dans le champ de vision 3D ont été conservés pour être annotés. À ce stade, 59 images US 3D correspondant aux reins gauche et/ou droit de 39 patients ont été sélectionnées. Toutes les autres ont été exclues du processus d'annotation, car les reins étaient hors du champ de vision. Les 59 images US 3D sélectionnées ont été annotées par deux radiographes indépendants ayant 10 et 11 ans d'expérience, à l'aide du logiciel 3D Slicer [163] (Version 4.11.20210226). Sept points de repère anatomiques ont été identifiés par les trois radiologistes impliqués dans le processus d'annotation, comme étant à la fois simples à localiser et suffisamment représentatifs du rein pour une tâche de recalage global (non-nécessairement précis). Pour leur marquage, chaque annotateur a procédé comme suit. Tout d'abord, une transformation spatiale a été appliquée pour centrer le rein, suivie d'une rotation 3D pour que les axes principaux du rein s'alignent sur les axes spatiaux. Ensuite, 7 points de repère anatomiques ont été marqués sur le plan longitudinal (Figure 3.3). Le premier point correspondait au centre du bassin rénal. Il a plus loin été utilisé pour évaluer l'erreur sur cible de recalage (TRE - pour "Target Registration Error" en anglais), grâce à une meilleure concordance inter-annotateurs pour sa localisation. Les 6 autres points se trouvaient à la surface du rein et ont été utilisés pour estimer un recalage global servant d'initialisation pour les approches d'affinement. Le rein a été segmenté manuellement suivant les orientations axiale, sagittale et coronale, et l'outil d'interpolation de volume de 3D Slicer a été utilisé pour produire de manière interactive la segmentation 3D du rein. L'orientation axiale a été suivie en premier lieu, avec des segmentations à 5 mm environ d'intervalle entre les coupes. Les autres orientations ont été utilisées

pour améliorer la qualité des segmentations, en particulier lorsque les frontières du rein suivant l'orientation axiale n'étaient pas claires. La segmentation autour du hile rénal est connue pour être imprécise [164]. Les annotateurs ont suivi une définition de sa limite similaire à celle dans [165], et ont exclu des contours du rein, les vaisseaux sanguins de la zone du hile rénal.

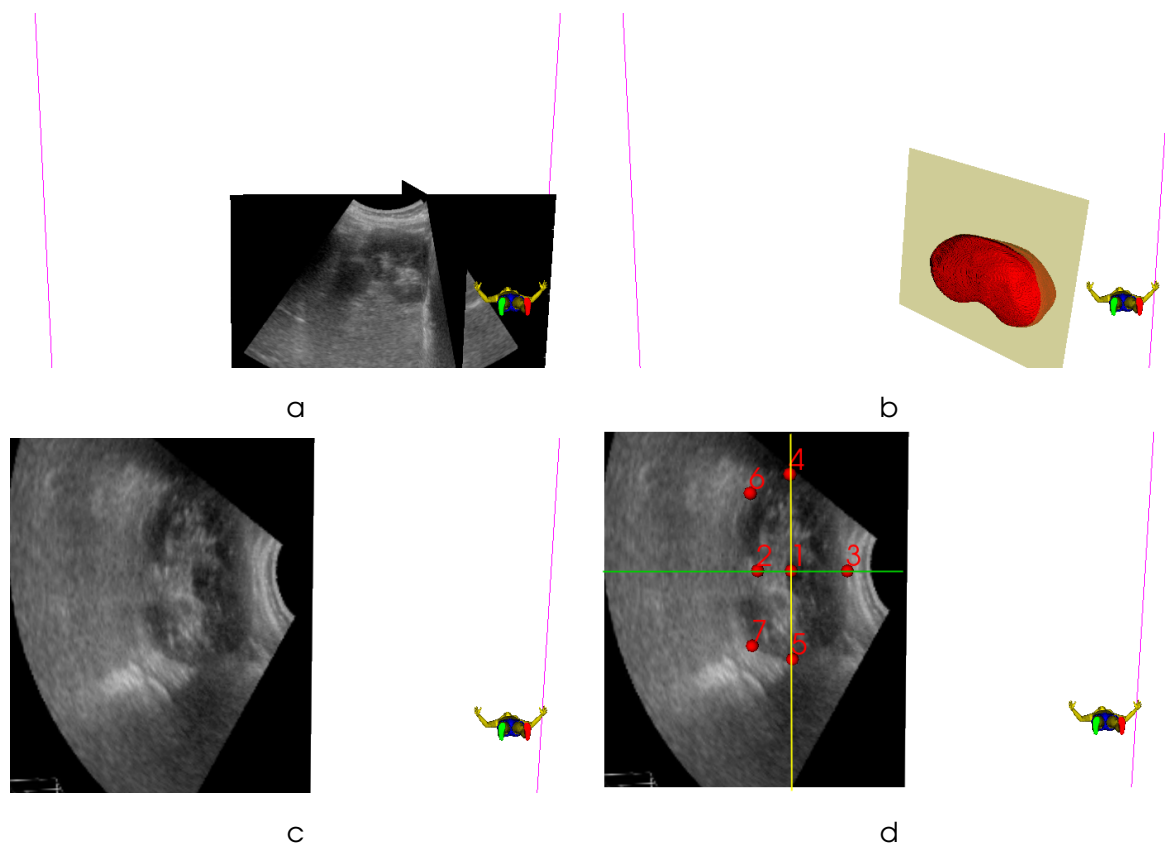


FIGURE 3.3 – Étapes du marquage des points de repère dans un volume échographique du jeu de données TRUSTED. (a) l'image US 3D dans sa pose initiale. (b) plan sur lequel les points sont marqués, par rapport au rein dans sa pose initiale dont la segmentation est présentée ici pour faciliter la visualisation. (c) le plan des points de repère dans l'image US 3D après centrage et rotation de celle-ci pour que les axes principaux du rein soient alignés avec les axes spatiaux. (d) les sept points de repère marqués.

Les images tomodensitométriques des 48 patients ont été annotées manuellement selon un processus similaire à celui des images échographiques. Seule la prise en compte de la présence des deux reins (gauche et droit) sur la même image était nécessaire, contrairement aux images US qui n'en contiennent qu'un. La Figure 3.4 illustre un cas d'annotations manuelles dans une image CT et dans une image US 3D.

Pour créer les segmentations de référence, l'algorithme STAPLE [166], largement utilisé en imagerie médicale, a permis de combiner automatiquement les

segmentations manuelles de reins fournies par les deux radiographes. Comme position de référence pour chacun des points de repère, la moyenne (milieu) des positions fournies par les deux radiographes a été utilisée. STAPLE possède un hyperparamètre important (β) qui régule le niveau de lissage spatial. Pour les images CT, ce paramètre n'était pas sensible, car sa variation avait très peu d'effet sur les segmentations fusionnées, de sorte que la valeur par défaut de $\beta_{CT} = 2.5$ (proposée dans [166]) a été utilisée. Cependant, en raison d'une concordance inter-annotateurs plus faible pour la segmentation des images US, la valeur de β avait une plus grande influence sur le résultat de fusion des segmentations. Il était donc nécessaire de l'ajuster. La valeur optimale de β_{US} a été estimée sur la base d'une hypothèse de déformation faiblement non linéaire du tissu rénal [167, 65]. Ce qui implique une faible variation de volume du rein d'une modalité à l'autre. Une recherche de valeur sur grille a été effectuée, où pour chaque $\beta_{US} \in \{1.0, 2.5, 3.0, 5.0, 10.0\}$, l'algorithme STAPLE était exécuté sur les données échographiques (image + deux segmentations), puis la différence relative de volume (en %) entre la sortie de STAPLE et la segmentation vérité de terrain (VT) dans l'image CT était calculée. La valeur optimale de β_{US} a été choisie comme étant celle conduisant à la concordance de volume la plus élevée en moyenne, soit $\beta_{US} = 2.5$.

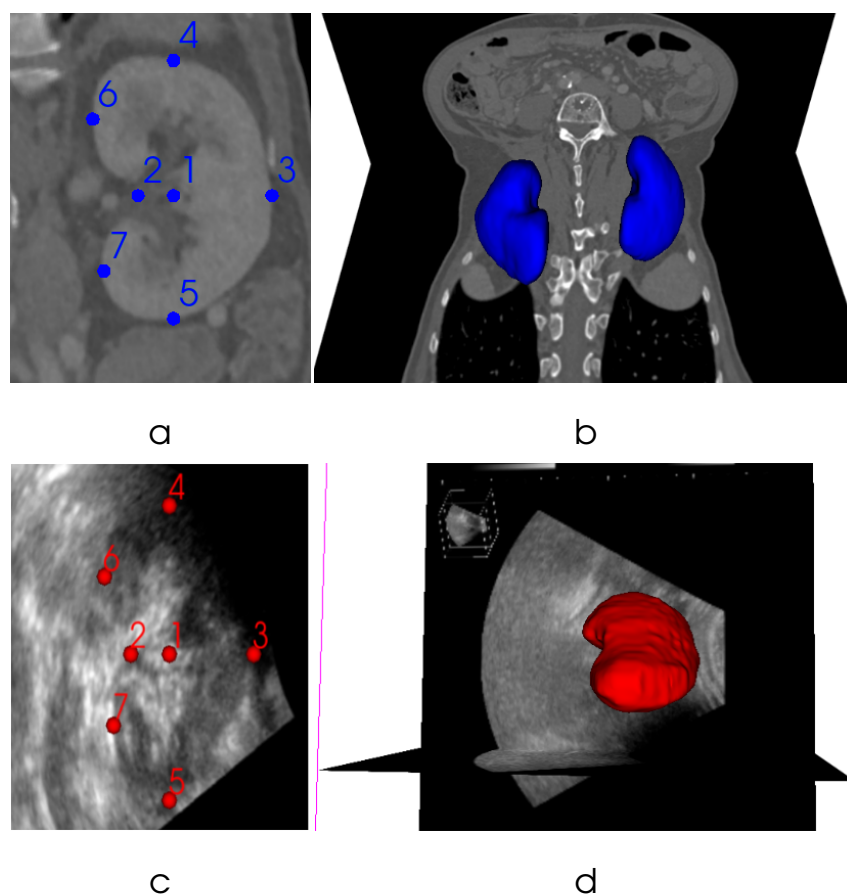


FIGURE 3.4 – Annotations des images dans le jeu de données TRUSTED. (a) et (b) points de repère coplanaires et segmentations manuelles (respectivement) des deux reins dans une image CT. (c) et (d) points de repère coplanaires et segmentations manuelles (respectivement) du rein dans une image US.

3.2.3 Disponibilité du jeu de données

Le jeu de données TRUSTED sera prochainement publié par le journal Nature - Scientific Data, et en libre accès.

Le répertoire contenant le jeu de données est structuré ainsi :

Le répertoire racine

Le répertoire racine contient deux sous-dossiers :

- **CT_DATA**, contient les images CT et leurs annotations.
- **US_DATA**, contient les images US et leurs annotations.
- **Initial_Transforms_from_noisy_landmarks** (optionnel), contient les transformations initiales calculées à partir des points de repère perturbés, utilisées pour l'analyse de la sensibilité à l'initialisation.

- **README.txt**, qui fournit une description de la structure du jeu de données, ainsi que les identifiants des patients anonymes affectés aux blocs de validation croisée (la liste nommée X_{cvi} donne, pour la modalité X : CT ou US, les identifiants des patients utilisés dans l'ensemble de test pour le i -ème bloc).
- **data_use_agreement.txt**, qui est la licence d'utilisation à remplir par toute personne demandant à télécharger cet ensemble de données.

Le sous-répertoire *CT_DATA*

Il contient huit sous-dossiers :

- *CT_images* : contient les images CT au format de fichier d'image Nifti [168]. Il y a 48 fichiers Nifti, chacun stockant une image CT d'un patient de dimension de $512 \times 512 \times 586 \pm 219$ voxels avec espacement de $0.81 \pm 0.08 \times 0.81 \pm 0.08 \times 1.05 \pm 0.28$ mm. Chaque fichier est une version anonymisée du fichier de données DICOM brut produit par le scanner CT, et par conséquent, l'information d'intensité de l'image n'est pas modifiée par rapport aux données brutes. Le nombre au début de chaque nom de fichier ici correspond à l'identifiant du patient.
- *CT_landmarks* : contient les annotations des points de repère anatomiques. Il contient trois sous-dossiers : *Annotator1* et *Annotator2* contiennent respectivement les coordonnées 3D des annotations de points de repère fournis par chaque annotateur. **GT_estimated_IdksCT** contient les coordonnées 3D des points de repère vérité terrain (VT), estimées à partir de celles produites par les deux annotateurs, tel que décrit dans la section 3.3. Dans ces sous-dossiers, il y a 96 ($48 \text{ individus} \times 2 \text{ reins par individu}$) fichiers .txt, chacun contenant les coordonnées 3D des 7 points de repère anatomiques, pour les reins gauche et droit du patient. Les fichiers texte sont nommés *XL1_IdkCT.txt* et *XR1_IdkCT.txt*, indiquant les coordonnées des points de repère pour les reins gauche et droit, pour le patient d'ID X , fournies par l'annotateur 1. Ces fichiers contiennent les coordonnées sur 7 lignes et 3 colonnes (délimitées par des espaces). Les lignes sont triées, de sorte que la i -ème ligne correspond aux coordonnées du point de repère numéro i , selon la Figure 3.4-a et c.
- *CT_landmarks_inimg*, les points de repère ci-dessus transformés de manière à appartenir au système de coordonnées CT
- *CT_tbackldk_transforms*, les transformations utilisées ci-dessus
- *CT_masks* : contient les annotations de segmentation sous forme de masques 3D binaires avec la même résolution spatiale que les fichiers d'image Nifti

dans *CT_images*. Il contient trois sous-dossiers : *Annotator1* et *Annotator2* contiennent respectivement les masques de segmentation fournis par l'annotateur 1 et par l'annotateur 2. *GT_estimated_masksCT* contient les masques de segmentation vérité terrain, estimées à l'aide de l'algorithme STAPLE, comme décrit dans la section 3.3. Chacun de ces 3 sous-dossiers contient 48 images Nifti binaires, chacune préfixée par l'ID du patient. Les images Nifti ont une dimension de $512 \times 512 \times 586 \pm 219$ voxels et un espacement de $0.81 \pm 0.08 \times 0.81 \pm 0.08 \times 1.05 \pm 0.28$ mm, correspondant exactement aux données d'image dans *CT_images*. Dans chaque fichier de masque, les valeurs de voxel 0 et 1 correspondent respectivement aux voxels étiquetés à l'extérieur et à l'intérieur du rein.

- *CT_meshes* : contient les maillages de surface, extraits automatiquement des masques de segmentation, en utilisant l'algorithme Marching Cubes [169], tel que décrit dans la section 3.3. Ce répertoire contient 3 sous-dossiers : *Annotator1*, *Annotator2* et *GT_estimated_meshesCT*, qui contiennent respectivement les maillages correspondant aux masques dans *CT_masks/Annotator1*, *CT_masks/Annotator2* et *CT_masks/GT_estimated_masksCT*. Les maillages sont stockés en utilisant le standard format .obj, dont la dénomination ressemble à celle utilisée pour les points de repère.
- *CT_meshes_inimg* : contient les maillages ci-dessus transformés pour appartenir au même système de coordonnées que l'image CT
- *CT_tbackmesh_transforms* : les transformations utilisées ci-dessus

Le sous-répertoire *US_DATA*

Ce sous-répertoire est organisé de la même manière que *CT_DATA*, mais ne contient que 5 sous-dossiers : *US_images*, *US_landmarks*, *US_masks*, *US_meshes* et *US_meshes_inimg*. Les conventions de dénomination et de format de fichier utilisées sont les mêmes que pour les données CT. Les données dans *US_meshes_inimg* sont toutes égales à la matrice diagonale $([-1, -1, 1, 1])$.

3.3 Analyse de la concordance

inter-annotateurs

Une analyse de la concordance inter-annotateurs sur les segmentations et les points de repère produits dans les deux modalités d'imagerie a été réalisée afin de mesurer la cohérence entre les annotateurs. Cette analyse quantitative des annotations une contribution scientifique importante, car elle permet entre

autres d'évaluer la difficulté qu'il peut y avoir pour les experts humains de se mettre d'accord sur ce qui serait un contour du rein ou la position d'un point, notamment dans les images échographiques 3D. Elle permet de poser le challenge déjà au niveau humain et introduit les opportunités d'amélioration que ce nouveau jeu de données apportera à travers le développement de nouvelles méthodologies. Pour cette analyse, les annotations de chaque radiographie (Ann.1 et Ann.2) ont été comparées aux annotations de référence estimées et considérées comme vérité terrain (voir la Section). Le Tableau 3.1 résume les résultats d'analyse des segmentations, en indiquant les valeurs moyennes et les écarts-types (entre crochets) de trois métriques d'évaluation : le score de Dice (DSC), la distance moyenne de surface du plus proche voisin (NN dist.) et la distance de Hausdorff à 95% (HD95) [170]. Les résultats indiquent une concordance inter-annotateurs légèrement plus élevée pour les segmentations dans la modalité CT par rapport à l'échographie, quoique la concordance pour les segmentations est globalement élevée dans les deux modalités. Le Tableau 3.2 résume les résultats d'analyse des points de repère, en indiquant pour chaque point de repère la distance moyenne entre la VT estimée et les positions annotées (écart-type entre crochets). Les positions des points de repère VT ont été calculées en moyennant les positions fournies par chaque annotateur. Par conséquent, les deux annotateurs ont la même distance par rapport à la VT pour les points de repère (comme indiqué dans le Tableau 3.2). En général, il y a une concordance plus élevée pour les points de repère dans la modalité CT, probablement en raison d'un meilleur contraste d'image, et une meilleure concordance pour le point de repère 1 (centre du bassinnet du rein) probablement en raison d'une meilleure visibilité anatomique à l'intérieur du rein. En ce qui concerne les images échographiques, la concordance est plus faible pour les points de repère 4 à 7 par rapport aux points de repère 1 à 3. Rappelons que ces points étaient situés sur un plan sagittal du rein, sélectionné par l'annotateur. Par conséquent, les déviations d'orientation par rapport à ce plan ont produit un effet de levier avec des différences de position plus importantes aux extrémités du rein.

Modalité	Manuelle Annotateur	DSC en %	NN dist. en mm	HD95 dist. en mm
CT	Ann.1	95.87 (0.78)	0.61 (0.15)	1.34 (0.50)
	Ann.2	95.68 (0.69)	0.62 (0.16)	1.35 (0.48)
US	Ann.1	93.45((2.09)	0.69 (0.39)	2.03 (1.31)
	Ann.2	93.97 (1.89)	0.67 (0.44)	1.92 (1.44)

TABLE 3.1 – Statistiques de concordance inter-annotateurs pour la segmentation des reins dans le jeu de données TRUSTED. Pour chaque modalité (CT et US), des segmentations de référence ont été calculées à partir des segmentations des annotateurs (2 annotateurs par image, une segmentation par annotateur) à l’aide de l’algorithme STAPLE. La concordance inter-annotateurs a été mesurée en comparant la segmentation de référence avec celle de chaque annotateur (DSC, NN dist. et HD95 dist.). La moyenne et l’écart-type (entre parenthèses) sont présentés dans le Tableau (chaque rein annoté dans l’ensemble de données correspond à un échantillon).

Modalité	LM1 dist.	LM2 dist.	LM3 dist.	LM4 dist.	LM5 dist.	LM6 dist.	LM7 dist.
CT	1.82 (1.65)	2.84 (1.81)	2.93 (2.40)	2.10 (1.27)	2.38 (1.75)	3.55 (2.64)	2.81 (1.88)
US	2.23 (1.64)	2.64 (1.98)	3.34 (2.50)	2.77 (2.03)	2.89 (2.15)	3.78 (2.55)	3.36 (2.73)

TABLE 3.2 – Statistiques de concordance inter-annotateurs pour les points repère rénaux dans le jeu de données TRUSTED. Pour chaque modalité (CT et US), les positions des points de repère 3D de référence ont été calculées en moyennant celles fournies par les deux annotateurs (2 annotateurs par image, 7 point de repère par annotateur). La concordance entre les annotateurs a été mesurée en comparant les positions des points de repère VT avec celles fournies par chaque annotateur (distance absolue en mm). La moyenne et l’écart-type (entre parenthèses) sont présentés dans le Tableau. Il convient de noter que, par construction, la distance absolue par rapport à la vérité terrain était la même pour chaque annotateur.

3.4 Validation du jeu de données pour la formation de modèles de segmentation automatique

Dans cette section, plusieurs modèles de réseaux de neurones profonds de référence pour la segmentation automatique d'images médicales sont entraînés et évalués à l'aide du jeu de données TRUSTED. L'objectif est de montrer l'utilité de l'ensemble de données TRUSTED pour l'entraînement et l'évaluation des modèles, mais aussi, d'exposer les limites de l'état de l'art dans les tâches cliniques de segmentation de rein dans les modalités US 3D et CT.

3.4.1 Les modèles et la méthodologie d'évaluation

Trois modèles/méthodes populaires utilisant les réseaux de neurones convolutifs (CNNs) ont été inclus dans cette comparaison : les modèles UNet [171], VNet [172], et la méthode nnU-Net [97] ; ainsi qu'un modèle de "Vision Transformer" (ViT) : CoTr [173]. Tous les modèles ont été formés et évalués dans le cadre d'une validation croisée à 5 blocs en veillant à ce que les données d'un même patient ne soient pas présentes dans différents blocs. Ainsi, les 39 patients (dont les 59 images de reins sont issues) ont d'abord été classés de manière aléatoire, puis ils ont été répartis de manière égale dans chaque bloc. La formation a été réalisée dans deux contextes : (i) cible unique, où les étiquettes d'entraînement correspondaient aux segmentations vérité de terrain obtenues en fusionnant les segmentations manuelles avec l'algorithme STAPLE, et (ii) cible double, où les deux segmentations manuelles fournies par les deux annotateurs ont été utilisées comme cibles d'entraînement à poids égal. La motivation de comparer ces deux approches d'entraînement était d'étudier si l'information sur l'incertitude des étiquettes, contenue implicitement dans les cibles doubles, contribuait à la performance des modèles.

Toutes les images d'entraînement ont été sous-échantillonnées en $128 \times 128 \times 128$ voxels afin de garantir que tous les modèles puissent être entraînés sous les deux versions dans un délai raisonnable. Pour sélectionner le mode d'interpolation pour ce sous-échantillonnage, l'erreur de quantification après sous-échantillonnage a été évaluée en utilisant le chevauchement spatial (score de Dice) entre les masques VT originaux et les masques VT soumis à un sous-échantillonnage en $128 \times 128 \times 128$ voxels, puis de nouveau à un sur-échantillonnage de $128 \times 128 \times 128$ vers leur taille d'origine. Parmi différents

modes d'interpolation : interpolation par zone ("area" en anglais), interpolation par plus proche voisin ("nearest" en anglais) et interpolation trilinéaire ("trilinear" en anglais), celui qui a donné les scores DICE moyens les plus élevés était le mode trilinéaire, avec 94.8 % et 95.7 % pour la tomodensitométrie et les ultrasons, respectivement. Le mode d'interpolation trilinéaire a donc été utilisé pour le sous-échantillonnage.

Comme cela a été montré dans [174, 175], la normalisation de l'intensité des images d'entrée améliore les performances des modèles neuronaux. Elle a été appliquée aux images US 3D et CT en tant qu'étape de prétraitement. L'approche utilisée consistait à soustraire la valeur moyenne du canal de chaque canal d'entrée, puis à diviser le résultat par l'écart type du canal, tel que cela est communément fait. L'amélioration des performances grâce à la normalisation de l'intensité a également été confirmée dans notre cas, puisque de moins bons résultats ont été obtenus en effectuant des essais d'entraînement/d'évaluation sans cette normalisation. Des techniques standard d'augmentation artificielle des données ont été utilisées pendant l'entraînement des modèles afin d'éviter un sur-apprentissage des données en quantité limitée par les modèles, conduisant à une mauvaise généralisation de ces derniers sur de nouveaux exemples. L'augmentation artificielle des données utilisait des transformations géométriques 3D aléatoires (retournement, rotation, translation, mise à l'échelle et distorsion affine) et des perturbations aléatoires de l'intensité des images (mise à l'échelle de l'intensité, décalage, bruitage gaussiens, lissage gaussien et variation de contraste). Les modèles nnU-Net et CoTr ont été entraînés avec les implémentations des auteurs : une fonction de perte combinée Dice-entropie croisée, avec régularisation par décroissance des poids au coefficient ("weight decay") 10^{-5} , une optimisation par la méthode de Descente du Gradient Stochastique, un taux d'apprentissage ou "learning rate" de 1% appliqué au gradient des poids à chaque itération, et un nombre maximal d'époques (ensemble d'itérations couvrant toutes les données d'entraînement) égal à 1000. Pour 3D UNet et VNet, le framework Monai [176] a été utilisé avec les paramètres d'apprentissage par défaut : une fonction de perte combinée Dice-entropie croisée, avec régularisation par décroissance des poids au coefficient 10^{-5} , une optimisation par la méthode de Descente du Gradient Stochastique, un taux d'apprentissage de 100% appliqué au gradient des poids à chaque itération, et un nombre maximal d'époques égal à 1000. Il est à noter que l'entraînement de 3D UNet et VNet avec une fonction de perte combinée Dice-entropie croisée et un taux d'apprentissage par itération de 10^{-1} ou moins, convergerait très lentement, d'où l'utilisation d'un taux d'apprentissage de 100%.

Après l'inférence, les masques de segmentation prédits avec la taille $128 \times$

128×128 voxels ont été post-traités avec une analyse des composantes connexes afin d'éliminer toutes les régions faussement positives et facilement détectables. Plus précisément, pour les images de tomodensitométrie, les régions étiquetées ayant les deux plus grandes composantes connexes par rapport à l'ensemble du volume ont été retenues comme régions des deux reins dans les images CT. Pour les images échographiques, où un seul rein est visible à la fois, la plus grande composante connexe a été retenue comme étant celle représentant la région du rein. Ensuite, les masques nettoyés ont été sur-échantillonnés par interpolation trilineaire de $128 \times 128 \times 128$ vers les dimensions des images d'origine. Enfin, les régions des reins estimées ont été comparées aux vérités terrain à l'aide de deux métriques bien établies : le score de Dice et la distance de surface moyenne des plus proches voisins (NN dist.). La distance des plus proches voisins a été calculée en utilisant les maillages de surface générés automatiquement par application de l'algorithme d'extraction de surface 3D "Marching Cubes" [169] sur les masques de segmentation.

3.4.2 Résultats quantitatifs de segmentation automatique

Le Tableau des résultats (Tableau 3.3) compare les performances des méthodes de segmentation. Les sections horizontales du haut et du bas montrent respectivement les résultats pour les modalités CT et US. Les trois métriques sont présentées par modèle et par modalité. Pour chaque configuration d'entraînement, il est indiqué : la performance moyenne, calculée sur l'ensemble des blocs de validation croisée, l'écart-type inter-blocs (entre parenthèses), l'intervalle de confiance (IC) bootstrap à 95 % de la moyenne (en utilisant les paramètres par défaut de l'implémentation `scipy.stats.bootstrap()`), et la p-valeur en comparaison avec la meilleure configuration en moyenne (en utilisant l'approximation normale du test de rang signé de Wilcoxon). En gras, la méthode automatique avec la meilleure performance moyenne pour chaque métrique.

En général, la méthode nnU-Net utilisant des cibles d'entraînement doubles a obtenu les meilleures performances en moyenne pour la segmentation du rein dans les images échographiques et de tomodensitométrie. La méthode la moins performante est 3D UNet dans les deux modalités. Les performances de segmentation dans la modalité échographique étaient généralement moins bonnes que celles dans les images de tomodensitométrie, avec une différence moyenne des scores de Dice sur tous les modèles, d'environ 9%, p-valeur = 1.17 % (en utilisant l'approximation normale du test de rang signé de Wilcoxon). Ces résultats illustrent bien que ces méthodes de segmentations automatiques, quoique

		TDM ou CT								
Méthode de segmentation	Cible d'entraînement	DSC ↑ en %	IC en % DSC	p-valeur DSC	NN dist.↓ en mm	IC en mm NN dist.	p-valeur NN dist.	HD95 dist.↓ en mm	IC en mm HD95 dist.	p-valeur HD95 dist.
3DUNet	Double	81.32 (3.16)	78.276-84.115	0.0431	2.29 (0.60)	1.820-2.892	0.0431	7.18 (2.26)	5.383-9.543	0.0431
3DUNet	Unique	79.63 (4.17)	75.334-82.676	0.0431	2.33 (0.57)	1.842-2.818	0.0431	7.05 (1.99)	5.424-8.963	0.0431
VNet	Double	85.66 (4.81)	81.371-89.588	0.0431	1.76 (0.77)	1.114-2.468	0.0431	4.76 (2.26)	2.918-7.019	0.0431
VNet	Unique	85.65 (3.96)	81.963-88.879	0.0431	1.71 (0.76)	1.089-2.412	0.0431	4.68 (2.22)	2.923-6.881	0.0431
nnU-Net	Double	90.09 (3.02)	87.041-92.427	-	1.27 (0.54)	0.910-2.001	-	3.69 (2.11)	2.312-6.564	-
nnU-Net	Unique	86.79 (6.65)	77.638-91.137	0.224	1.74 (1.42)	0.978-3.827	0.685	4.82 (3.97)	2.255-10.383	0.685
CoTr	Double	86.53 (3.51)	82.922-89.127	0.0431	1.52 (0.60)	1.138-2.348	0.0431	4.44 (2.31)	3.143-7.770	0.0431
CoTr	Unique	87.24 (2.70)	84.336-89.352	0.0431	1.39 (0.54)	1.002-2.090	0.0431	3.97 (2.07)	2.679-6.746	0.0431
Annotateur 1	-	95.87 (0.31)	95.591-96.142	0.0431	0.61 (0.04)	0.579-0.645	0.0431	1.34 (0.15)	1.223-1.483	0.0431
Annotateur 2	-	95.69 (0.23)	95.514-95.924	0.0431	0.62 (0.05)	0.563-0.655	0.0431	1.35 (0.12)	1.224-1.450	0.0431

		US								
Méthode de segmentation	Cible d'entraînement	DSC ↑ en %	IC en % DSC	p-valeur DSC	NN dist.↓ en mm	IC en mm NN dist.	p-valeur NN dist.	HD95 dist.↓ en mm	IC en mm HD95 dist.	p-valeur HD95 dist.
3DUNet	Double	72.24 (2.58)	70.204-74.747	0.0431	3.82 (0.33)	3.533-4.118	0.0431	11.96 (1.27)	10.867-13.035	0.0431
3DUNet	Unique	70.51 (23.02)	67.889-73.066	0.0431	3.98 (0.57)	3.542-4.587	0.0431	12.14 (1.69)	10.864-13.901	0.0431
VNet	Double	79.17 (3.02)	75.982-81.367	0.345	2.63 (0.61)	2.165-3.258	0.0431	8.26 (2.14)	6.793-10.437	0.0431
VNet	Unique	79.50 (4.19)	75.250-82.701	0.500	2.68 (0.58)	2.316-3.318	0.0431	8.62 (2.06)	7.350-10.841	0.0431
nnU-Net	Double	80.70 (4.83)	73.757-83.639	-	2.30 (0.51)	1.945-2.891	-	7.32 (1.77)	6.128-9.221	-
nnU-Net	Unique	79.37 (2.64)	76.575-81.370	0.500	2.35 (0.33)	2.106-2.718	0.500	7.69 (1.36)	6.687-9.120	0.224
CoTr	Double	77.79 (3.61)	72.534-79.882	0.0431	2.67 (0.51)	2.229-3.104	0.079	8.73 (1.61)	7.204-9.997	0.079
CoTr	Unique	75.63 (5.60)	67.680-78.992	0.0431	2.81 (0.76)	2.283-3.630	0.0431	8.85 (2.33)	7.053-11.333	0.0431
Annotateur 1	-	93.80 (0.58)	93.300-94.317	0.0431	0.68 (0.12)	0.595-0.797	0.0431	2.02 (0.37)	1.709-2.351	0.0431
Annotateur 2	-	94.19 (0.32)	93.868-94.444	0.0431	0.66 (0.11)	0.575-0.779	0.0431	1.91 (0.36)	1.616-2.255	0.0431

TABLE 3.3 – Évaluation quantitative des performances de segmentation sur des images CT et 3D US à l’aide du jeu de données TRUSTED. Le Tableau est divisé en deux sections : en haut - les résultats de segmentation du rein dans les images CT, en bas - les résultats de segmentation du rein dans les images 3D US. Les signes (↑ / ↓) indiquent si des valeurs plus élevées ou plus basses sont meilleures. Quatre méthodes automatiques ont été comparées (3DUNet, VNet, nnU-Net et CoTr). Les performances ont été évaluées sur les données de test en contexte de validation-croisée en comparant les masques vérité terrain avec les prédictions (métriques : Dice, distance NN et distance HD95). Les résultats en gras indiquent la méthode automatique avec la meilleure performance moyenne pour chaque métrique. Les performances de la segmentation manuelle sont présentées en bas de Tableau.

compétitives, performant moins bien avec des images échographiques 3D par rapport aux images de tomodensitométrie. Des recherches supplémentaires sont donc nécessaires pour combler l’écart, grâce à de meilleures architectures ou de meilleures configurations d’apprentissage. L’ensemble de données TRUSTED soutiendra ces efforts de recherche. En ce qui concerne l’utilisation de cibles d’entraînement uniques ou doubles, aucune preuve significative n’a été trouvée pour étayer l’hypothèse selon laquelle l’entraînement sur des cibles doubles intégrant l’information d’incertitude sur les annotations, améliore la

performance en termes de segmentation (La p-valeur du score de Dice pour la configuration "nnU-Net Unique" comparée à "nnU-Net Double" est de 22.4 % pour les images CT et de 50 % pour les images US). Le Tableau 3.3 montre également les performances quantitatives des segmentations manuelles de chaque annotateur humain par rapport à la vérité terrain. Elles sont significativement supérieures (p-valeurs inférieures à 5%) à la meilleure méthode automatique (nnU-Net). Cela indique que, avec l'ensemble de données TRUSTED, il existe toujours un écart important entre la segmentation humaine et la segmentation automatique de pointe. Toutefois, il faut reconnaître que les masques vérité terrain étant dérivés des annotations humaines, la performance au niveau humain peut être artificiellement favorisée.

3.4.3 Résultats qualitatifs de segmentation automatique

La Figure 3.5 montre, pour deux exemples de patients (418 et 680), les résultats de segmentation des reins dans les images de tomodensitométrie, avant le post-traitement par analyse des composantes connexes mentionné dans la section 3.4. Chaque ligne de la Figure correspond à une méthode de segmentation, les deux premières correspondent aux segmentations manuelles effectuées par l'un ou l'autre des annotateurs (Ann.1 et Ann.2), et les autres correspondent aux segmentations par les méthodes automatiques à base de réseaux de neurones profonds. Deux types de visualisations sont représentés. Une coupe coronale 2D dans les images CT avec les contours des reins segmentés et une vue des masques volumiques 3D de segmentation. La visualisation en coupe 2D permet de mettre en évidence les détails de segmentation à l'échelle de cette coupe, alors que la visualisation des masques 3D donne suivant un angle de vue, un aperçu de l'information spatiale 3D des segmentations. Les segmentations automatiques sont représentées en vert, tandis que les segmentations de référence (vérité terrain) dans les images CT sont représentées en bleu. Pour les méthodes automatiques, les résultats de la meilleure configuration d'entraînement (en termes de score de Dice moyen sur les blocs de validation) entre cibles uniques ou cibles doubles, sont présentés. Les bonnes performances des méthodes automatiques pour la segmentation dans les images de tomodensitométrie, telles qu'indiquées dans le Tableau 3.3, se reflètent sur la Figure 3.5. Ce qui se traduit aussi par une faible différenciation entre les masques volumiques vérité terrain (en bleu) et automatique (en vert) dans les vues en 3D. Toutefois, quelques erreurs locales de segmentation automatique non négligeables sont visibles sur les vues 3D : (3D UNet, TDM_418), (3D UNet, TDM_680) et (VNet, TDM_680).

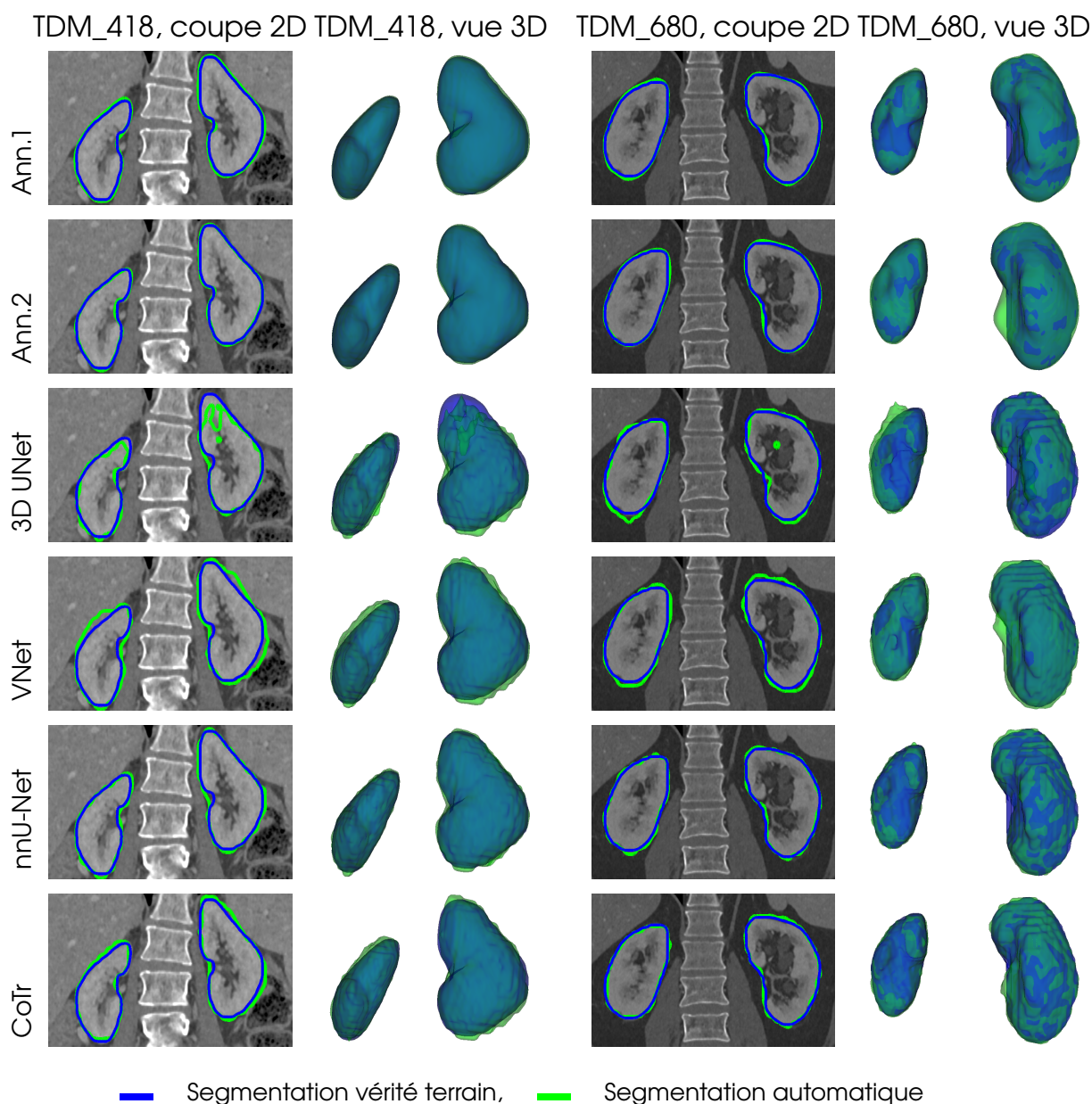


FIGURE 3.5 – Résultats qualitatifs montrant deux exemples de segmentations dans les images CT du jeu de données TRUSTED (patients 418 et 680) avant post-traitement par analyse des composantes connexes. Les première et troisième colonnes montrent des coupes coronales 2D dans les images TDM (ou CT), avec les segmentations automatiques en vert et les segmentations vérité terrain superposées en bleu. Les deuxième et quatrième colonnes montrent des vues des masques 3D de segmentation automatique en vert avant le post-traitement éliminant les faux positifs grossiers, et les masques de segmentation vérité terrain en bleu. Les deux premières lignes (Ann.1 et Ann.2) montrent les segmentations de chaque annotateur, et les lignes suivantes montrent pour chaque méthode de segmentation automatique, le résultat avec la meilleure configuration d’entraînement en moyenne entre cible unique et cibles doubles.

La Figure 3.6 montre, pour deux exemples de patients (418R : patient 418, rein droit, et 680L : patient 680, rein gauche), les résultats de segmentation des reins dans les images échographiques, avant le post-traitement par analyse des composantes connexes mentionné dans la section 3.4. Les lignes correspondent aux méthodes de segmentation, telles que dans le paragraphe précédent des résultats qualitatifs dans les images TDM (ou CT). Deux types de visualisations sont également représentés. Une coupe coronale 2D dans les images US 3D avec les contours de rein segmentés et une vue des masques volumiques 3D de segmentation. La visualisation en coupe 2D donnant des détails de segmentation à l'échelle de cette coupe et la vue 3D une information spatiale plus globale. Les segmentations automatiques sont représentées en vert et les segmentations de référence (vérité terrain) dans les images US 3D en rouge. Pour les méthodes automatiques, les résultats de la meilleure configuration d'entraînement (cible unique versus cibles doubles) sont présentés. Les performances des méthodes automatiques pour la segmentation dans les images échographiques 3D sont plus variées, tel qu'indiqué dans le Tableau 3.3. En général, elles sont moins bonnes dans les zones d'ombre où le contraste de l'image (nécessaire pour délimiter précisément l'organe) est beaucoup plus faible : (3D UNet, US_418R), (VNet, US_418R), (3D UNet, US_680L), (VNet, US_680L), (nnU-Net, US_680L) et (CoTr, US_680L). Dans le cas du rein US_418R, la meilleure performance de nnU-Net est observée, suivie de CoTr malgré une région faussement positive qui sera écartée grâce au post-traitement par analyse des composantes connexes (voir section 3.4). Dans ce cas, US_418R, 3D-Unet et VNet sont moins robustes au faible contraste dans la zone du pôle supérieur du rein et dans la zone du hile rénal dont la difficulté de délimitation précise a déjà été évoquée dans la section 4.10. Dans le cas du rein US_680L, toutes les méthodes automatiques sous-segmentent, d'une certaine façon, la zone à droite qui est faiblement contrastée.

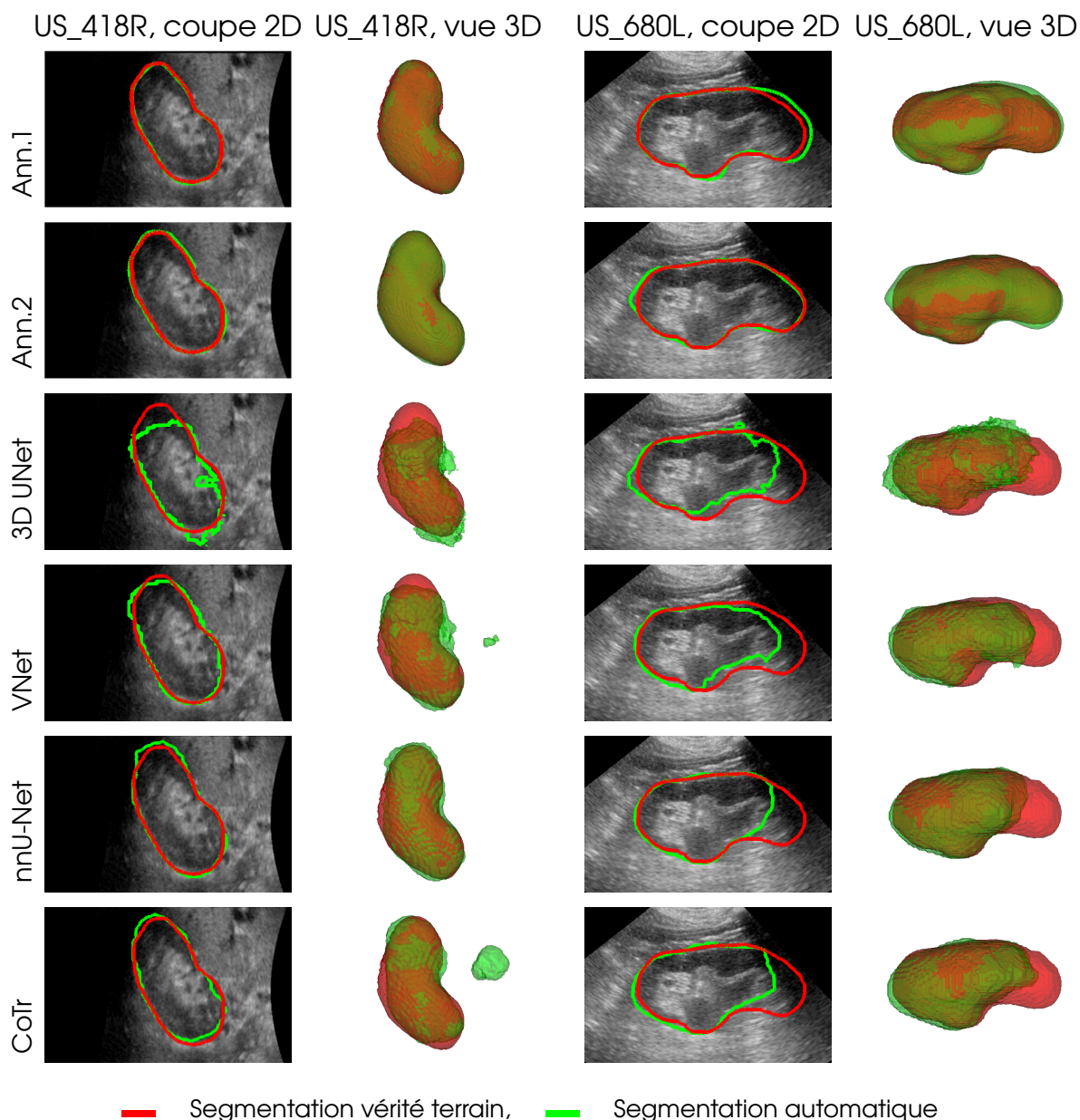


FIGURE 3.6 – Résultats qualitatifs montrant deux exemples de segmentations dans les images US 3D du jeu de données TRUSTED (418R : patient 418, rein droit, et 680L : patient 680, rein gauche) avant post-traitement par analyse de composantes connexes. Les première et troisième colonnes montrent des coupes coronales 2D dans les images US 3D, avec les segmentations automatiques en vert et les segmentations vérité terrain superposées en rouge. Les deuxième et quatrième colonnes montrent des vues des masques 3D de segmentation automatique en vert avant le post-traitement éliminant les faux positifs grossiers, et les masques de segmentation vérité terrain en rouge. Les deux premières lignes (Ann.1 et Ann.2) montrent les segmentations de chaque annotateur, et les lignes suivantes montrent pour chaque méthode de segmentation automatique, le résultat avec la meilleure configuration d’entraînement en moyenne entre cible unique et cibles doubles.

3.5 Validation du jeu de données pour la formation de modèles de recalage

Dans cette section, quatre méthodes concurrentes de recalage inter-modal d'images sont comparées à l'aide du jeu de données TRUSTED. Les méthodes de recalage les plus précises utilisent un processus d'affinement (affinement du recalage) après qu'un recalage initial approximatif ait été fourni (initialisation du recalage). Les méthodes dont il est question ici sont utilisées pour affiner le recalage.

3.5.1 Méthodologie d'évaluation

La première méthode est l'algorithme *Iterative Closest Point* (ICP) [115], qui malgré son âge, reste une méthode compétitive en termes de précision et de vitesse d'exécution [177]. La méthode estime la transformation de recalage en utilisant les nuages de points issus de segmentation de l'organe dans les images. La seconde méthode est *Bayesian Coherent Point Drift* (BCPD) [117], qui améliore la méthode populaire *Coherent Point Drift* (CPD) en termes de vitesse d'optimisation et de bassin de convergence plus large. Les méthodes ICP et BCPD ont besoin des surfaces segmentées de l'organe. Dans ces expérimentations, les segmentations ont été effectuées par la méthode nnU-Net entraînée avec des cibles doubles, sur la base de ses bonnes performances dans les deux modalités d'imagerie, tel que décrit dans la section 3.4. Puis les surfaces ont été générées à partir des segmentations par la version Lewiner de l'algorithme *Marching Cubes* [169]. La troisième méthode de recalage provient de l'outil *ImFusion Suite 2.45.3* (ImFusion GmbH, Munich, Allemagne), qui offre une méthode d'optimisation itérative de pointe, spécialement conçue pour le recalage inter-modal d'images échographiques [178, 38]. Il s'agit d'une approche de recalage iconique. Deux variantes ont été comparées ici à l'aide de différentes mesures de similarité entre les images : *Local Normalized Cross Correlation* (iLNCC) [123] et *Linear Correlation of Linear Combination* (iLC2) [65]. La quatrième méthode est le modèle de recalage d'images médicales *TransMorph* (TrMph) [76]. Il s'agit d'une adaptation récente du réseau à convolution *VoxelMorph* [75], mais qui utilise les "Vision Transformers" (ViTs). Les modèles VoxelMorph et TransMorph (avec des images en entrée et des déformations en sortie) ne nécessitent aucun champ de déformation fourni pendant l'entraînement ; ils génèrent plutôt eux-mêmes de pseudo-étiquettes (champs de déformation). Pour cela ils sont souvent qualifiés de non-supervisés ou encore d'auto-supervisés

[75].

Les méthodes de recalage nécessitent un modèle hypothétique de la transformation spatiale à estimer (rigide, affine, non-rigide). Sur la base de travaux antérieurs qui montrent que le mouvement physiologique du rein n'induit pas de déformation non linéaire significative [167, 179, 65], deux modèles de transformation ont été comparés : mouvement rigide avec 6 degrés de liberté (DDL), et déformation affine, avec 12 DDL. L'expérimentation de la méthode ICP a été conduite avec son implémentation dans le framework VTK [180]. Les méthodes iconiques iLNCC et iLC2 ont été expérimentées via l'outil *ImFusion Suite 2.45.3*, avec les paramètres par défaut. Pour expérimenter TransMorph sous hypothèse de transformation affine, la version originale du modèle TransMorph_affine proposée par les auteurs a été entraînée et évaluée (https://github.com/junyuchen245/TransMorph_Transformer_for_Medical_Image_Registration), en utilisant les mêmes blocs de validation-croisée que les modèles de segmentation automatique (voir Section 3.4.2). Les images CT et US 3D appariées et leurs segmentations vérité terrain ont été utilisées pendant l'entraînement, et la perte de Dice a été minimisée comme cela a été mentionné dans l'article original [76]. Pour évaluer le modèle sur les ensembles de test, seules les images ont été utilisées pour inférer la transformation de recalage. Pour expérimenter TransMorph sous hypothèse de transformation rigide, les dernières couches de l'architecture TransMorph_affine, qui génèrent les paramètres de mise à l'échelle et de cisaillement, ont été simplement supprimées. Pour expérimenter BCPD, les valeurs suivantes ont été utilisées pour les paramètres, en accord avec leurs noms dans le dépôt GitHub des auteurs (<https://github.com/ohirose/bcpd>) :

- " -w0.1 -l1e9 -g0.1 -ux -K70 -J300 -n50 -N30 -p -d7 -e0.3 -f0.3 -o " , en contexte de recalage rigide
- " -w0.1 -l1e4 -g0.1 -ux -K70 -J300 -n50 -N30 -p -d7 -e0.3 -f0.3 -o " , en contexte de recalage affine

Initialisation du recalage

Pour une comparaison équitable, toutes les méthodes de recalage présentées ici ont été initialisées avec la même approche de recalage global. Elle utilisait les points de repère vérité terrain estimés : #3 : centre-latéral, #4 : pôle supérieur et #5 : pôle inférieur (voir la Figure 3.4). Puis, calculait à partir d'eux, la transformation de similarité 3D la mieux adaptée au sens des moindres carrés. Il s'agit d'une initialisation "manuelle" du recalage (points de repère vérité terrain). L'usage d'une initialisation manuelle du recalage est encore la pratique courante en contexte clinique. Dans le chapitre 4, une approche automatique pour le recalage global du rein dans les images US 3D et CT est proposée.

La sensibilité des méthodes d’affinement à la précision du recalage initial a été évaluée à l’aide d’une expérience semi-synthétique. Dans cette expérience, les positions des points de repère dans le système de coordonnées de l’image US 3D ont été perturbées de manière aléatoire par des quantités variables de bruit blanc gaussien, afin de simuler une imprécision variable du recalage initial. Cette expérience est présentée dans la section 3.5.2.

Métriques d’évaluation

Quatre métriques complémentaires ont été utilisées pour évaluer la précision des méthodes de recalage :

- **Target Registration Error** (TRE) ou erreur sur cible de recalage, en mm. Le centre du bassin rénal (point de repère #1) vérité terrain a été identifié comme cible fiable, pour évaluer le recalage à l’aide du jeu de données TRUSTED. En effet, parmi les sept points de repère localisés dans les deux modalités, il est l’unique situé à l’intérieur de l’organe. Il est donc un bon candidat pour rendre compte des déformations internes du rein. De plus, il a été localisé par les annotateurs avec la meilleure concordance dans les deux modalités (distance moyenne par rapport à la vérité terrain de 1.82 mm dans les images CT et 2.23 mm sur les images US 3D). Enfin, il n’a été utilisé dans aucune des étapes et par aucune des méthodes de recalage dans ces expérimentations. Ainsi, la TRE évalue la distance entre la position cible 3D du point de repère #1 (dans l’image CT) et sa position 3D prédite en transformant l’image US 3D selon le recalage estimé.
- **Score de Dice** (DSC), en %. Il évalue le chevauchement spatial entre la segmentation vérité terrain dans l’image CT et celle dans l’image US 3D transformée selon le recalage estimé.
- **Distance de surface moyenne des plus proches voisins** (NN dist.), en mm. Elle évalue la distance moyenne des plus proches voisins entre la surface de la segmentation vérité terrain dans l’image CT et celle dans l’image US 3D transformée selon le recalage estimé.
- **Distance de Hausdorff 95** (HD95). Elle évalue le 95e centile de la distance de Hausdorff entre la surface de la segmentation vérité terrain dans l’image CT et celle dans l’image US 3D transformée selon le recalage estimé.

3.5.2 Résultats quantitatifs de recalage

Le Tableau 3.4 montre les performances quantitatives des méthodes d’affinement du recalage. Le Tableau est divisé en deux blocs principaux : Bloc supérieur

- résultats des méthodes sous hypothèse de transformation rigide (6 DoF), Bloc inférieur - résultats des méthodes sous hypothèse de transformation affine (12 DoF). Toutes les méthodes d’affinement ont été initialisées avec les mêmes transformations spatiales (transformée de similarité calculée à partir des points de repère annotés manuellement # 3, # 4 et # 5). Les quatre mesures sont présentées pour chaque méthode. Le Tableau montre, pour chaque configuration, la performance moyenne sur l’ensemble des blocs, l’écart-type entre les blocs (entre parenthèses), l’intervalle de confiance bootstrap à 95 % de la moyenne (en utilisant les paramètres par défaut de l’implémentation `scipy.stats.bootstrap()`), et la p-valeur en comparaison avec la meilleure configuration en moyenne (en utilisant l’approximation normale du test de rang signé de Wilcoxon). Pour chaque métrique, les éléments en gras montrent la méthode ayant obtenu les meilleures performances en moyenne (à l’exclusion des méthodes utilisant les segmentations vérité terrain).

Dans les sections horizontales supérieures des deux blocs du tableau (montrant la performance des méthodes qui peuvent être utilisées en pratique), BCPD sous hypothèse de transformation affine montre le meilleur équilibre global de précision à travers les métriques, avec score de Dice moyen de 84.10%, NN dist de 2.17 mm, HD95 de 7.32 mm et TRE de 4.47 mm. Il est suivi par TransMorph sous hypothèse de transformation affine, avec des moyennes en score de Dice de 77.67 %, en NN dist. de 2.38 mm, en HD95 de 6.34 mm et en TRE de 3.73 mm. En général, une amélioration relativement faible est observée entre la version rigide et la version affine pour toutes les méthodes. Elle peut s’expliquer en partie par le fait que le recalage initial utilise une transformée de similarité, qui prend déjà compte certains effets non rigides (la mise à l’échelle isotrope). TransMorph montre les meilleures performances en termes de TRE (<3.80 mm) et HD95 (6.34 mm), mais pas en termes de score de Dice (rigide : 76.64%, affine : 77.67%) et de NN dist (rigide : 2.42 mm, affine : 2.38 mm). Il n’y a pas d’explication expérimentale, mais il semble que la version affine (et rigide) de TransMorph capture bien les relations globales dans les données, conduisant à un meilleur alignement global des structures anatomiques, ce à quoi les métriques TRE et HD95 sont sensibles. Mais cette version de TransMorph a plus de difficulté avec l’alignement précis au niveau du voxel, ce qui entraîne un moins bon alignement local, auquel les métriques score de Dice et NN dist sont plus sensibles. Les méthodes ICP et BCPD obtiennent de meilleurs résultats que les méthodes iconiques iLNCC et iLC2 qui, elles, ont des performances moyennes inférieures pour toutes les mesures. Ce qui illustre deux potentiels aspects. Premièrement, ces méthodes iconiques n’ont pas été en mesure d’exploiter les informations de recalage supplémentaires potentielles contenues dans les valeurs d’intensité de

voxel. Deuxièmement, elles ont eu du mal à converger vers la solution correcte, par rapport aux méthodes utilisant les surfaces.

Si l'on considère la section inférieure du tableau 3.4 montrant les performances des méthodes de recalage utilisant les surfaces vérité terrain, BCPD sous hypothèse de transformation affine a fourni les meilleurs résultats pour toutes les métriques, et en outre les performances ont dépassé celles de BCPD utilisant les segmentations automatiques de nnU-Net : La TRE moyenne était de 4.10 mm et le score moyen de Dice de 87.10%. Les résultats sont conformes à ceux de [181], qui a évalué les performances de l'ICP avec surfaces issues de segmentation manuelle sur un ensemble de données d'image fantôme et obtient TRE moyenne de 4.7 mm. L'amélioration des résultats obtenue sur cet ensemble de données en utilisant une segmentation manuelle justifie la nécessité de développer des méthodes de segmentation automatique plus performantes, en vue d'atteindre ou de dépasser les performances de recalage avec segmentation manuelle. Par conséquent, le jeu de données TRUSTED peut être utilisé pour évaluer et améliorer les méthodes de segmentation automatique en tenant compte d'une application concrète en aval : le recalage.

Hypothèse de transformation rigide													
Méthode d'affinement	Méthode de segmentation	TRE ↓ en mm	IC en mm	p-valeur TRE	Dice score ↑ en %	IC en %	p-valeur Dice score	NN dist. en mm	IC en mm	p-valeur NN dist.	HD95 dist. ↓ en mm	IC en mm	p-valeur HD95 dist.
ICP (S)	nnU-Net	5.30 (0.44)	4.938-5.696	0.0431	80.15 (1.44)	78.964-81.563	0.0431	2.45 (0.24)	2.318-2.799	0.0431	8.58 (1.07)	7.945-10.119	0.0431
BCPD (S)	nnU-Net	4.87 (0.49)	4.348-5.231	0.0431	81.86 (2.46)	79.940-84.487	0.0431	2.33 (0.22)	2.195-2.660	0.138	7.92 (1.20)	7.097-9.169	0.0431
ILNCC (I)	/	7.48 (2.04)	5.799-9.286	0.0431	72.74 (4.68)	68.492-76.464	0.0431	2.95 (0.35)	2.711-3.321	0.0431	12.70 (3.11)	10.190-15.463	0.0431
iLC2 (I)	/	6.42 (1.78)	4.965-7.935	0.0431	74.46 (4.37)	70.482-78.241	0.0431	2.65 (0.24)	2.479-2.920	0.0431	10.83 (1.63)	9.209-12.155	0.0431
TransMorph	/	3.24 (0.31)	2.984-3.515	-	76.64 (2.09)	74.786-78.346	0.0431	2.42 (0.26)	2.253-2.744	0.0431	6.34 (0.80)	5.802-7.237	0.8927
ICP (S)	STAPLE-VT	4.56 (0.57)	4.183-5.134	0.0431	81.98 (0.95)	81.218-82.915	0.0431	2.19 (0.21)	2.053-2.494	0.8927	5.63 (0.69)	5.161-6.382	0.0431
BCPD (S)	STAPLE-VT	4.66 (0.51)	4.125-4.990	0.0431	82.04 (0.86)	81.337-82.803	0.0431	2.20 (0.21)	2.061-2.495	0.6858	5.68 (0.66)	5.226-6.452	0.0431
Hypothèse de transformation affine													
Méthode d'affinement	Méthode de segmentation	TRE ↓ en mm	IC en mm	p-valeur TRE	DSC ↑ en %	IC en %	p-valeur DSC	NN dist. en mm	IC en mm	p-valeur NN dist.	HD95 dist. ↓ en mm	IC en mm	p-valeur HD95 dist.
ICP (S)	nnU-Net	5.34 (0.59)	4.901-5.958	0.0431	80.60 (1.27)	79.638-81.991	0.0431	2.49 (0.25)	2.340-2.843	0.0431	8.53 (1.26)	7.726-10.192	0.0431
BCPD (S)	nnU-Net	4.47 (0.62)	3.830-4.943	0.0431	84.10 (1.85)	82.899-86.418	-	2.17 (0.40)	1.896-2.602	-	7.32 (1.28)	6.272-8.547	0.0431
ILNCC (I)	/	6.32 (0.66)	5.870-7.125	0.0431	74.78 (1.29)	73.505-75.843	0.0431	3.10 (0.45)	2.680-3.480	0.0431	11.24 (1.09)	10.358-12.268	0.0431
iLC2 (I)	/	6.13 (1.42)	5.050-7.686	0.0431	75.80 (3.20)	73.600-79.626	0.0431	2.81 (0.36)	2.487-3.104	0.0431	10.79 (1.51)	9.061-11.840	0.0431
TransMorph	/	3.73 (0.30)	3.475-4.003	0.0431	77.67 (1.63)	76.154-79.074	0.0431	2.38 (0.26)	2.204-2.670	0.138	6.34 (0.79)	5.765-7.142	-
ICP (S)	STAPLE-VT	4.78 (0.64)	4.256-5.034	0.0431	86.88 (1.51)	85.186-87.965	0.0431	1.41 (0.19)	1.292-1.628	0.0431	3.92 (0.67)	3.489-4.795	0.0431
BCPD (S)	STAPLE-VT	4.10 (0.43)	3.667-4.404	0.0431	87.10 (1.14)	85.903-87.993	0.0431	1.50 (0.20)	1.370-1.724	0.0431	4.23 (0.54)	3.853-4.859	0.0431

TABLE 3.4 – Résultats quantitatifs sur le jeu de données TRUSTED, des méthodes d'affinement du recalage. Les signes (↑ / ↓) indiquent si des valeurs plus élevées ou plus basses sont meilleures. Pour les approches de recalage utilisant les surfaces, indiquées par '(S)', la méthode de génération desdites surfaces dans les modalités CT et US est indiquée dans la colonne 2.

Analyse de la robustesse à une initialisation imprécise du recalage

Dans le contexte clinique, il n'est pas souhaitable d'avoir une méthode d'affinement nécessitant une initialisation précise, parce qu'une initialisation précise en contexte clinique est non seulement coûteuse en temps, mais aussi difficile à obtenir. L'expérience décrite dans cette section analyse, à l'aide du jeu de données TRUSTED, la robustesse des méthodes d'affinement à une initialisation imprécise. Le processus se déroule comme suit : pour chaque paire de reins CT /US 3D, des bruits blancs aléatoires d'écart-types variés ont été appliqués aux trois points de repère du rein US avant leur utilisation pour l'initialisation du recalage. Les écart-types σ (ou std - pour "standard deviation" en anglais) allaient de 2 mm (faible bruit) à 10 mm (bruit important) par incrément de 2 mm. Pour chaque σ , le recalage a été initialisé à l'aide de la transformation de similarité optimale suivant les points de repère bruités, puis chaque méthode d'affinement du recalage a été exécutée. Pour générer différents échantillons de bruit, le processus a été répété 5 fois pour chaque valeur d'écart-type. Au total, 11 800 recalages ont été effectués : 4 méthodes de recalage, 2 hypothèses de transformation spatiale, 5 valeurs d'écart-types, 5 échantillons de bruit et 59 paires d'images). A des fins de comparaison, les résultats avec les points de repère non bruités ont été rapportés comme correspondant à $\sigma = 0$ mm ou std = 0 mm.

La Figure 3.7 montre pour les cas 418R et 680L les positions relatives des reins dans les images CT et US 3D après recalage initial par les points de repère bruités. Rappelons que pour chaque valeur de l'écart-type σ , 5 échantillons aléatoires de recalage initial ont été simulés. La Figure n'en présente qu'un par écart-type.

Les résultats de cette expérience sont présentés par métrique par la Figure 3.8. Les valeurs de l'écart-type σ sont indiquées par *std* ("standard deviation" en anglais). Pour chaque boîte à moustaches, les barres inférieure, centrale et supérieure indiquent les 1er, 2e (médiane) et 3e quartiles, les cercles blancs indiquent la moyenne et les moustaches indiquent les valeurs extrêmes non considérées comme aberrantes. Les points indiquent les valeurs aberrantes, et les nombres au-dessus de chaque boîte à moustaches indiquent le nombre de valeurs aberrantes qui se situent au-delà de l'intervalle de l'axe des ordonnées. De manière générale, les performances d'affinement de recalage se dégradent avec l'augmentation de σ (std), bruit introduit dans le recalage initial.

Lorsque $\sigma = 0$ (std=0), les résultats sont identiques à ceux présentés dans le Tableau 3.4, mais présentés sous forme de boîte à moustaches. Les scores médians de Dice sont tous supérieurs à 75 %, les TRE, NN dist et HD95 médianes sont inférieures à 5 mm, 3 mm et 10 mm respectivement. Toutes les méthodes ont de fortes queues, indiquant une minorité non négligeable de paires d'images

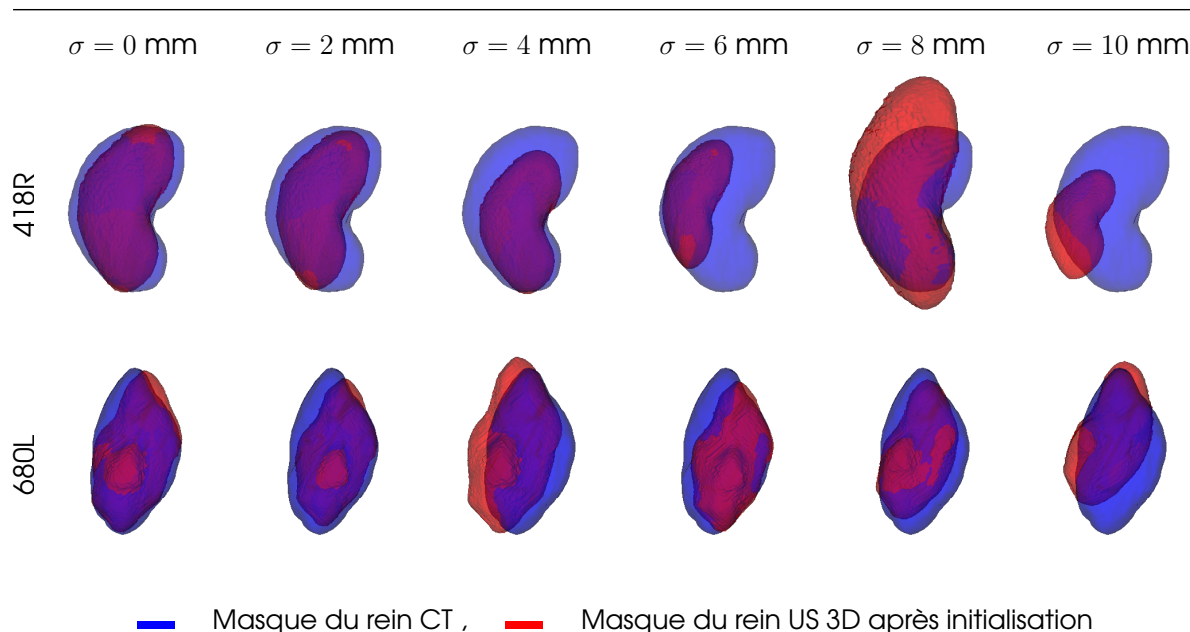


FIGURE 3.7 – Positions relatives des reins dans les images CT et US 3D après recalage initial perturbé par bruit blanc avec différentes valeurs d'écart-type σ en mm. Pour chaque écart-type, cinq valeurs de bruit sont échantillonnées, mais une seule est observée.

ayant produit des erreurs élevées. Les 3e quartiles (*i.e* les bords supérieurs de la boîte) montrent qu'au moins 25% des cas ont des valeurs de TRE supérieures à 5 mm ou un score de Dice inférieur à 85%, pour toutes les méthodes. Ce qui indique que, même lorsque les méthodes de recalage bénéficient d'une initialisation relativement bonne (en utilisant des points de repère manuels), le jeu de données TRUSTED pose un défi méthodologique pour réduire la proportion de cas présentant des erreurs de recalage élevées.

Lorsque le bruit d'initialisation du recalage augmente, les performances des méthodes iconiques (iLNCC et iLC2) et TransMorph se dégradent plus que celles des méthodes utilisant les surfaces (ICP, BCPD). Pour BCPD, le modèle de transformation affine produit de meilleurs résultats pour le DSC, la NN dist. et la HD95 dist., par rapport au modèle de transformation rigide. La différence de résultats en modèles affine et rigide est plus faible si l'on considère la TRE. Cela peut s'expliquer par le fait que l'emplacement du point de repère utilisé pour la TRE (centre du bassin rénal) est souvent proche du centroïde du rein et par conséquent, le recalage de ce point avec les modèles rigide et affine produit généralement des déplacements similaires.

A mesure que le bruit augmente, toutes les méthodes présentent une proportion accrue de paires d'images mal recalées, ce qui montre clairement l'importance de la précision de l'initialisation sur le recalage final. Les mêmes tendances sont

visibles pour les 4 métriques. Les résultats indiquent qu'il est nécessaire d'améliorer toutes les méthodes, en particulier les méthodes les plus performantes, afin de réduire la sensibilité à un bruit d'initialisation élevé.

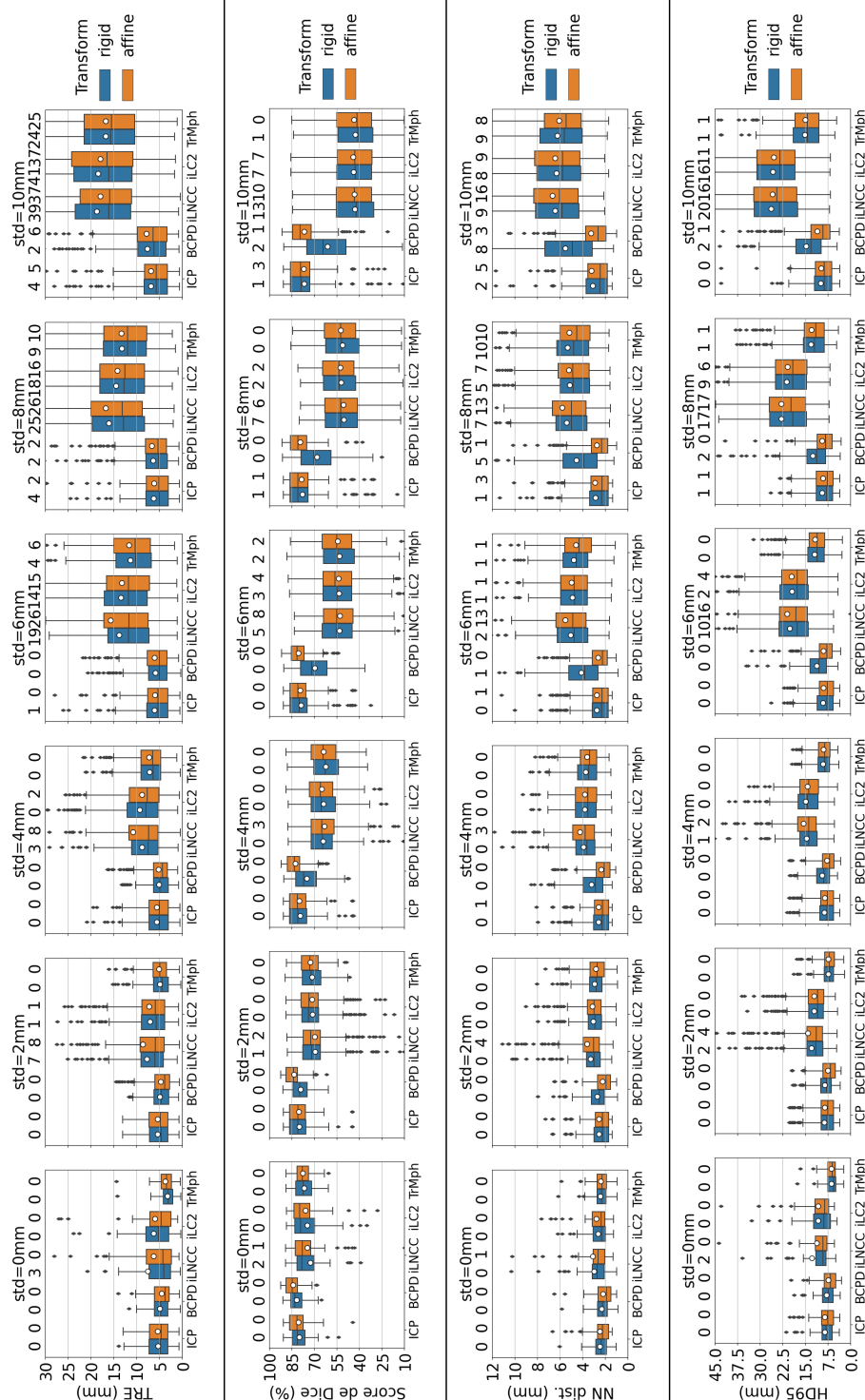


FIGURE 3.8 – Sensibilité des méthodes d'affinement du recalage aux erreurs d'initialisation sur TRUSTED, mesurée par NN dist, pour chaque écart-type (std). Les boîtes à moustaches indiquent les 3 quartiles, la moyenne (boule blanche), les valeurs extrêmes (moustaches) et aberrantes (pointillés et nombre sur les boîtes).

3.5.3 Résultats qualitatifs de recalage

La Figure 3.9 montre pour les cas 418R (patient 418, rein droit) et 680L (patient 680, rein gauche), les résultats de recalage précis avec initialisation manuelle (initialisation manuelle combinée à un affinement de recalage). Pour les affinements par ICP et BCPD, les segmentations automatiques dans les images US 3D et CT ont été fournies par nnU-Net entraîné avec cibles doubles. Les colonnes correspondent aux méthodes d’affinement du recalage. Les lignes correspondent aux hypothèses de transformation rigide ou affine, pour lesquelles une représentation en damier (lignes 1, 3, 5, 7) et une vue 3D (lignes 2, 4, 6, 8) des masques vérité terrain des reins après recalage, sont présentées. Les représentations en damier montrent sur une section de l’image CT fixe, deux morceaux remplacés par ceux de l’image US 3D recalée. Chaque représentation en damier porte 4 annotations : le contour et le point bleus sont la segmentation vérité terrain du rein et le point de repère vérité terrain #1 dans l’image CT. Le contour et le point rouges sont la segmentation vérité terrain du rein et le point de repère vérité terrain #1 dans l’image US 3D recalée selon la transformation estimée. Pour le cas 418R, les méthodes ICP et BCPD montraient un meilleur alignement tant des points de repère #1 que des surfaces vérité terrain qui, pour le rappeler, ne sont pas celles ayant été utilisées pour le recalage (recalage calculé avec les segmentations de nnU-Net). Les résultats de ICP et BCPD sous hypothèse de transformation affine étaient meilleurs que ceux sous hypothèse de transformation rigide. Pour le cas 680L, la précision de recalage était généralement moins bonne que pour le cas 418R, principalement en raison d’une ombre importante qui réduisait considérablement le contraste des bords du rein. Il est également difficile de conclure sur la qualité du recalage en se basant uniquement sur la métrique TRE. Par conséquent, les métriques de superposition des surfaces, comme le score de DICE et la NN dist., sont à prendre en compte, en plus de la TRE, lors de l’évaluation des méthodes sur cet ensemble de données.

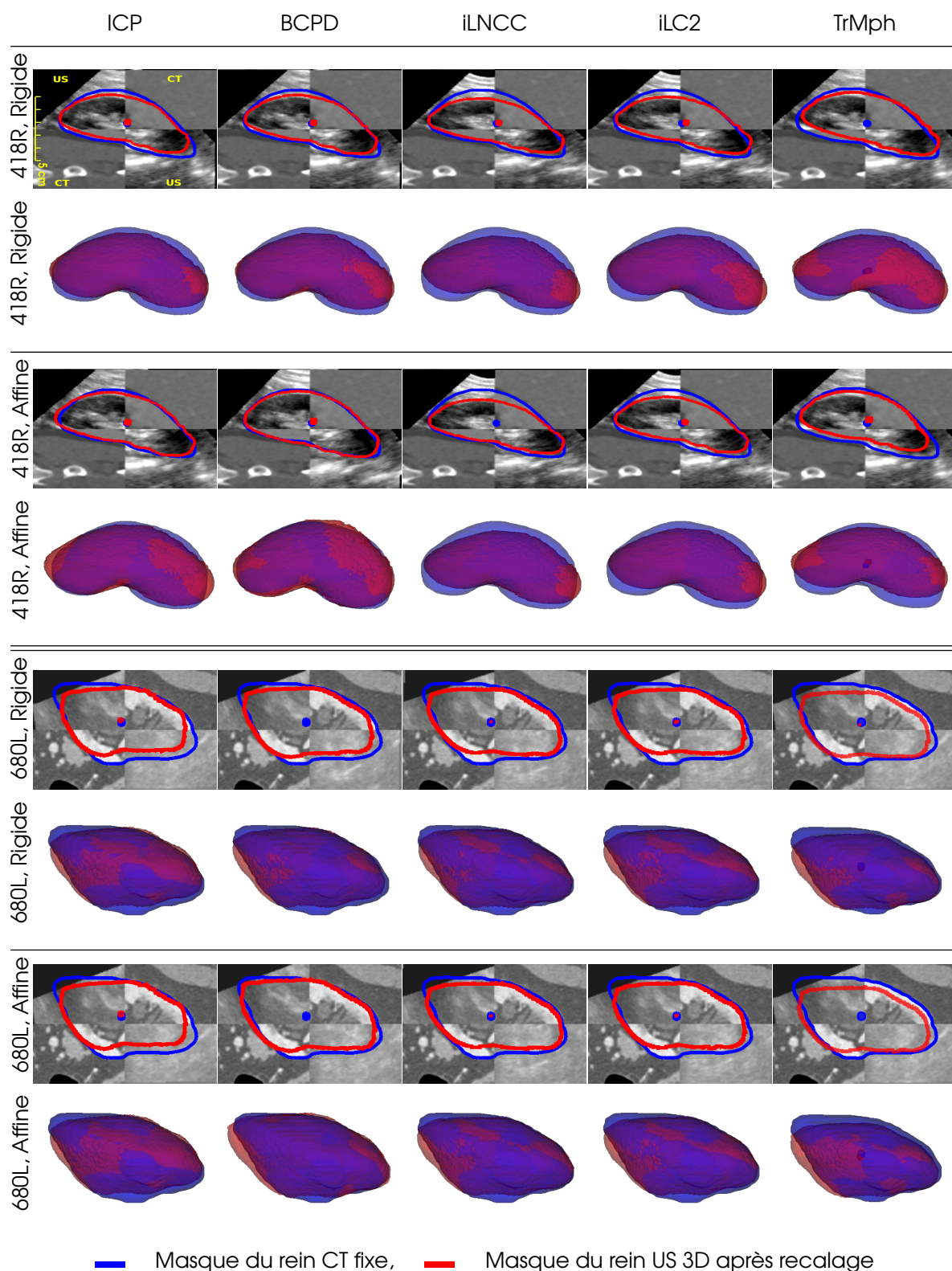


FIGURE 3.9 – Résultats qualitatifs montrant les performances de recalage sur les paires de reins 418R et 680L du jeu de données TRUSTED. Chaque colonne correspond à une méthode d’affinement de recalage. Les lignes montrent pour chaque hypothèse de transformation, les représentations en damier avec les annotations vérité terrain après recalage, et les vues 3D des régions vérité terrain des reins dans les deux modalités.

3.6 Conclusion

Ce chapitre présentait la première contribution scientifique importante de cette thèse. Il s'agit du jeu de données publiques TRUSTED, constitué d'images échographiques et tomodensitométriques appariées de reins de patients humains, dans lesquelles deux radiographes expérimentés ont indépendamment produit des annotations. Les annotations étaient constituées de masques de segmentation manuelle de rein et d'un ensemble de sept points de repère anatomiques du rein dans les deux modalités d'images. Des annotations vérité terrain ont été produites à partir des annotations manuelles et les concordances inter-annotateurs vis-à-vis des vérités terrain calculées ont été systématiquement évaluées. Le jeu de données TRUSTED a été évalué avec succès quant à sa capacité à servir pour l'entraînement, l'évaluation et la comparaison de méthodes de l'état de l'art actuel dans les domaines cibles, notamment celles utilisant des modèles à base de réseaux de neurones profonds. Une analyse de la robustesse des méthodes de recalage à une initialisation imprécise a également été conduite avec le jeu de données TRUSTED. Ces différentes expérimentations constituent le point de départ du potentiel du jeu de données TRUSTED en termes de développement méthodologique et d'applications en procédures cliniques guidées par l'image.

4. RECALAGE GLOBAL AUTOMATIQUE DES IMAGES ÉCHOGRAPHIQUES 3D ET TOMODENSITOMÉTRIQUES DU REIN

Résumé du chapitre

Ce chapitre propose le premier système entièrement automatique pour le recalage précis d'images échographiques tridimensionnelles et scanner du rein, visant à améliorer le guidage des interventions rénales. La méthode développée s'affranchit de l'initialisation manuelle, jusque-là nécessaire dans les approches de pointe existantes, en combinant une stratégie de recalage global innovante avec un affinement robuste. Les expérimentations conduites dans ce chapitre s'appuient sur le jeu de données TRUSTED proposé au chapitre précédent.

Dans un premier temps, une expérimentation utilisant des segmentations automatiques pour générer des hypothèses de recalage a été réalisée. Cette méthode reposait sur des caractéristiques de surface extraites du rein pour produire des hypothèses de recalage initial, mais elle a révélé plusieurs défis. La symétrie pôle supérieur - pôle inférieur du rein entraîne des ambiguïtés de recalage, et les segmentations automatiques manquent parfois de détails précis, compliquant l'obtention d'une initialisation stable et fiable. La résolution de ces limitations a conduit à une nouvelle approche, mieux performante et plus robuste.

La solution utilise des réseaux de neurones pour détecter et étiqueter sept points de repère anatomiques coplanaires, dans les images. Ces points sont utilisés dans une variante déterministe de la méthode RANSAC pour produire deux hypothèses de recalage global, qui correspondent aux ambiguïtés causées par la symétrie pôle supérieur - pôle inférieur du rein. Un réseau de neurones résout ensuite cette ambiguïté en identifiant la meilleure hypothèse de recalage global. Enfin, un affinement du recalage avec la méthode Bayesian Coherent

Point Drift (BCPD) est appliqué. Testée sur un ensemble de 59 paires d'images 3D échographiques et scanner du rein, la méthode proposée, combinée à l'affinement BCPD, atteint une erreur sur cible de recalage (Target Registration Error - TRE) de 5.78 mm sur le centre du pelvis rénal et une distance moyenne de 2.42 mm entre les surfaces recalées. Ces performances comparables à celles des méthodes semi-automatiques (initialisation manuelle et affinement automatique), surpassent plusieurs systèmes automatiques de pointe, qui nécessitent en outre une segmentation de l'organe à l'initialisation.

Le chapitre a deux sections principales. La première décrit des expérimentations menées avec les approches automatiques de pointe pour le recalage global utilisant les surfaces. La deuxième section présente la nouvelle approche automatique de recalage global utilisant sept points de repère anatomiques coplanaires. Elle constitue une avancée importante vers l'automatisation des systèmes complets de recalage précis.

4.1 Recalage global automatique utilisant les surfaces

Il s'agit d'approches automatiques de recalage utilisant les segmentations des organes ou des structures que l'on souhaite aligner. En contexte de recalage multimodal, cette information intermédiaire (masque de segmentation) permet de briser le gap entre les distributions d'intensité des différentes modalités, en les plongeant dans un domaine identique. Ce qui rend possible l'utilisation de métriques classiques d'optimisation telles que l'erreur quadratique moyenne. En pratique, les surfaces segmentées sont discrétisées et représentées en nuages de points 3D. Les caractéristiques globales de la géométrie de ces nuages sont alors mises en correspondances, puis alignées.

4.1.1 L'information de surface

La surface provient de la segmentation de la structure d'intérêt. Il s'agit de délimiter la région précise occupée par cette dernière. Plusieurs approches pour réaliser la segmentation automatique dans les images médicales ont été développées avant que l'utilisation des réseaux de neurones profonds devienne le standard [182]. Désormais, la quasi-totalité des approches utilise des modèles de réseaux de neurones profonds (DNNs). Ainsi, les images CT et US 3D sont traitées par deux réseaux de neurones convolutifs (CNNs), pour la segmentation

de la région du rein dans chaque modalité. La segmentation est suivie d'un maillage de la surface et d'une génération de nuages de points pour chaque rein (Figure 4.1). Pour ce qui est des images US 3D, le champ de vision limité des appareils d'échographie 3D actuels ne permet de voir qu'un seul rein sur l'image. Le côté du rein associé à l'image US 3D est connu *a priori* dans pratiquement tous les scénarios d'application. Par conséquent, le CNN pour la modalité US 3D est entraîné pour effectuer une segmentation sémantique du rein (étiquetage binaire) sans avoir besoin de distinguer le rein gauche ou droit. L'image US 3D est d'abord redimensionnée à une résolution canonique ($256 \times 256 \times 256$), puis passée en entrée à un CNN de segmentation sémantique. Toute architecture de CNN compétitive peut être utilisée, et le choix a été porté sur la méthode nnU-Net [183], en vertu de ces récentes très bonnes performances en matière de segmentation d'images médicales. Le réseau étiquette chaque voxel comme étant à l'intérieur ou à l'extérieur de l'organe. L'algorithme Marching Cubes [184] est ensuite utilisé pour extraire des maillages de surface à partir des segmentations, puis les nuages de points associés sont générés en utilisant les sommets du maillage (un sommet par voxel à la résolution canonique).

Contrairement à la modalité échographique, les deux reins sont normalement présents dans une image CT abdominale. Tout d'abord, l'image CT est redimensionnée à une résolution canonique ($256 \times 256 \times 256$), puis passée en entrée à un CNN entraîné pour effectuer une segmentation sémantique des reins (CNN actuellement implémenté avec un nnU-Net). Comme les images CT ont un système de coordonnées spatiales standard, le rein gauche ou droit est facilement identifié comme la plus grande région segmentée sur la moitié gauche ou droite de l'image CT. L'algorithme Marching Cubes est ensuite utilisé pour extraire le maillage et le nuage de points du rein CT associé au rein dans l'image US.

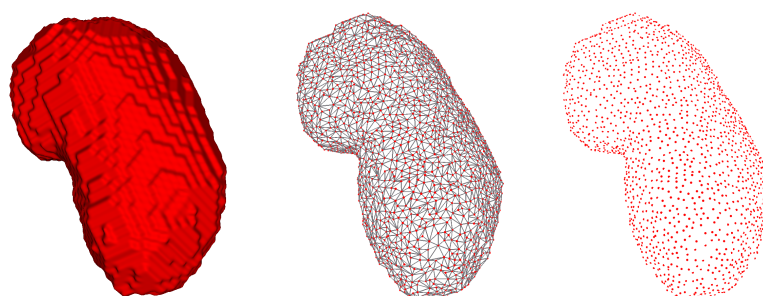


FIGURE 4.1 – De gauche à droite, visualisation du masque de segmentation automatique du rein dans un volume échographique, puis du maillage et du nuage de points associés.

4.1.2 Recalage global automatique utilisant l'Analyse en Composantes Principales (ACP) des géométries

Les composantes principales des nuages de points issus des deux modalités peuvent être utilisées afin d'estimer une fonction de recalage global $g : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ qui fait correspondre automatiquement le nuage échographique (mobile) avec le nuage scanner (fixe). M_m et M_f désignent les maillages 3D issus des segmentations de région de l'échographie et du scanner, respectivement.

Tout d'abord, chaque maillage 3D est pivoté de sorte que ses trois axes principaux, classés par variance décroissante, s'alignent sur les axes x , y et z du système de coordonnées auquel le maillage appartient. R_m et R_f désignent les matrices de rotation pour l'image mobile et l'image fixe, respectivement.

Ensuite, les points d'intersection entre chaque maillage et les axes x , y sont déterminés (Figure 4.2). Ce qui produit 4 points de repère géométriques par nuage de points, deux par axe. Notons que les intersections avec le troisième axe principal ne sont pas nécessaires dans la mesure où la transformation recherchée est une similarité en 3D, qui peut être estimée avec minimalement 3 paires de points. Les points de repère associés à M_m et M_f respectivement sont désignés par $L_m \in \mathbb{R}^{4 \times 3}$ et $L_f \in \mathbb{R}^{4 \times 3}$.

Enfin, g est estimée comme la transformation de similarité des moindres carrés qui aligne L_m avec L_f , tel que décrit dans [185].



FIGURE 4.2 – Nuage de points CT (vert) et US 3D (rouge), leurs deux premiers axes principaux, et les points de repère géométriques déterminés aux intersections nuage-axe.

Les sous-matrices de rotation estimées ne sont pas définies de manière unique en raison des ambiguïtés de signe des axes (vecteurs) principaux. Il y a quatre

solutions possibles de transformations globales distinctes, chacune équivalente à une rotation de 180 degrés du rein autour de ses axes principaux primaire et secondaire (Figure 4.3). L'ensemble de ces 4 hypothèses est désigné par : $\mathcal{T} = \{\mathbf{T}_-, \mathbf{T}_+, \mathbf{T}_+^+, \mathbf{T}_-^+\}$, où les indices et exposants inconnus indiquent si l'orientation est correcte suivant le premier axe principal et le second axe principal, respectivement.

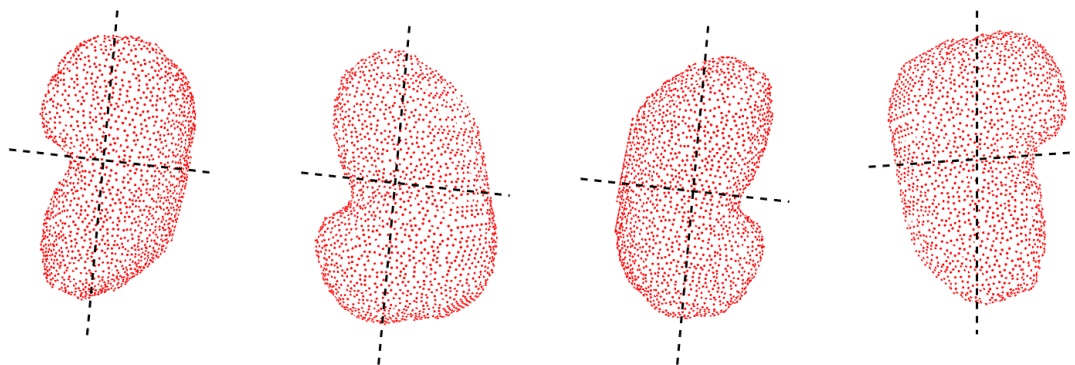


FIGURE 4.3 – Positions du nuage de points US suivant les 4 hypothèses possibles de recalage global issues de l'approche par ACP.

Une stratégie automatique naïve pour gérer ces ambiguïtés consiste à utiliser les quatre solutions pour initialiser un recalage fin, puis conserver la solution finale avec l'erreur d'ajustement la plus faible, par exemple la plus petite distance moyenne du plus proche voisin (nn-dist - pour "nearest neighbour distance" en anglais) entre les nuages de segmentations automatiques, après recalage. Le succès de cette approche de sélection dépend de la qualité de la segmentation automatique.

4.1.3 Recalage global automatique utilisant des descripteurs locaux de surface

Les approches dont il est question ici calculent des descripteurs locaux de points (ou des descripteurs de patches de points) dans les nuages US 3D et CT, puis établissent des appariements entre ceux-ci par recherche du plus proche voisin dans l'espace des descripteurs calculés (Figure 4.4), enfin un algorithme de type RANSAC ("RANdom SAmple Consensus" en anglais) [106] ou TEASER++ [111] est utilisé pour estimer la transformation (similarité) optimale qui aligne les deux nuages de points en minimisant les erreurs de correspondance.

Les méthodes "Fast Point Feature Histograms" (FPFH) [110] et "Distinctive 3D local deep descriptors" (DIP) [129] ont la réputation de pouvoir s'adapter à

différentes géométries et à diverses résolutions de nuages de points, offrant une certaine flexibilité.

La méthode FPFH calcule des descripteurs qui sont des caractéristiques des histogrammes de distributions de primitives géométriques locales calculées entre chaque point et ses voisins. C'est une approche conçue pour être rapide. Les descripteurs sont simples à calculer et invariants aux transformations rigides. La méthode DIP repose sur l'apprentissage profond pour générer des descripteurs locaux 3D des points. La méthode vise à apprendre des représentations locales à la fois générales et distinctives.

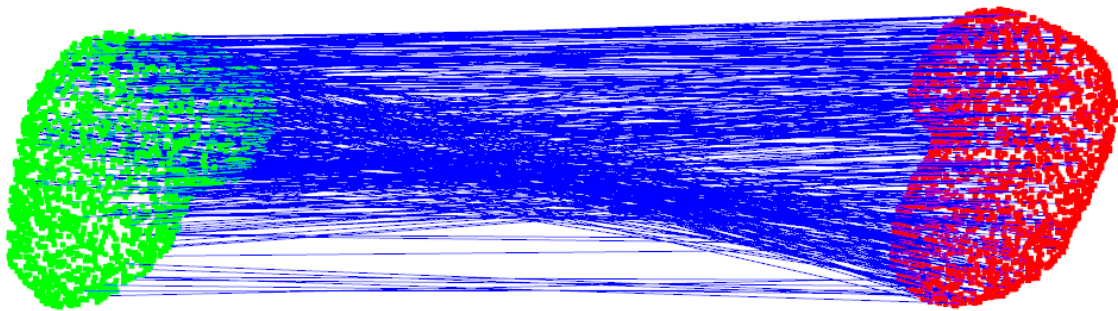


FIGURE 4.4 – Exemple d'appariement de points des surfaces du rein dans l'image CT (vert) et dans l'image US 3D (rouge), par recherche du plus proche voisin dans l'espace des descripteurs DIP.

4.1.4 Expériences

Rappelons que ce chapitre aborde le recalage global comme méthode d'initialisation d'un recalage précis ou affiné. Les expérimentations visent à évaluer les performances des méthodes de recalage global utilisant les surfaces, sur l'ensemble des deux étapes : initialisation du recalage, puis affinement du recalage. Le jeu de données TRUSTED est utilisé dans un contexte de validation croisée à 5 blocs. La partition en blocs est exactement la même que celle utilisée pour les expérimentations dans le chapitre 3. Pour les approches utilisant les descripteurs locaux (FPFH, DIP), deux méthodes robustes d'estimation de la transformation de recalage global sont expérimentées, à la suite du calcul des descripteurs : RANSAC (Random Sample Consensus) [106] et TEASER++ [111]. Les hyper-paramètres de FPFH et DIP, notamment l'échelle des caractéristiques (rayon de voisinage) ont été optimisés sur chaque ensemble d'entraînement de validation croisée, puis bloqués avant l'évaluation sur chaque ensemble

de test. Tous les réseaux de segmentation automatique ont été entraînés sur chaque ensemble d'entraînement et utilisés en inférence sans optimisation sur l'ensemble de test correspondant. Enfin, toutes les méthodes de recalage global ont été utilisées comme initialisation pour la méthode d'affinement appelée "Bayesian Coherent Point Drift" (BCPD) [117]. Cette dernière utilise les surfaces et a obtenu de bonnes performances dans des tâches similaires de recalage d'images médicales, comme le recalage de vaisseaux du foie [62].

Les différentes configurations comparées sont donc :

1. **(PCA, nndist)**. Estimation de la transformation de recalage global utilisant l'ACP tel que décrit dans la section 4.1.2.
2. **(FPFH, RANSAC)**. Estimation de la transformation de recalage global utilisant un appariement des points dans l'espace des descripteurs FPFH calculés, suivi d'un calcul par l'algorithme RANSAC.
3. **(DIP, RANSAC)**. Estimation de la transformation de recalage global utilisant un appariement des points dans l'espace des descripteurs locaux DIP appris, combiné à RANSAC.
4. **(FPFH, TEASER)**. Estimation de la transformation de recalage global utilisant un appariement des points dans l'espace des descripteurs FPFH calculés, combiné à TEASER++ [111].
5. **(DIP, TEASER)**. Estimation de la transformation de recalage global utilisant un appariement des points dans l'espace des descripteurs locaux DIP appris, combiné à TEASER++.
6. **(Manuelle)**. Estimation de la transformation de recalage global utilisant la transformation de similarité calculée à partir des points de repère de vérité terrain. Cette configuration sert de standard à atteindre. Elle est incluse pour évaluer l'écart de performance entre les méthodes automatiques et la méthode manuelle.

4.1.4.1 Les métriques d'évaluation

Les principales métriques d'évaluation utilisées sont :

- La distance de surface moyenne des plus proches voisins (**nndist**), en mm. Cette mesure quantifie le degré de chevauchement spatial entre l'organe dans les deux modalités après recalage. Elle a été implémentée en convertissant les segmentations CT et US 3D vérité terrain en maillages à l'aide de l'algorithme Marching Cubes. Ensuite, le maillage US 3D a été transformé spatialement en fonction du recalage calculé avec les segmentations automatiques, et la distance moyenne au plus proche voisin entre tous les

sommets a été calculée par rapport au maillage CT. La métrique $nndist$ quantifie la superposition des surfaces après recalage.

- Erreur sur cible de recalage (TRE, pour "Target Registration Error"), en mm. La position vérité terrain de chaque point de repère numéro $X \in \{1, \dots, 7\}$ dans l'image CT (suivant la convention de numérotation indiquée dans la Figure 4.9) est utilisée comme cible de recalage. **TRE-X** désigne la distance entre le point de repère X dans l'image US 3D transformé spatialement en fonction du recalage calculé et le point de repère cible X dans l'image CT fixe.

Tel qu'indiqué dans le chapitre 3 décrivant les analyses du jeu de données TRUSTED, il a été difficile, même pour des experts annotateurs, de placer les points de repère appariés numéros 2 à 7 dans les images US 3D et CT avec une grande précision. Ce qui n'entame en rien la valeur de ces annotations (points de repère manuels) puisqu'elles servent à développer des méthodes de recalage global vues comme initialisation d'un recalage affiné. Une grande précision dans leur localisation n'est donc pas nécessaire. Par contre, le point de repère numéro 1 (le centre du pelvis rénal) a été placé dans les deux modalités avec précision (concordance inter-annotateurs moyenne : 1.82 mm dans les CT et 2.23 mm dans les US). Ainsi, les différents TRE-X ne renvoient pas la même information. TRE-1 est beaucoup plus sensible à l'erreur sur la cible de recalage qu'est le pelvis rénal. TRE-4,6 et TRE-5,7 sont plus sensibles à la qualité de l'orientation "pôle supérieur - pôle inférieur" du rein dans l'image US 3D après recalage. Plus ces valeurs sont faibles, plus l'orientation "pôle supérieur - pôle inférieur" est correcte. TRE-2 et TRE-3 sont plus sensibles à la qualité de l'orientation transversale du rein dans l'image US 3D après recalage. Plus ces valeurs sont faibles, plus l'orientation transversale est correcte.

4.1.4.2 Résultats quantitatifs

Les résultats quantitatifs sont présentés par les graphiques de la Figure 4.5, et les résumés statistiques dans le Tableau 4.1.

Les graphiques de la Figure 4.5 montrent la distribution des métriques de recalage ($nndist$, TRE-1, TRE-4 et TRE-5) pour chacune des méthodes étudiées dans cette section. Les axes des ordonnées représentent différents seuils de métriques et les axes des abscisses indiquent le nombre de reins pour lesquels la métrique évaluée était inférieure au seuil de métrique. Nous constatons que les courbes de **(DIP, TEASER)** et **(PCA, $nndist$)** sont généralement plus proches de la courbe de **Manuelle** (la référence) que celles des autres méthodes.

S'agissant du Tableau 4.1, les colonnes présentent les métriques de performance pour chaque configuration par rapport à la valeur moyenne (Avg - pour

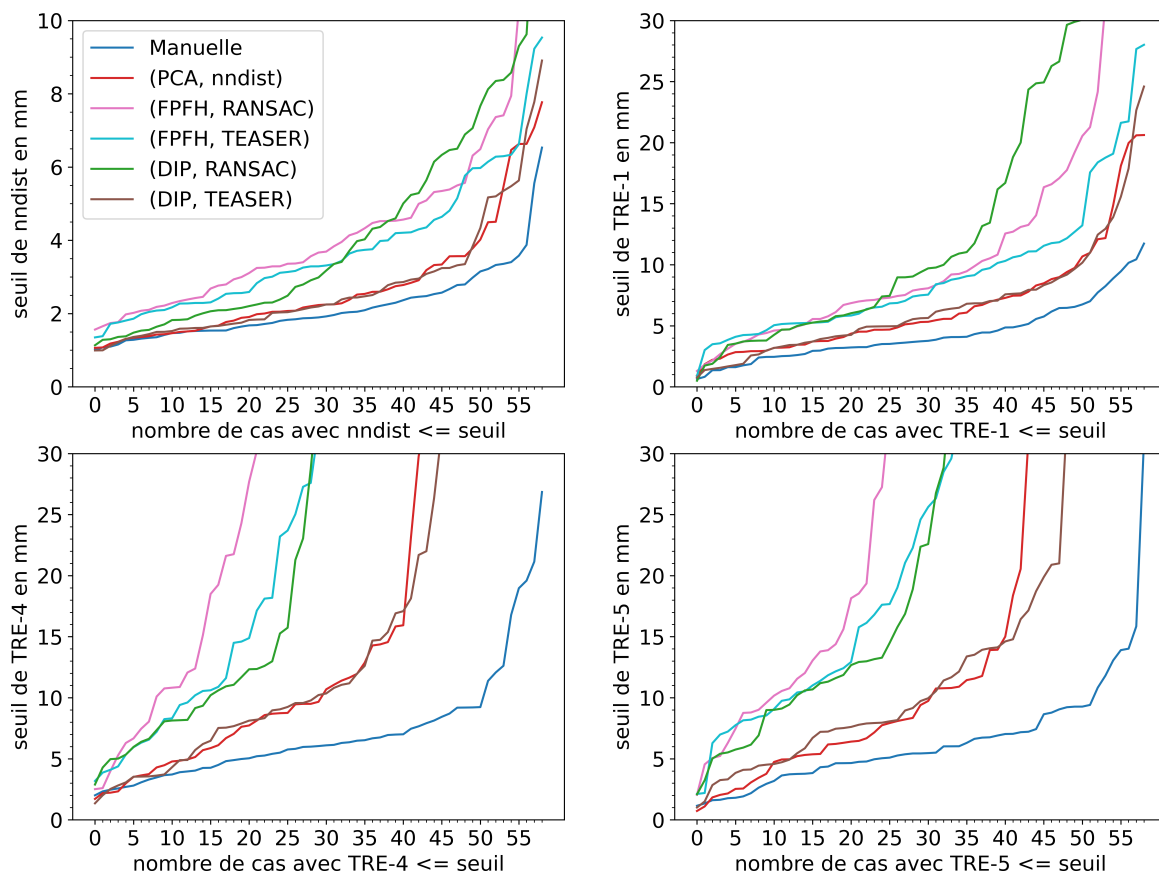


FIGURE 4.5 – Performances des méthodes de recalage global utilisant les surfaces, combinées à l’affinement par BCPD. Les distributions de distance du plus proche voisin (nndist) et de TRE-1 sont présentées sur la ligne du haut. Les distributions de TRE-4 et de TRE-5 sont présentées sur la ligne du bas. Les axes des ordonnées représentent différents seuils de métriques et les axes des abscisses représentent le nombre de reins pour lesquels la métrique évaluée après recalage était inférieure au seuil de métrique.

"average" en anglais), l’écart-type (Std - pour "standard deviation" en anglais) et le troisième quartile (q3). Ce dernier a été inclus, car il peut clairement révéler l’existence de pics forts dans les distributions de performance. Étant donné le contexte de validation croisée, les résultats ont été calculés en utilisant les moyennes par blocs. Les méthodes comparées ici ont toutes été évaluées après affinement par la méthode BCPD, puisque l’objectif de cette contribution est de proposer une méthode automatique d’initialisation du recalage qui, en combinaison avec une méthode d’affinement, produit un alignement précis de l’organe dans les deux modalités (TRE < 5 mm en contexte clinique). Le choix de la méthode BCPD pour l’affinement s’explique par ses bonnes performances rapportées dans [186], en comparaison à d’autres méthodes populaires d’affinement. Les caractères gras (dans le tableau) indiquent la meilleure valeur suivant chaque métrique. La configuration **(PCA, nndist)** présente la meilleure

performance en moyenne pour les métriques *ndist* (2,7 mm) et *TRE-1* (6,57 mm), ainsi qu’une performance supérieure en termes de *q3* pour *TRE-3* (17,54 mm). La configuration **(DIP, TEASER)** offre les meilleures performances pour les autres métriques de *TRE* (*TRE-2*, *TRE-4*, *TRE-5*, *TRE-6* et *TRE-7*), tant en moyenne (de 11,14 mm à 29,05 mm) qu’en termes de *q3* (de 13,11 mm à 34,34 mm), ainsi que le meilleur *q3* pour *TRE-1* (7,96 mm) et *ndist* (également 2,7 mm).

Ces résultats montrent que **(DIP, TEASER)** et **(PCA, *ndist*)** sont deux méthodes compétitives pour le recalage global du rein, parmi les méthodes utilisant l’information de surface comparées dans cette étude.

Résultats quantitatifs sans affinement du recalage

Ce paragraphe présente, pour chaque configuration d’initialisation, la contribution de l’affinement BCPD dans l’ensemble du processus de recalage précis (initialisation et affinement). A cet effet, le Tableau 4.2 donne les résultats de performances isolées, des méthodes de recalage global automatique. La structure du Tableau 4.2 est identique à celle du Tableau 4.1. L’observation des deux tableaux montre que l’affinement a un impact particulièrement marqué sur la méthode **(DIP, TEASER)**, avec des réductions notables dans les métriques *ndist* et *TRE*. Par exemple, la moyenne de la métrique *ndist* passe de 3.94 mm (sans affinement) à 2.7 mm (avec affinement). Les écarts-types (*Std*) ont également tendance à diminuer après affinement pour la configuration **(DIP, TEASER)**. Cela suggère que l’affinement non seulement réduit les erreurs moyennes de **(DIP, TEASER)**, mais apporte aussi une plus grande stabilité à l’ensemble du processus de recalage initialisé avec **(DIP, TEASER)**.

Méthodes	nndist↓ (mm)			TRE-1 ↓ (mm)			TRE-4 ↓ (mm)			TRE-5 ↓ (mm)		
	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3
(Manuelle)	2.16	0.4	2.16	4.34	0.53	4.45	7.1	2.35	7.86	6.4	1.54	7.51
(PCA, nndist)	2.7	0.57	2.82	6.57	1.66	8	37.92	18.53	52.06	33.58	15.46	47.6
(FPFH, RANSAC)	4.62	0.99	5.09	11.7	2.72	13.28	65.82	6.53	72.11	61.7	7.86	67.65
(FPFH, TEASER)	3.76	0.44	3.75	9.26	1.31	8.94	50.21	7.76	55.43	46.9	5.36	49.62
(DIP, RANSAC)	4.16	0.74	4.57	15.46	4.46	20.09	52.82	6.6	56.83	44.01	4.94	45.65
(DIP, TEASER)	2.7	0.69	3.02	6.61	1.73	7.96	29.05	10.76	32.73	25.93	8.4	34.34

Méthodes	TRE-2↓ (mm)			TRE-3 ↓ (mm)			TRE-6 ↓ (mm)			TRE-7 ↓ (mm)		
	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3
(Manuelle))	5.94	1.36	6.9	5.13	0.48	4.98	8.78	1.97	9.79	6.27	1.15	6.68
(PCA, nndist)	12.09	3.33	15.04	17.54	5.96	22.84	29.87	12.37	36.38	26.59	9.56	36.42
(FPFH, RANSAC)	23.41	2.38	24.96	37.83	4.78	40.63	51.37	3.7	54.74	54.71	7.02	60.93
(FPFH, TEASER)	21.68	3.22	22.85	33.26	5.21	38.31	47.89	5.39	49.3	44.03	4.49	47.49
(DIP, RANSAC)	21.82	5.28	26.41	32.66	6.27	39.32	43.26	8.24	51.47	35.65	3.33	36.79
(DIP, TEASER)	11.14	3.29	13.11	17.65	3.54	17.11	23.5	5.07	29.08	20.4	6.31	21.98

TABLE 4.1 – Comparaison quantitative des méthodes automatiques de recalage global utilisant les surfaces, combinées avec l’affinement BCPD. Le symbole (↓) indique que les faibles valeurs sont les meilleures. Les résultats sont représentés sur deux blocs de tableau : en haut - les résultats des métriques nndist, TRE-1, TRE-4 et TRE-5 ; en bas - les résultats des métriques TRE-2, TRE-3, TRE-6 TRE-7. Les méthodes sont divisées en deux groupes : les résultats de la méthode **Manuelle**, utilisant les points de repère manuels ; c’est la référence utilisée en comparaison avec les méthodes automatiques. Puis, les résultats des méthodes comparées. Les colonnes présentent respectivement la moyenne (Avg), l’écart-type (Std) et le troisième quartile (q3) entre les blocs de validation croisée, des métriques nndist, TRE-1, TRE-4 et TRE-5 pour chaque méthode. Les valeurs en gras indiquent la meilleure valeur de la mesure.

Méthodes	nndist↓ (mm)			TRE-1 ↓ (mm)			TRE-4 ↓ (mm)			TRE-5 ↓ (mm)		
	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3
(Manuelle)	2.36	0.28	2.65	2.28	0.30	2.37	3.88	0.66	3.95	2.77	0.31	2.93
(PCA, nndist)	3.21	0.30	3.44	6.34	0.07	7.15	37.34	16.85	51.56	34.13	14.71	46.90
(FPFH, RANSAC)	5.86	0.75	6.28	21.85	3.53	21.76	72.13	4.41	76.36	70.02	7.90	73.96
(FPFH, TEASER)	5.96	0.71	6.58	16.02	1.27	16.36	60.08	8.26	66.38	53.88	3.59	55.54
(DIP, RANSAC)	5.29	0.80	5.60	26.56	4.50	30.46	60.46	6.64	64.06	52.59	5.35	57.66
(DIP, TEASER)	3.94	1.17	4.50	10.05	1.94	11.26	33.98	10.98	37.32	28.45	7.78	36.99

Méthodes	TRE-2↓ (mm)			TRE-3 ↓ (mm)			TRE-6 ↓ (mm)			TRE-7 ↓ (mm)		
	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3	Avg	Std	q3
(Manuelle)	3.46	0.68	3.86	3.11	0.32	3.37	5.39	0.87	5.24	4.12	0.55	4.35
(PCA, nndist)	12.32	2.92	15.01	17.90	5.52	23.06	29.85	11.32	37.86	27.70	8.46	35.12
(FPFH, RANSAC)	31.15	3.70	34.44	41.56	4.61	42.84	56.18	4.13	58.85	60.04	7.92	68.83
(FPFH, TEASER)	25.07	3.19	25.03	36.21	4.36	40.18	51.35	5.11	51.75	46.42	5.23	48.33
(DIP, RANSAC)	31.45	3.46	33.95	37.28	7.70	42.24	49.92	7.33	58.12	44.51	2.88	46.49
(DIP, TEASER)	13.73	3.63	17.57	19.11	3.01	18.64	26.34	6.50	32.89	22.46	6.25	23.74

TABLE 4.2 – Comparaison quantitative des méthodes automatiques de recalage global utilisant les surfaces, mais sans affinement. Le symbole (↓) indique que les faibles valeurs sont les meilleures. Les résultats sont représentés sur deux blocs de tableau : en haut - les résultats des métriques nndist et TRE-1, TRE-4, TRE-5 ; en bas - les résultats des métriques TRE-2, TRE-3, TRE-6, TRE-7. Les colonnes présentent respectivement la moyenne (Avg), l'écart-type (Std) et le troisième quartile (q3) entre les blocs de validation croisée des métriques nndist, TRE-1, TRE-4 et TRE-5, pour chaque méthode. Les valeurs en gras indiquent la meilleure valeur de la mesure.

4.1.4.3 Résultats qualitatifs

Les Figures 4.6 et 4.7 montrent chacune, pour deux reins, des visualisations de recalage avec les méthodes globales utilisant les surfaces, après affinement avec BCPD. Dans chacune des figures, les lignes 1 et 3 montrent des visualisations en damier représentant en première diagonale le rein dans l'image fixe CT et en seconde diagonale le rein dans l'image US 3D transformée dans le système de coordonnées du CT, par la transformation estimée. Sur ces représentations en damier, sont superposées les segmentations vérité terrain dans chaque modalité et le centre du pelvis rénal (point de repère numéro 1). Les lignes 2 et 4 quant à elles présentent des vues 3D des segmentations vérité terrain dans le système de coordonnées du CT après recalage. Ces exemples ont été sélectionnés systématiquement en triant les reins en fonction de la difficulté de recalage, puis en prenant un exemple du 1er quart des meilleures performances générales (118L), un exemple du 2nd quart (20R), un exemple du troisième (14L) et un exemple du quart des pires performances générales (94R). La difficulté de recalage de chaque paire d'images a été évaluée en calculant la valeur médiane, sur les méthodes, de la métrique *nndist* pour ladite paire. Les symboles ✓ et ✗ indiquent si les orientations "pôle inférieur-pôle supérieur" et transversale du rein dans l'image US 3D recalée, sont correctes.

On peut constater une bonne performance dans 3 cas sur 4 pour les méthodes (**DIP, TEASER**) et (**PCA, nndist**).

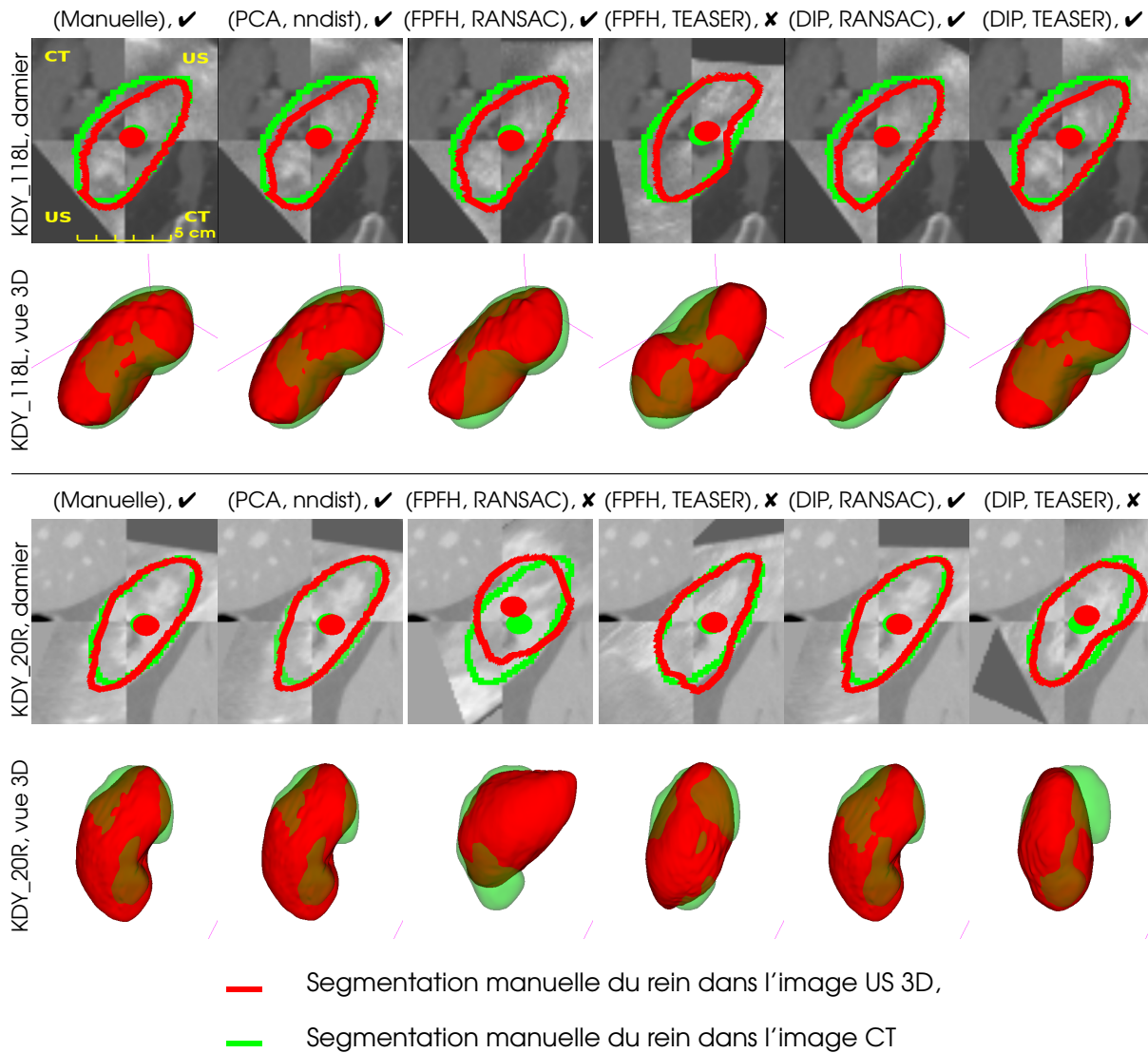


FIGURE 4.6 – Résultats qualitatifs illustrant les performances de recalage pour les exemples (KDY_118L) et (KDY_20R) à l'aide d'une visualisation en damier (deux morceaux du CT sur la première diagonale, deux morceaux de l'image US recalée sur la deuxième diagonale) et d'une vue 3D. Chaque colonne correspond à une méthode de recalage global suivie d'un affinement BCPD. Sur les damiers, quatre annotations sont observables : Contour vert : segmentation vérité terrain du rein dans le CT. Point vert : position vérité terrain du point de repère numéro 1 dans le CT. Contour rouge : segmentation vérité terrain du rein dans l'image US déplacée dans le système de coordonnées du CT selon le recalage estimé. Point rouge : position vérité terrain du point de repère numéro 1 dans l'image US, déplacé dans le système de coordonnées du CT selon le recalage estimé. Les symboles ✓ et ✗ indiquent si les orientations "pôle inférieur-pôle supérieur" et transversale sont correctes.

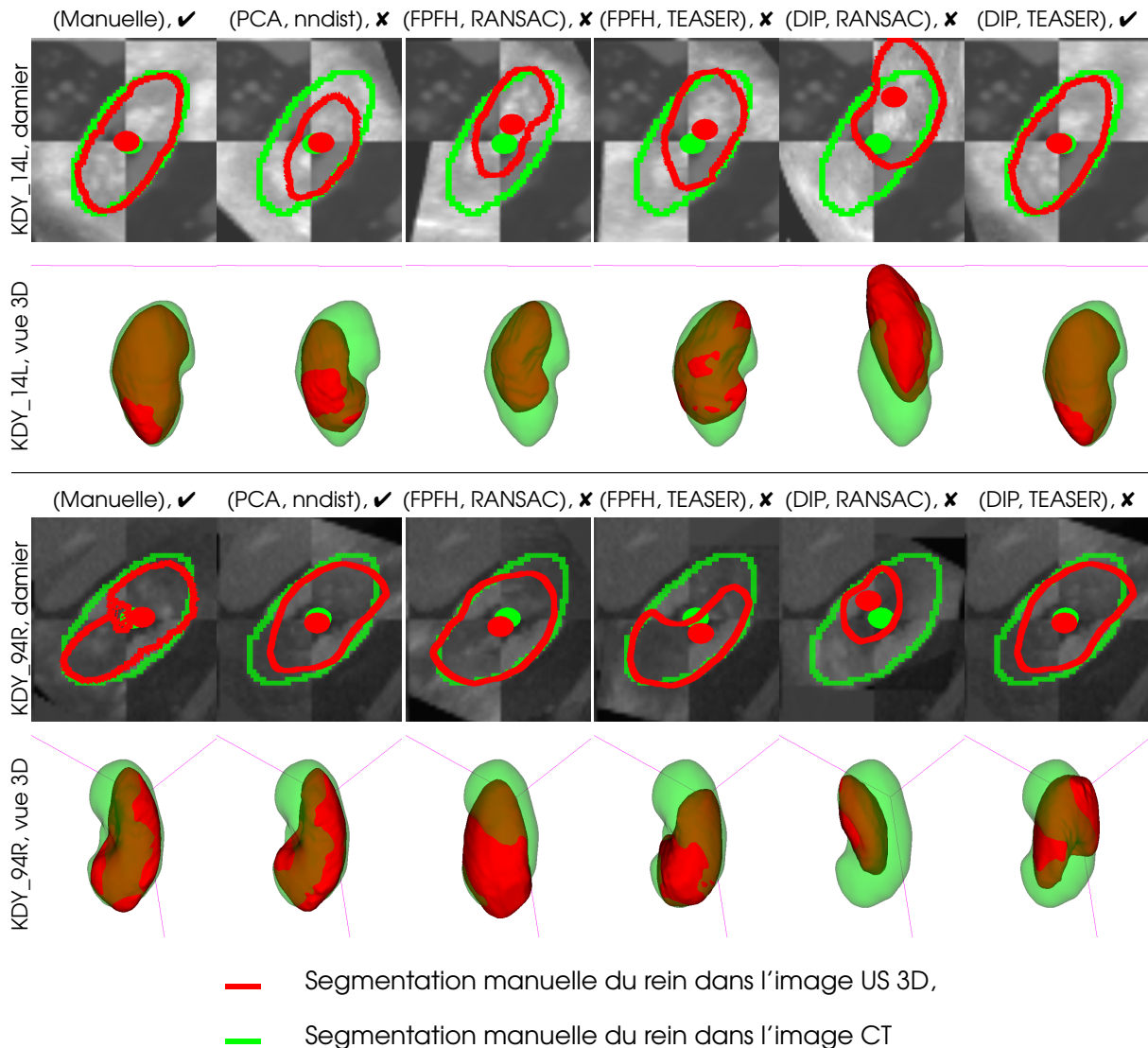


FIGURE 4.7 – Résultats qualitatifs illustrant les performances de recalage pour les exemples (KDY_14L) et (KDY_94R) à l'aide d'une visualisation en damier (deux morceaux du CT sur la première diagonale, deux morceaux de l'image US recalée sur la deuxième diagonale) et d'une vue 3D. La description est la même que celle de la Figure 4.6.

4.1.5 Conclusion et discussion

Dans cette section, deux principales approches de recalage global automatique utilisant l'information de surface, ont été présentées. Dans le cas du recalage du rein dans les images US 3D et CT, cette information de surface permet de briser le gap entre les deux domaines de modalité. Toutefois, ces méthodes sont dépendantes des segmentations du rein dans les images. Cette segmentation n'est pas une tâche simple, notamment dans les images US 3D, tel que l'ont montré les résultats du chapitre 3.

De plus, l'observation des résultats qualitatifs de recalage (Figures 4.6 et 4.7) a révélé deux types d'erreur d'orientation, dont au moins une survient au moins une fois dans les 4 exemples présentés, pour chaque méthode. La première était une erreur d'orientation transversale observée clairement 6 fois sur les 12 cas d'erreur flagrante dans ces figures : 118L (**FPFH,TEASER**), 20R (**FPFH,TEASER**), 20R (**DIP,TEASER**), 14L (**DIP,RANSAC**), 94R (**FPFH,RANSAC**), 94R (**FPFH,TEASER**). Malgré la symétrie transversale dans la forme globale du rein, cette erreur peut être facilement résolue si, sur l'axe transversal de rein, la région locale de l'artère rénale est strictement distinguée de la région locale opposée (Figure 2.8). La différence de courbure de la surface de rein dans les deux régions peut être utilisée. En effet, dans la région de l'artère rénale, la surface du rein est concave, alors que dans la région en face (en restant sur l'axe transversal) elle est convexe. La deuxième erreur est celle de l'orientation "pôle supérieur - pôle inférieur". Elle est observable clairement 8 fois sur les 12 cas d'erreur flagrante dans ces figures : 20R (**FPFH,RANSAC**), 14L (**PCA,ndst**), 14L (**FPFH,RANSAC**), 14L (**FPFH,TEASER**), 14L (**DIP,RANSAC**), 94R (**FPFH,TEASER**), 94R (**DIP,RANSAC**), 94R (**DIP,TEASER**). Cette seconde erreur est plus difficile à résoudre géométriquement. En effet, en plus de la symétrie globale dans la forme du rein sur son axe "pôle supérieur - pôle inférieur", les régions locales pôle supérieur et pôle inférieur sont fortement symétriques. Cette double symétrie (globale et locale) rend difficile la résolution de cette seconde ambiguïté en utilisant des outils de géométrie.

Dans la deuxième section de ce chapitre, une nouvelle approche automatique pour le recalage global automatique du rein dans les images US 3D et CT est proposée. Celle-ci fait face aux principales difficultés de la modalité échographique (fenêtre limitée, faible contraste, ombre acoustique,...), en évitant la dépendance par rapport aux segmentations, et en résolvant le problème d'ambiguïté causé par les symétries globales et locales dans la forme du rein.

4.2 Une approche de recalage global automatique par points de repère étiquetés

4.2.1 Objectifs et méthode générale

Dans cette section, est présentée, une nouvelle méthode utilisant des points de repère anatomiques pour le recalage global automatique du rein dans les images échographiques 3D et tomodensitométriques. L'approche est motivée par un défi lié à deux difficultés inhérentes aux images échographiques du rein.

- La taille limitée de la fenêtre échographique : il y a peu de couverture échographique sur les organes (ou structures) voisin(e)s. Pourtant, cette information peut servir à la compréhension de l'orientation du rein dans les images US.
- Le faible contraste et les ombres présentes en raison des os des côtes.

Ainsi, il peut être très difficile, même pour un expert en imagerie médicale, de comprendre correctement l'orientation rotationnelle du rein à partir des seules données de l'image échographique, en raison de la symétrie pôle supérieur - pôle inférieur. Cette difficulté est significativement réduite si une image appariée complémentaire, issue d'une autre modalité (comme la tomodensitométrie), est disponible. Celle-ci révèle des informations anatomiques spécifiques au patient, sur les structures et organes environnants, tels que les vaisseaux principaux, le foie, la rate. Les informations de la tomodensitométrie et de l'échographie peuvent alors être mentalement corroborées pour comprendre la relation spatiale globale entre les images. Une modélisation informatique de ce processus mental permet, grâce au recalage automatique, d'utiliser l'échographie dans des procédures rénales où il n'est pas simple de l'utiliser seule, et d'y joindre aussi souvent que nécessaire un complément d'information visuelle de meilleure qualité (CT). Le pipeline de l'approche proposé pour le recalage global automatique du rein dans les images US 3D et CT est illustré par la Figure 4.8, et comporte les étapes suivantes :

1. **Localisation et labellisation de points de repère anatomiques.** Grâce à deux réseaux de neurones profonds, cette étape reproduit l'approche manuelle de localisation développée dans le chapitre 3, en résolvant deux challenges. Premièrement, l'établissement de correspondance entre les points homologues. Nécessaire pour l'estimation de la transformation qui

aligne au mieux l'organe dans les deux modalités d'imagerie, cette correspondance est usuellement établie à l'aide de caractéristiques calculées sur les images ou sur les géométries. Le problème de localisation est défini ici comme un problème de segmentation sémantique multi-classes, attribuant ainsi des étiquettes aux points, celles-ci servant à l'appariement de ces derniers. Le second challenge est la ressemblance entre les régions locales des pôles supérieur et inférieur du rein dans la plupart des images US, rendant difficile la distinction de ces derniers. La stratégie à ce niveau est de ne pas distinguer les étiquettes des points symétriques, ce qui, tel qu'on le verra dans les expérimentations, améliore la capacité de détection. La résolution de l'ambiguïté, elle, est résolue à une étape ultérieure.

2. **Estimation d'hypothèses de transformation.** Par une adaptation de l'algorithme RANSAC, cette étape prend en compte la contrainte d'appariement des points détectés suivant leurs étiquettes, le nombre relativement faible de ces points permettant une exploration exhaustive (et non aléatoire) des possibilités, et la présence d'ambiguïté par symétrie. Cette adaptation est nommée "Labeled Exhaustive Sample Consensus" : LESAC (consensus sur échantillonnage exhaustif étiqueté).
3. **Résolution des ambiguïtés des hypothèses.** Par un réseau de neurones profond de segmentation sémantique, mais adapté pour classifier les paires multimodales d'images 3D de rein suivant que l'orientation de l'organe y est la même ou pas. Ce réseau est nommé FlipNet.

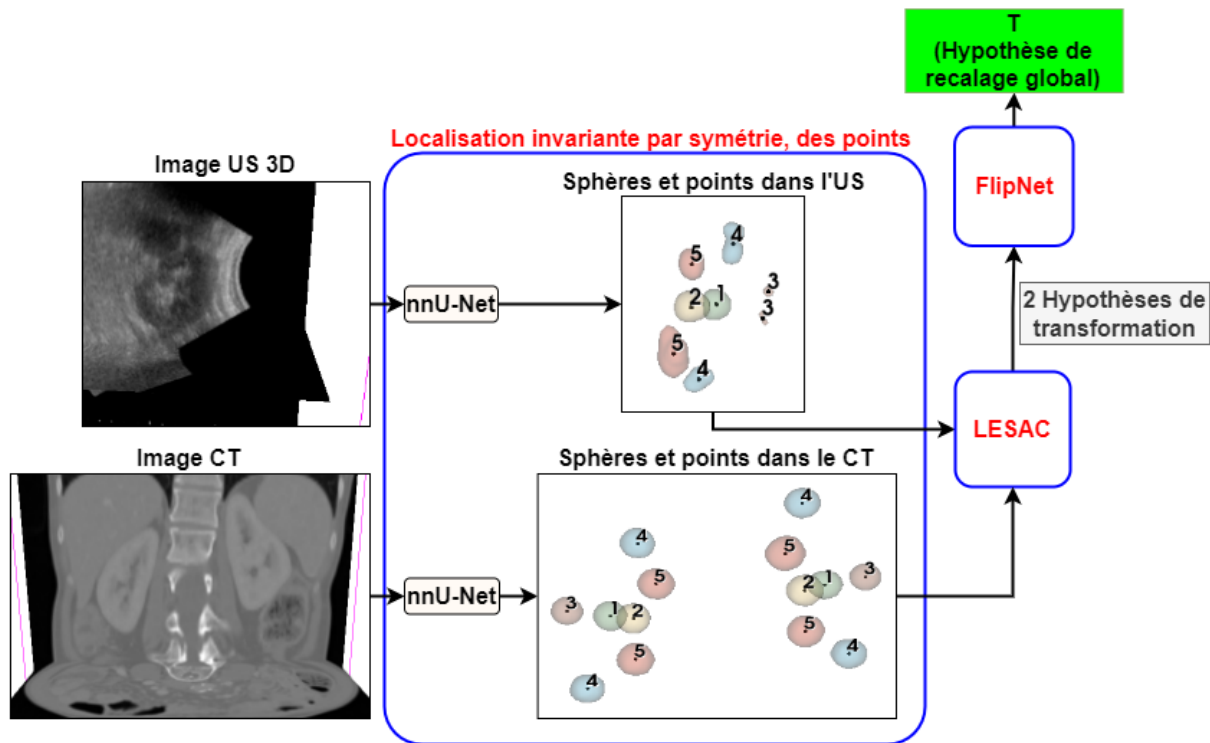


FIGURE 4.8 – Pipeline de la méthode proposée pour le recalage global automatique du rein dans les images US 3D et CT. Un module de localisation et labellisation des points par deux réseaux de neurones profonds; un module d’estimation d’hypothèses de transformations optimales mais ambiguës par symétrie, avec l’algorithme LESAC; et un module de sélection de l’hypothèse correcte par le classifieur FlipNet.

4.2.2 Les données

La méthode de recalage global proposée s’appuie sur le jeu de données TRUSTED décrit au chapitre 3. Les images US 3D, CT, et les annotations de 7 points de repère anatomiques coplanaires (FIGURE 4.9) de tous les reins annotés dans les deux modalités (59 reins) ont été utilisées pour former et évaluer cette nouvelle méthodologie de recalage global.

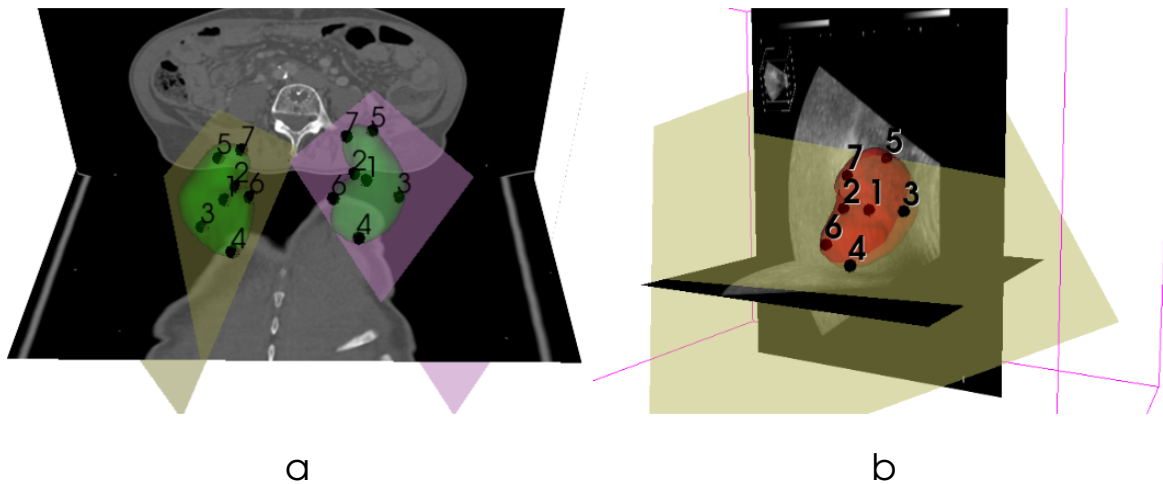


FIGURE 4.9 – Visualisations des données utilisées pour former la méthode proposée (issues du jeu de données TRUSTED). Vues 3D des 7 points de repère coplanaires du rein dans une image CT (a) et dans une image US 3D (b).

4.2.3 Localisation et étiquetage des points de repère

Deux DNNs ont été entraînés pour segmenter des régions locales entourant chaque point de repère. Cette représentation est similaire à divers modèles de régression retrouvés dans [133, 187, 67], mais au lieu de régresser des "cartes thermiques" ou des coordonnées spatiales, une région binaire (masque) est inférée par point de repère. Les valeurs de masque, nulles ou non nulles, représentent des voxels situés à l'intérieur ou à l'extérieur du voisinage local du point en utilisant un rayon sphérique de r mm (valeur par défaut $r=10$), valeur qui n'a pas été optimisée expérimentalement avec les données de test.

Deux difficultés principales ont été rencontrées pour la localisation des points de repère dans les images US. D'une part, la fenêtre limitée des images US 3D impliquait souvent une couverture incomplète de la région du rein. D'autre part, la symétrie pôle supérieur - pôle inférieur du rein (Figure 4.9) faisait de la différenciation des points de repère très similaires (4 vs. 5 et 6 vs. 7) un défi pour les DNNs. Deux stratégies ont été testées. La première dénommée **7L7L** ("7 landmarks 7 labels" en anglais, ou 7 points de repère 7 étiquettes) représente la stratégie dans laquelle chacun des 7 points de repère est étiqueté différemment. La charge de résoudre l'ambiguïté des pôles est laissée aux réseaux de localisation. La seconde dénommée **7L5L** ("7 landmarks 5 labels" en anglais, ou 7 points de repère 5 étiquettes) représente la stratégie dans laquelle le besoin de différenciation des pôles par les DNNs est écarté. Les étiquettes des points similaires ont été fusionnées, et il en est résulté 5 étiquettes sémantiques : étiquettes 1 à 3 pour les points 1 à 3 respectivement, étiquette 4 pour les points 4 et 5, et étiquettes 5 pour les points 6 à 7 (Figure 4.10).

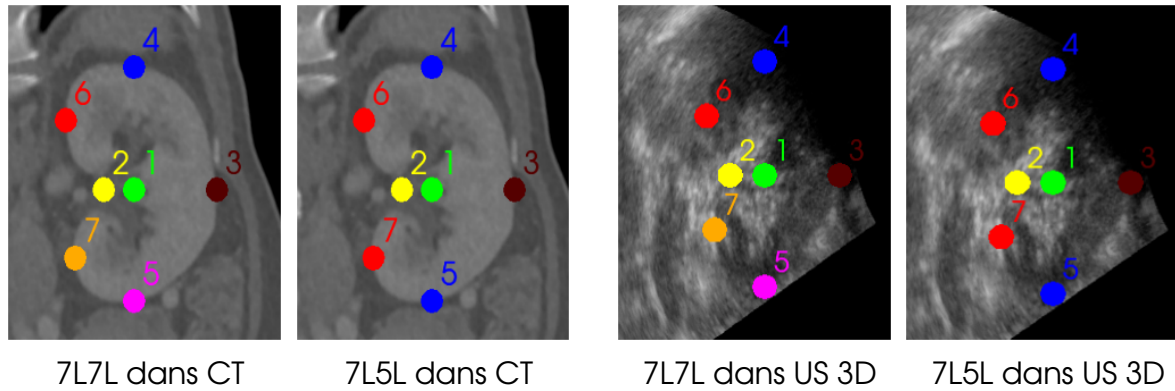


FIGURE 4.10 – Coupes sagittales 2D et étiquettes des points de repère de rein suivant la stratégie d’étiquetage spécifique par point (7L7L), et suivant la stratégie d’étiquettes fusionnées (7L5L).

N’importe quelle méthode de pointe fondée sur les DNNs peut être utilisée pour segmenter les régions des points de repère. La méthode nnU-Net [97] a été choisie en raison de ses bonnes performances générales pour la segmentation d’images médicales. Les DNNs ont été entraînés de manière supervisée, en utilisant l’auto-configuration "3d_fullres" et des paramètres standards de nnU-Net : une fonction de perte combinée Dice-entropie croisée, un taux d’apprentissage de 1% appliqué au gradient des poids à chaque itération, les normalisations d’intensité de type "CT norm" pour les images CT et de type "Zscore norm" pour les images US 3D. Après inférence, les positions des points sont estimées comme centroïdes de régions estimées et différenciées par une analyse de composantes connexes. A chaque point, l’étiquette de la région dont il est issu est attribuée. Cet étiquetage des points est ensuite utilisé pour appairer les points homologues dans les deux modalités.

Les ensembles de tous les points détectés dans le système de coordonnées de l’image US 3D (image mobile) et CT (image fixe) sont désignés par $\mathcal{P}$ et $\mathcal{Q}$, respectivement. Chacun de ces points possède une étiquette entière notée $L(\mathbf{p} \in \mathcal{P}) \in \mathbb{Z}$ et $L(\mathbf{q} \in \mathcal{Q}) \in \mathbb{Z}$. Dans la pratique, à une étiquette peut correspondre zéro, un ou plusieurs points dans $\mathcal{P}$ ou $\mathcal{Q}$ (Figures 4.11 et 4.12). En effet, dans les images US par exemple (Figure 4.12), un point peut ne pas être détecté en raison d’une ombre acoustique. De plus, les erreurs de faux positifs des DNNs peuvent entraîner des points parasites. Quoique ces erreurs soient beaucoup plus fréquentes avec les images US 3D, quelques rares cas ont été identifiés avec les images CT (Exemple 84, Figure 4.11). Par conséquent, la détermination de la transformation spatiale à partir des points détectés n’est pas triviale. Dans l’étape suivante, un algorithme simple, rapide et déterministe inspiré de RANSAC [106] est utilisé pour déterminer la transformation optimale de recalage qui gère

ces défis.

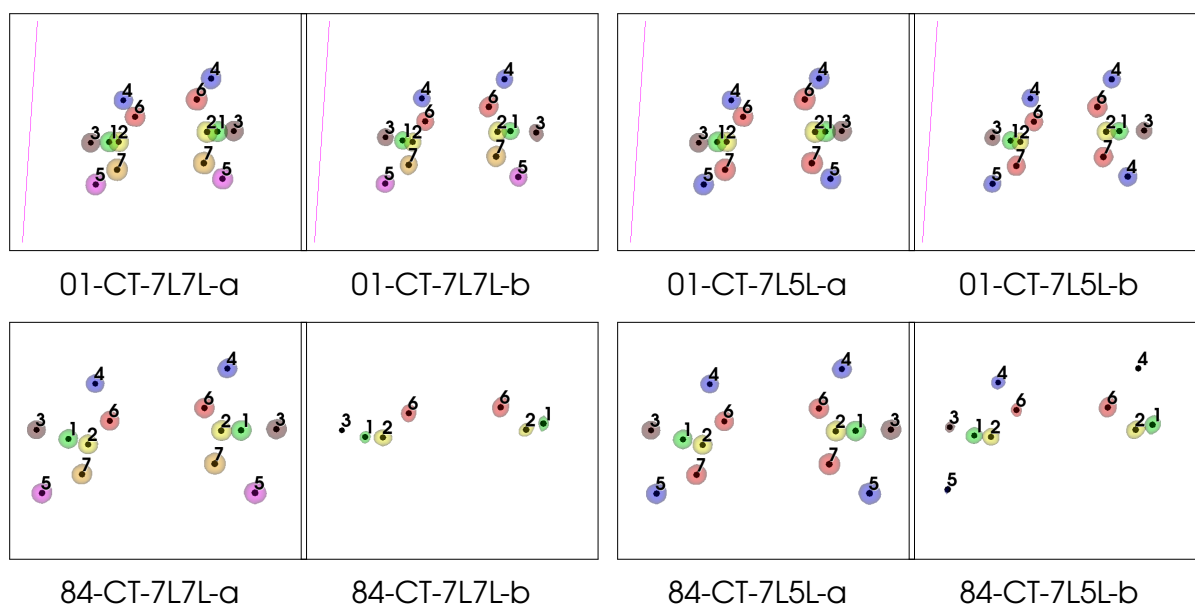


FIGURE 4.11 – Visualisations des points de repère rénaux détectés et étiquetés dans deux exemples CT (01 et 84). La nomenclature de chaque image est la suivante : "indice de l'individu-modalité de l'image-configuration de détection des points-x (x=a pour vérité terrain, x=b pour nnU-Net)".

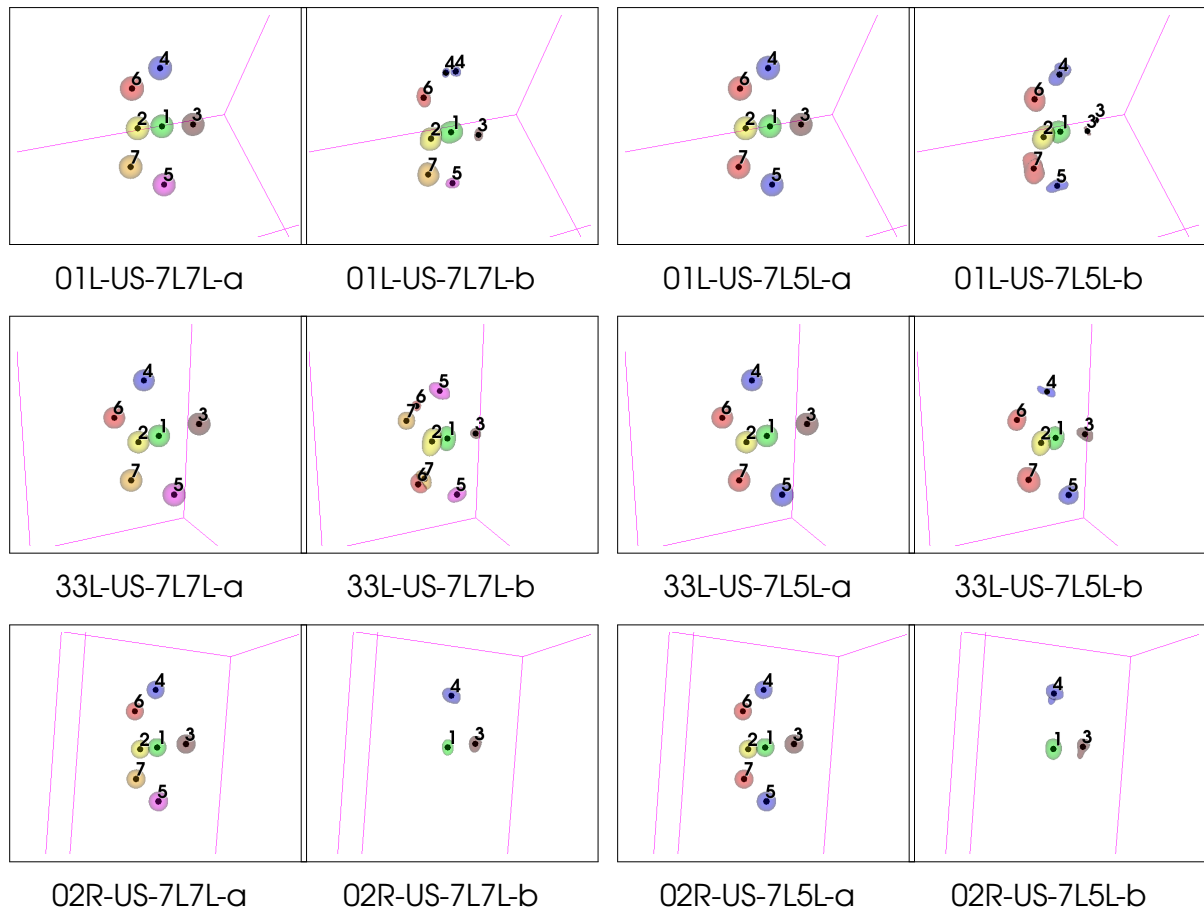


FIGURE 4.12 – Visualisations des points de repère rénaux détectés et étiquetés dans trois exemples US (01L, 33L et 02R). La nomenclature de chaque image est la suivante : "indice de l'individu-modalité de l'image-configuration de détection des points-x (x=a pour vérité terrain, x=b pour nnU-Net)".

4.2.4 Génération d'hypothèses de recalage global

RANSAC est une méthode bien établie pour estimer de manière robuste une transformation spatiale qui aligne deux ensembles de points pouvant contenir des valeurs aberrantes. Elle procède par échantillonnage aléatoire d'un sous-ensemble minimal de paires de points appariés, calcule une transformation spatiale associée à ce sous-ensemble, puis évalue le score de cette transformation en fonction du nombre de paires de points géométriquement compatibles avec la transformation. Le processus est répété de nombreuses fois, et la transformation avec le score le plus élevé est renvoyée.

Pour exécuter un algorithme de type RANSAC, il faut se donner une forme de transformation spatiale à estimer. L'objectif étant d'obtenir un recalage global pas nécessairement précis, mais qui peut être amélioré par un algorithme de recalage précis, la forme de transformation géométrique utilisée ici est une

similarité. En dimension 3, elle nécessite un minimum de $M=3$ paires de points pour être estimée ($M = 7$ degrés de liberté / 3 coordonnées), car ayant 7 degrés de liberté, dont 3 en translation, 3 rotations et 1 en mise à échelle. Toutefois, l'approche proposée peut être généralisée à d'autres formes de transformations linéaires. Une limitation de RANSAC classique est qu'il ne renvoie qu'une seule hypothèse de recalage, qui peut être incorrecte si les ensembles de points présentent une symétrie. Un moyen de réduire ce risque consiste à échantillonner de manière exhaustive les sous-ensembles minimaux, plutôt qu'aléatoirement, puis de renvoyer toutes les hypothèses avec des scores élevés. Puisque le nombre de paires de points manipulées est relativement faible (moins d'une dizaine, y compris les faux positifs) et que les paires valides sont contraintes par les étiquettes, l'échantillonnage exhaustif est rapidement exécutable. Ainsi, par exemple, avec 7 points détectés dans chacune des deux modalités, et sans contraintes d'étiquettes, l'espace de recherche comporte 18 424 transformations, alors qu'avec des contraintes d'étiquettes spécifiques sur chacun des 7 points, il est réduit à 35.

Définitions :

Une paire de points de repère US 3D - CT est dite valide si les deux points ont la même étiquette. Un échantillon de paires valides est dit valide s'il ne contient pas de répétition d'étiquettes. Un échantillon est minimal s'il contient un nombre minimal de paires nécessaires pour estimer la transformation cherchée. Enfin, une hypothèse est valide si elle est issue d'un échantillon minimal valide.

L'adaptation de RANSAC proposée et qui utilise l'échantillonnage exhaustif avec des contraintes d'étiquettes, est résumée par l'algorithme 1. Il a été baptisé LESAC, pour "Labeled Exhaustive Sample Consensus" ("consensus d'échantillonnage exhaustif étiqueté"). Il renvoie toutes les hypothèses valides de transformation $\mathcal{H}$, une pour chaque échantillon minimal valide constitué de paires valides, ainsi que leurs scores (au sens de RANSAC) associés $\mathcal{I}$ et les erreurs quadratiques moyennes (RMSE) $\mathcal{E}$.

À partir de l'ensemble d'hypothèses retourné, les meilleures hypothèses de recalage sont sélectionnées séquentiellement à l'aide d'une approche gloutonne. Tout d'abord, l'hypothèse $H_{i*} \in \mathcal{H}$ avec le score le plus élevé $s_{i*} = \max_{s \in \mathcal{S}}$ est sélectionnée. Ensuite, toutes les transformations dans $\mathcal{H}$ qui sont géométriquement "similaires" de H_{i*} sont éliminées, générant un ensemble d'hypothèses résiduelles $\mathcal{H}' \subset \mathcal{H}$. Une fois l'ensemble $\mathcal{H}'$ formé, la meilleure hypothèse suivante dans $\mathcal{H}'$ est sélectionnée, et le processus de sélection gloutonne se poursuit jusqu'à ce que plus aucune hypothèse ne puisse être sélectionnée ou qu'un

nombre maximum max_H d'hypothèses ait été sélectionné. Pour le rein, sachant qu'il a normalement une symétrie pôle supérieur - pôle inférieur, un critère d'arrêt anticipé de $max_H = 2$, a été imposé.

La similarité entre deux hypothèses de transformation (par exemple H_1 et H_2) telle que définie ci-dessus, a été évaluée en utilisant leurs composantes rotations (R_1 et R_2) de la manière suivante. D'abord, la rotation représentant la différence entre R_1 et R_2 est définie par : $R_d = R_2 \times \text{inv}(R_1)$. Ensuite, la valeur absolue θ de l'angle de rotation (en degrés) de R_d est calculée à l'aide de la formule de Rodrigues [188], qui convertit une matrice de rotation en représentation axe/angle. θ varie alors de 0 (lorsque $R_1 = R_2$) à 180 degrés. Les hypothèses de transformations H_1 et H_2 sont dites "similaires" si θ dépasse un seuil fixé (le seuil de 90 degrés a été utilisé par défaut).

Algorithm 1 LESAC : Génération d’hypothèses de recalage avec l’algorithme LESAC

Data: Ensembles des points fixes et mobiles $\mathcal{Q}$ et $\mathcal{P}$, ainsi que leurs étiquettes associées $L(\mathbf{q} \in \mathcal{Q}) \in \mathbb{Z}$ et $L(\mathbf{p} \in \mathcal{P}) \in \mathbb{Z}$

Result: L’ensemble des hypothèses de recalage $\mathcal{H}$, l’ensemble des nombres d’appariements cohérents associées $\mathcal{I}$ et l’ensemble des erreurs moyennes quadratiques associées (RMSEs) $\mathcal{E}$

$\mathcal{H} \leftarrow \emptyset, \mathcal{I} \leftarrow \emptyset, \mathcal{E} \leftarrow \emptyset$

```

forall  $\mathbf{p}_1, \dots, \mathbf{p}_M, \mathbf{q}_1, \dots, \mathbf{q}_M$  such that  $L(\mathbf{p}_1) = L(\mathbf{q}_1), \dots, L(\mathbf{p}_M) = L(\mathbf{q}_M)$  do
    Estimation de la transformation optimale  $\mathbf{H}$  alignant  $(\mathbf{p}_1, \dots, \mathbf{p}_M)$  à  $(\mathbf{q}_1, \dots, \mathbf{q}_M)$ 
     $s \leftarrow 0$  ; // Nombre d’appariements cohérents associés à  $\mathbf{H}$ 
    forall  $\mathbf{p} \in \mathcal{P}$  do
         $\mathbf{p}' \leftarrow \mathbf{H}(\mathbf{p})$ 
         $\mathcal{S} \leftarrow \emptyset$  ; // Ensemble d’appariements cohérents
         $\mathcal{Q}' = \{\mathbf{q} \in \mathcal{Q} \mid L(\mathbf{q}) = L(\mathbf{p})\}$ 
         $\mathbf{q} = \arg \min_{\mathbf{q}_i \in \mathcal{Q}'} \|\mathbf{q}_i - \mathbf{p}'\|$  ; // Appariement au plus proche voisin
        if  $\|\mathbf{q} - \mathbf{p}'\| < \tau$  then
             $\mathcal{S} \leftarrow \mathcal{S} \cup \{(\mathbf{p}, \mathbf{q})\}$  ; // Ajout de la paire à l’ensemble des
                appariements cohérents
             $s \leftarrow s + 1$ 
        end
    end
    if  $s \geq M$  then
        Estimation de la transformation optimale  $\mathbf{H}'$  et de l’erreur moyenne qua-
            dratique associée (RMSE)  $e$  en utilisant les appariements cohérents  $\mathcal{S}$ 
         $\mathcal{H} \leftarrow \mathcal{H} \cup \{\mathbf{H}'\}$ 
         $\mathcal{I} \leftarrow \mathcal{I} \cup \{s\}$ 
         $\mathcal{E} \leftarrow \mathcal{E} \cup \{e\}$ 
    end
end
return  $\mathcal{H}, \mathcal{I}, \mathcal{E}$ 

```

4.2.5 Résolution des ambiguïtés des hypothèses de recalage

Un module de résolution des ambiguïtés des hypothèses a été construit. Il est constitué d'un réseau de neurones entraîné sous forme de classifieur pour indiquer si les reins dans deux images hypothétiquement globalement recalées, ont la même orientation ou pas. Ce classifieur a été appelé FlipNet (Figure 4.13). Il s'agit d'un réseau à convolution (CNN) prenant en entrée une image 3D à deux canaux : l'image CT et l'image US 3D transformée suivant le recalage estimé et re-échantillonnée de manière à avoir la même résolution et le même espacement (distance entre deux centres de voxels adjacents) de voxels que l'image CT. Toutes les images CT et US 3D ont été au préalable redimensionnées en taille $256 \times 256 \times 256$. Le réseau retourne un masque volumique de même dimension que les images, dont le nombre de voxels non-nuls d'un canal spécifique indique le nombre de "votes" en faveur d'une orientation correcte du rein dans l'image US 3D (*i.e.* orientation similaire à celle du rein dans le CT). Le nombre de voxels nuls de ce canal, lui, indique le nombre de votes en faveur d'un retournement des pôles supérieur et inférieur dans l'image US 3D transformée.

FlipNet résolvant une tâche de classification d'images 3D, toute architecture de classification de pointe pourrait être envisagée, comme EfficientNet [189]. Toutefois, il a été constaté qu'il était difficile d'entraîner de tels classificateurs à partir de zéro pour cette tâche, probablement en raison du nombre limité de données d'entraînement. L'objectif de classification a été modifié de sorte que, au lieu de prédire directement en sortie une classe globale ("correct", "retourné") par image 3D à deux canaux, FlipNet prédisait une carte d'étiquettes (masque de segmentation) volumique de même dimension que chacun des canaux de l'image d'entrée, avec un canal spécifique pour exprimer une sorte de vote en faveur ou en défaveur d'une orientation correcte. Ainsi un CNN initialement dédié à la segmentation a été utilisé à la place d'un classifieur standard, et le framework choisi pour cela a été nnU-Net une fois de plus.

Après l'inférence par FlipNet, le score de classification correspondant à une hypothèse de transformation H_{i*} est déterminé comme étant le maximum entre le nombre de voxels en faveur d'une orientation "correcte" dans la carte d'étiquettes en sortie correspondante et le nombre de voxels en faveur d'un retournement dans la carte d'étiquettes en sortie correspondant à l'autre hypothèse (H_{-i*}).

Les données d'entraînement de Flipnet sont construites comme suit. Pour chaque image CT (redimensionnée en taille $256 \times 256 \times 256$), l'hypothèse correcte est générée (*i.e.* une image 3D à 2 canaux), en lui appliquant la transformation

de similarité des moindres carrés calculée à partir des points de repère vérité terrain (manuels) dans les images US 3D et CT. L'hypothèse "retournée" est ensuite générée en échangeant les points de repère CT auto-similaires (4 avec 5, et 6 avec 7) et en appliquant la transformation de similarité des moindres carrés associée. Les deux hypothèses sont ensuite associées à des cartes d'étiquettes de segmentation sémantique dont un canal code les classes "correct" ou "retourné".

Deux types de cartes d'étiquettes de sortie ont été expérimentés pour la formation de FlipNet.

— **FlipNetA : apprentissage multitâche.**

Cette version de FlipNet est formée pour résoudre conjointement la classification telle que décrite ci-dessus et la localisation de points de repère ; une forme d'apprentissage multitâche [190]. Durant la phase d'entraînement, le réseau apprend des caractéristiques pertinentes pour ces tâches connexes, ce qui peut améliorer la généralisation. Dans cette configuration, les cartes d'étiquettes comportent 6 canaux numérotés de 1 à 6. Les canaux 1 à 5 présentent une carte de même type que celle utilisée pour la localisation des points de repère version 7L5L dans les CT. Ces 5 canaux codent la tâche de localisation. Le sixième canal code la tâche de classification. Ses voxels non-nuls sont en faveur de la classe "correct" tandis que ses voxels nuls sont en faveur de la classe "retourné". En conséquence, avec ce codage de carte d'étiquettes, FlipNet est entraîné à effectuer les deux tâches de classification et de localisation de points de repère. Il peut être entraîné à l'aide d'une architecture de segmentation existante. Les résultats avec nnU-Net sont présentés, en utilisant les mêmes paramètres d'entraînement que dans la section 4.2.

— **FlipNetB : masquage binaire.**

Cette version de FlipNet est similaire à FlipNetA, mais les cartes d'étiquettes ont uniquement le canal dédié à la classification.

En inférence, les entrées sont formées comme à l'entraînement, mais avec les hypothèses de transformation de similarité fournies par LESAC.

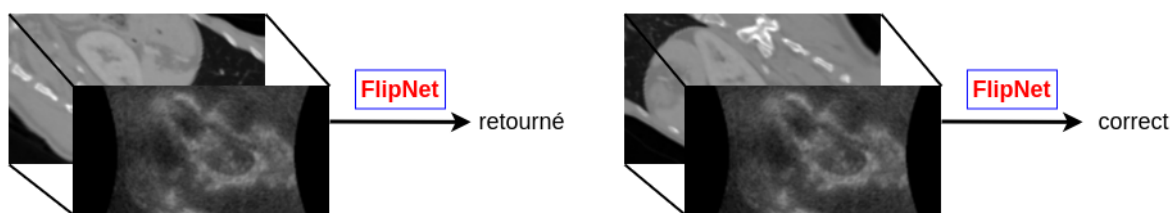


FIGURE 4.13 – Illustration du fonctionnement de FlipNet. En entrée, il prend une image 3D à deux canaux (l'image CT et l'image US 3D recalée suivant une hypothèse de recalage global issue de LESAC). En sortie, il retourne une carte d'étiquettes permettant de dire si l'hypothèse correspond à un rein correctement orienté ou un rein retourné.

4.2.6 Expériences

4.2.6.1 Configuration des expérimentations

Pour rappel, le recalage global est abordé dans ce chapitre comme méthode d'initialisation d'un recalage précis. Toutes les performances sont rapportées après affinement du recalage par la méthode BCPD, comme dans la section 4.1. Les différentes configurations de la méthode proposée sont : **(7L7L, LESAC)**, **(7L5L, LESAC, FlipNetA)** et **(7L5L, LESAC, FlipNetB)**. Chaque terme est noté tel que dans la description de la méthode en section 4.2 et indique une étape de l'approche de recalage globale proposée, la principale parmi ces approches étant **(7L5L, LESAC, FlipNetA)**. Le jeu de données utilisé reste TRUSTED et les blocs de validation croisée sont les mêmes que dans la section 4.1. Après estimation du recalage global, les masques de segmentation automatique produits par nnU-Net dans la section 4.1 ont été utilisés avec la méthode BCPD pour l'affinement du recalage, puis les évaluations ont été exécutées.

Les métriques d'évaluation sont les mêmes que dans la section 4.1, et leurs significations sont les mêmes.

- La distance de surface moyenne des plus proches voisins (**nndist**) (en mm) quantifie la superposition des surfaces après recalage
- TRE-1 (en mm) est sensible à l'erreur sur la cible de recalage qu'est le pelvis rénal
- TRE-4,6 et TRE-5,7 (en mm) sont sensibles à la qualité de l'orientation "pôle supérieur - pôle inférieur" du rein dans l'image US 3D après recalage
- TRE-2 et TRE-3 (en mm) sont sensibles à la qualité de l'orientation transversale du rein dans l'image US 3D après recalage

La principale configuration de l'approche proposée **(7L5L, LESAC, FlipNetA)** a été comparée à une série de méthodes de recalage global de référence

et concurrentes. Toutes ces méthodes de référence ont été configurées pour estimer une transformation de similarité elle-même utilisée comme initialisation de la méthode d’affinement BCPD. Les méthodes de référence sont :

- **(PCA, nndist)**. Seconde meilleure performance de recalage global utilisant les surfaces, parmi les méthodes évaluées dans la section 4.1.
- **(DIP, TEASER)**. Meilleure performance de recalage global utilisant les surfaces, parmi les méthodes évaluées dans la section 4.1.
- **(7L7L, RANSAC)**. Configuration dans laquelle le réseau de localisation des points de repère nnU-Net était entraîné pour localiser 7 points de repère en utilisant 7 étiquettes différentes. Puis, RANSAC était utilisé pour calculer le recalage à partir des paires de points de repère, appariés grâce à leurs étiquettes.
- **(Manuelle)**. La méthode utilisant les points de repère vérité terrain.

En plus des méthodes de référence ci-dessus, 7 configurations ablatives, incluant les autres configurations de l’approche proposée, ont été comparées afin d’évaluer l’impact de chaque aspect clé. Ces configurations sont les suivantes :

1. **(7L5L, LESAC, FlipNetA)**, principale configuration de l’approche proposée.
2. **(7L5L, LESAC, FlipNetB)**, similaire à la configuration 1, avec FlipNetB utilisé à la place de FlipNetA.
3. **(7L5L, LESAC, nndist)**, similaire à la configuration 1, avec FlipNetA remplacé par un critère géométrique pour la sélection de la meilleure hypothèse. Il s’agit de la distance du plus proche voisin entre le rein segmenté automatiquement dans les images US 3D et CT (comme dans **(PCA, nndist)**).
4. **(7L5L, RANSAC)**, similaire à la configuration 1, avec RANSAC utilisé pour estimer le recalage à partir des points de repère détectés avec 7L5L.
5. **(7L7L-2, LESAC, FlipNetA)**, similaire à la configuration 1, sauf que les deux hypothèses proposées par LESAC ont été générées à partir de points de repère comportant 7 étiquettes (pas de fusion d’étiquettes). Cette ablation a permis de comparer l’effet de la fusion d’étiquettes.
6. **(7L7L, LESAC)**, similaire à **(7L7L-2, LESAC, FlipNetA)**, sauf que LESAC a directement été utilisé pour estimer le recalage (hypothèse avec le score le plus élevé).
7. **(PCA, FlipNetA)**, similaire à **(PCA, nndist)**, avec FlipNetA utilisé au lieu de nndist, pour sélectionner l’hypothèse produite par PCA. Cette ablation a permis de comparer l’effet de la génération d’hypothèses à l’aide de notre approche se servant de points de repère, par rapport à l’ACP, en utilisant la même méthode pour la sélection finale des hypothèses (FlipNetA).

8. **(PCA, FlipNetB)**, similaire à **(PCA, nndist)**, avec FlipNetB utilisé pour sélectionner l'hypothèse produite par PCA.

4.2.6.2 Résultats quantitatifs

Les résultats quantitatifs des performances des méthodes de recalage global combinées avec l'affinement BCPD, sont présentés par les graphiques de la Figure 4.14, et les résumés statistiques dans le Tableau 4.3.

Les graphiques de la Figure 4.14 montrent la distribution des métriques de recalage (nndist, TRE-1, TRE-4 et TRE-5) pour chacune des principales méthodes. Les axes des ordonnées représentent différents seuils de métriques et les axes des abscisses indiquent le nombre de reins pour lesquels la métrique évaluée était inférieure au seuil de métrique. Nous constatons que les courbes de **(7L5L, LESAC, FlipNetA)** sont généralement plus proches de la courbe de **Manuelle** (la référence) que celles des autres méthodes, en particulier vers le côté droit des courbes.

Le tableau 4.3 montre les résultats quantitatifs des performances des méthodes de recalage global combinées avec l'affinement BCPD. Les colonnes présentent les métriques de performance pour chaque configuration par rapport à la valeur moyenne (Avg), l'écart-type (Std) et le troisième quartile (q3). Les valeurs de troisième quartile ont été incluses, car elles peuvent clairement révéler l'existence de pics dans les distributions de performance. Les résultats ont été calculés en utilisant les moyennes par blocs (en contexte de validation-croisée à 5 blocs). Les caractères gras indiquent la meilleure valeur de métrique. Les p-valeurs ont été calculées par métrique, par rapport à chaque meilleure moyenne, à l'aide du test unilatéral des rangs de Wilcoxon. Les p-valeurs inférieures à 0.05 indiquent que les moyennes considérées sont significativement différentes de la meilleure moyenne.

Si l'on considère la métrique nndist, la meilleure performance moyenne a été obtenue avec **(7L5L, LESAC, FlipNetA)** (2.42 mm), qui était statistiquement meilleure que toutes les autres méthodes principales, à l'exception de **(7L7L, RANSAC)** (2.48 mm, $p=0.33$). Elle présentait également l'écart type inter-blocs le plus faible, ainsi que la meilleure performance q3. S'agissant de la métrique TRE-1 (associé au centre du pelvis rénal), parmi les principales méthodes, la meilleure performance moyenne a été obtenue avec **(7L5L, LESAC, FlipNetA)** (5.78 mm), qui a également obtenu la meilleure performance q3 (6.19 mm). Dans les études ablatives, l'erreur moyenne TRE-1 la plus faible a été obtenue par **(7L5L, LESAC, nndist)** (5.49 mm), mais elle n'était pas significativement meilleure que celle de **(7L5L, LESAC, FlipNetB)**, **(7L5L, RANSAC)** et **(7L5L, LESAC, FlipNetA)**.

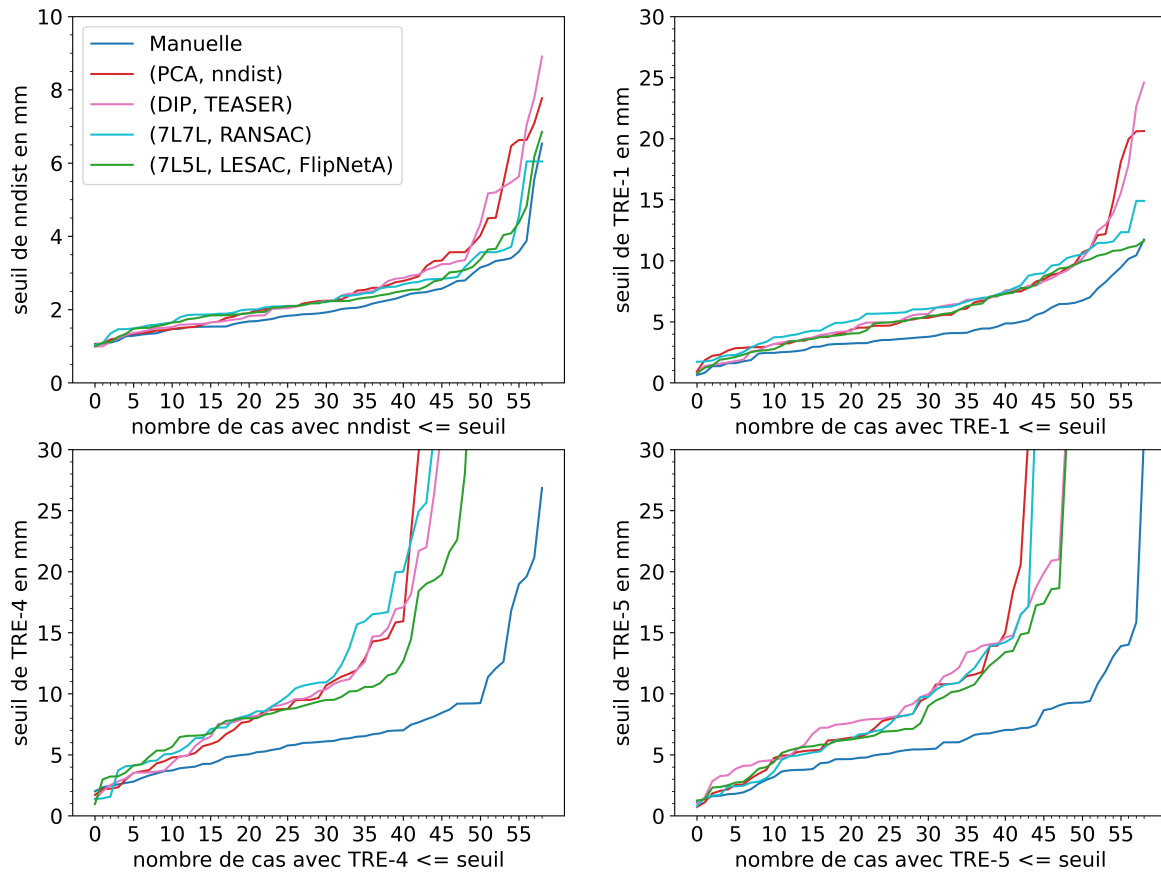


FIGURE 4.14 – Performances des méthodes automatiques de recalage global, combinées à l’affinement par BCPD. Les distributions de distance du plus proche voisin (nndist) et de TRE-1 sont présentées sur la première ligne. Les distributions de TRE-4 et de TRE-5 sont présentées sur la dernière ligne. Les axes des ordonnées représentent différents seuils de métriques et les axes des abscisses représentent le nombre de reins pour lesquels la métrique évaluée après recalage était inférieure au seuil de métrique.

La meilleure performance q3 TRE-1 parmi les principales méthodes et études ablatives a été obtenue avec **(7L5L, LESAC, FlipNetA)** (6.19 mm).

En ce qui concerne TRE-2, la meilleure performance moyenne parmi les principales méthodes a été obtenue par **(7L5L, LESAC, FlipNetA)** (8.09 mm), marginalement derrière **(7L5L, LESAC, FlipNetB)** (7.85) mm. Les différences entre **(7L5L, LESAC, FlipNetA)**, **(7L5L, LESAC, nndist)**, **(7L5L, RANSAC)** et **(7L5L, LESAC, FlipNetB)** n’étaient pas statistiquement significatives. Pour TRE-3, TRE-4, TRE-5, TRE-6 et TRE-7, la meilleure performance moyenne a été obtenue avec **(7L5L, LESAC, FlipNetA)** (11.26, 28.03, 25.24, 19.67 et 18.7 mm respectivement).

En résumé, bien qu’aucune méthode du tableau 4.3 ne soit la plus performante pour toutes les métriques, la principale configuration de l’approche proposée **(7L5L, LESAC, FlipNetA)** arrive en tête pour presque toutes les valeurs moyennes et est suivie de près par une autre configuration de l’approche

proposée (**7L7L-2, LESAC, FlipNetA**). La configuration (**7L5L, LESAC, nndist**) est également compétitive. En la comparant à (**7L7L-2, LESAC, FlipNetA**), des performances significativement moins bonnes sont observées de sa part, du point de vue de TRE-4, TRE-5, TRE-6 et TRE-7, mais pas du point de vue de TRE-1, TRE-2, TRE-3 et nndist. Cela peut s'expliquer par le fait que les points de repère 4 à 7 sont éloignés du plan de symétrie pôle supérieur - pôle inférieur du rein, contrairement aux points de repère 1 à 3, ce qui fait que les métriques TRE-4 à TRE-7 sont beaucoup plus sensibles aux recalages qui retournent incorrectement le rein. En outre, le critère de sélection d'hypothèses de recalage nndist requiert une segmentation automatique de l'organe dans les deux modalités, ce qui n'est pas le cas de FlipNetA et FlipNetB. (**7L7L, RANSAC**) présente des performances très similaires à celles de (**7L7L, LESAC**), comme prévu. Cependant, (**7L7L, RANSAC**) produit généralement plus de valeurs TRE extrêmes par rapport à (**7L5L, LESAC, FlipNetA**), comme le montre la Figure 4.14.

Résultats quantitatifs sans affinement du recalage

Ce paragraphe présente, pour chaque configuration d'initialisation, la contribution de l'affinement BCPD dans l'ensemble du processus de recalage précis (initialisation et affinement). A cet effet, le Tableau 4.4 donne les résultats de performances isolées, des méthodes de recalage global automatique. La structure du Tableau 4.4 est identique à celle du Tableau 4.3. L'observation des deux tableaux montre que l'affinement a un impact globalement positif sur l'ensemble du processus de recalage précis (initialisation et affinement), pour toutes les méthodes étudiées. De l'initialisation au recalage précis, chaque méthode voit ses performances améliorées ou maintenues pour la majorité des métriques, avec des réductions notables des erreurs moyennes. Par exemple, pour la méthode (7L5L, LESAC, FlipNetA), l'affinement réduit la moyenne de la métrique nndist de 3,80 mm à 2,42 mm, celle de TRE-1 de 6,83 mm à 5,78 mm, et celle de TRE-4 de 31,58 mm à 28,03 mm. De même, pour les métriques TRE-6 et TRE-7, l'affinement réduit respectivement les moyennes de 20,86 mm à 19,67 mm et de 21,75 mm à 18,70 mm, avec une diminution des écarts types, ce qui montre une meilleure stabilité et une meilleure robustesse.

Méthodes	nndist↓ (mm)				TRE-1 ↓ (mm)				TRE-4 ↓ (mm)				TRE-5 ↓ (mm)			
	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3
Manuelle	2.16	0.4	-	2.16	4.34	0.53	-	4.45	7.1	2.35	-	7.86	6.4	1.54	-	7.51
(PCA, nndist)	2.7	0.57	0.022	2.82	6.57	1.66	0.022	8	37.92	18.53	0.112	52.06	33.58	15.46	0.112	47.6
(DIP, TEASER)	2.7	0.69	0.040	3.02	6.61	1.73	0.022	7.96	29.05	10.76	0.446	32.73	25.93	8.4	0.554	34.34
(7L7L, RANSAC)	2.48	0.45	0.33	2.54	6.53	1.24	0.040	7.23	34.84	9.86	0.173	40.17	32.76	10.36	0.173	36.71
(7L5L, LESAC, FlipNetA)	2.42	0.41	-	2.52	5.78	0.9	0.250	6.19	28.03	7.96	-	35.2	25.24	6.13	-	31.31
(PCA, FlipNetB)	2.85	0.57	0.022	3.16	6.33	1.83	0.040	7.35	53.74	19.86	0.022	51.61	49.66	18.51	0.022	47.53
(PCA, FlipNetA)	2.73	0.59	0.022	2.82	6.57	1.7	0.040	7.18	40.29	16.47	0.040	51.89	36.43	14.94	0.069	50.33
(7L7L, LESAC)	2.73	0.72	0.069	2.85	6.52	1.28	0.040	7.34	34.76	9.97	0.173	40.43	32.85	10.5	0.173	37.03
(7L7L-2, LESAC, FlipNetA)	2.71	0.59	0.069	2.84	6.11	1.16	0.040	6.97	29.29	10.12	0.657	33.54	27.93	9.19	0.446	30.74
(7L5L, RANSAC)	2.62	0.41	0.022	2.72	5.68	1.09	0.250	6.34	46.11	13.61	0.022	51.6	41.57	11.69	0.022	50.97
(7L5L, LESAC, nndist)	2.51	0.48	0.112	2.82	5.49	1.51	-	6.98	41.87	12.43	0.022	52.2	38.74	10.06	0.022	46.41
(7L5L, LESAC, FlipNetB)	2.48	0.4	0.112	2.66	5.52	1.06	0.446	6.35	30.76	9.76	0.446	36.9	27.17	9.68	0.446	34.01

Méthodes	TRE-2 ↓ (mm)				TRE-3 ↓ (mm)				TRE-6 ↓ (mm)				TRE-7 ↓ (mm)			
	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3
Manuelle	5.94	1.36	-	6.9	5.13	0.48	-	4.98	8.78	1.97	-	9.79	6.27	1.15	-	6.68
(PCA, nndist)	12.09	3.33	0.022	15.04	17.54	5.96	0.022	22.84	29.87	12.37	0.040	36.38	26.59	9.56	0.040	36.42
(DIP, TEASER)	11.14	3.29	0.040	13.11	17.65	3.54	0.022	17.11	23.5	5.07	0.040	29.08	20.4	6.31	0.33	21.98
(7L7L, RANSAC)	8.55	1.97	0.069	9.37	13.53	1.8	0.069	14.72	24.57	5.18	0.069	29.94	24.44	6.41	0.112	29.54
(7L5L, LESAC, FlipNetA)	8.09	1.98	0.112	9.47	11.26	2.46	-	14.08	19.67	4.87	-	23.2	18.7	3.42	-	21.89
(PCA, FlipNetB)	12.88	4.17	0.022	15.1	20.35	5.91	0.022	24.05	38.66	14.88	0.022	37.16	35.86	11.82	0.022	33.44
(PCA, FlipNetA)	12.88	4.37	0.022	16.58	19.6	7.42	0.022	27.92	31.49	13.82	0.040	41.2	28.71	10.89	0.022	41.5
(7L7L, LESAC)	8.78	2.1	0.040	9.11	13.27	1.56	0.112	13.32	24.7	5.13	0.069	30.07	24.51	6.45	0.112	29.23
(7L7L-2, LESAC, FlipNetA)	8.88	1.99	0.039	9.38	13.15	1.06	0.069	13.52	22	5.92	0.33	26.16	20.96	6.7	0.250	23.3
(7L5L, RANSAC)	8.12	2.1	0.173	9.17	13.01	2.31	0.040	15.64	28.35	7.21	0.022	31.48	27.87	4.95	0.022	32.17
(7L5L, LESAC, nndist)	8.31	1.93	0.173	9.61	11.88	2.07	0.112	12.62	27.76	7.8	0.022	36.29	26.43	6.52	0.022	29.63
(7L5L, LESAC, FlipNetB)	7.85	1.9	-	9.46	11.81	2.31	0.173	13.73	20.95	4.44	0.446	23.18	20.33	5.21	0.250	24.57

TABLE 4.3 – Comparaison quantitative des méthodes de recalage global combinées avec l’affinement BCPD. Le symbole (↓) indique que les faibles valeurs sont les meilleures. Les résultats sont regroupés en deux blocs de métriques, le premier bloc contenant les métriques nndist, TRE-1, TRE-4, TRE-5, et deuxième bloc contenant les métriques TRE-2, TRE-3, TRE-6 et TRE-7. Dans chaque bloc, les méthodes sont divisées en trois groupes : Premièrement, les résultats de la méthode manuelle, utilisant les points de repère manuels ; c’est la référence utilisée en comparaison avec les méthodes automatiques. Deuxièmement, les résultats des principales méthodes comparées. Troisièmement, les résultats d’étude ablativité. Les colonnes présentent respectivement la moyenne (Avg), l’écart-type (Std), la p-valeur de la moyenne, le troisième quartile (q3) entre les blocs de validation croisée, des métriques et pour chaque méthode. Les valeurs en gras indiquent la meilleure valeur de la mesure. Les p-valeurs ("p-value" en anglais) ont été calculées par métrique, par rapport à chaque meilleure moyenne, à l’aide du test unilatéral des rangs de Wilcoxon.

Méthodes	nndist↓ (mm)				TRE-1 ↓ (mm)				TRE-4 ↓ (mm)				TRE-5 ↓ (mm)			
	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3
Manuelle	2.36	0.28	-	2.65	2.28	0.30	-	2.37	3.88	0.66	-	3.95	2.77	0.31	-	2.93
(PCA, nndist)	3.21	0.30	-	3.44	6.34	0.069	0.554	7.15	37.34	16.85	0.112	51.56	34.13	14.71	0.250	46.90
(DIP, TEASER)	3.94	1.17	0.112	4.50	10.05	1.94	0.021	11.26	33.98	10.98	0.343	37.32	28.45	7.78	0.657	36.99
(7L7L, RANSAC)	4.08	0.72	0.039	4.35	53.71	37.62	0.021	83.13	77.04	42.40	0.069	119.56	76.39	41.26	0.069	111.90
(7L5L, LESAC, FlipNetA)	3.80	0.40	0.021	3.89	6.83	1.07	0.039	7.45	31.58	6.46	-	37.46	29.40	5.51	-	33.39
(PCA, FlipNetB)	3.35	0.34	0.021	3.63	6.12	1.32	-	6.56	51.78	17.71	0.021	51.26	49.66	18.51	0.021	47.53
(PCA, FlipNetA)	3.26	0.29	0.069	3.44	6.39	1.25	0.112	6.55	38.83	15.22	0.342	50.82	36.43	14.94	0.172	50.33
(7L7L, LESAC)	4.03	0.74	0.039	4.21	7.66	1.00	0.021	8.48	37.00	9.75	0.172	41.01	36.09	9.55	0.172	39.95
(7L7L-2, LESAC, FlipNetA)	3.90	0.69	0.069	4.22	7.58	1.08	0.022	8.10	32.29	9.66	0.657	33.55	30.76	9.16	0.750	32.86
(7L5L, RANSAC)	3.92	0.52	0.039	4.39	6.93	1.40	0.172	8.04	49.26	11.87	0.021	56.53	45.68	10.06	0.021	55.52
(7L5L, LESAC, nndist)	3.62	0.28	0.021	3.78	6.42	1.28	-	7.50	43.65	10.35	0.021	48.79	41.26	8.50	0.022	47.40
(7L5L, LESAC, FlipNetB)	3.91	0.48	0.021	4.28	6.60	1.06	0.112	7.50	33.47	9.21	0.553	37.83	31.70	9.02	0.343	39.98

Méthodes	TRE-2 ↓ (mm)				TRE-3 ↓ (mm)				TRE-6 ↓ (mm)				TRE-7 ↓ (mm)			
	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3	Avg	Std	p-value	q3
Manuelle	3.46	0.68	-	3.86	3.11	0.32	-	3.37	5.39	0.87	-	5.24	4.12	0.55	-	4.35
(PCA, nndist)	12.32	2.92	0.021	15.01	17.90	5.52	0.069	23.06	29.85	11.32	0.069	37.86	27.70	8.46	0.039	35.12
(DIP, TEASER)	13.73	3.63	0.021	17.57	19.11	3.01	0.021	18.64	26.34	6.50	0.039	32.89	22.46	6.25	0.343	23.74
(7L7L, RANSAC)	55.63	38.78	0.021	88.09	61.02	36.60	0.021	84.71	69.45	39.99	0.039	109.96	69.69	40.31	0.039	103.15
(7L5L, LESAC, FlipNetA)	8.85	1.59	0.069	10.01	13.42	2.77	-	16.61	20.86	3.10	-	23.31	21.75	3.66	-	24.97
(PCA, FlipNetB)	13.04	3.40	0.021	15.76	20.95	5.02	0.021	24.67	37.81	13.24	0.021	37.13	36.62	10.51	0.021	34.64
(PCA, FlipNetA)	13.18	3.60	0.021	16.81	20.08	6.87	0.021	27.97	31.21	13.15	0.069	42.22	29.80	9.33	0.039	40.01
(7L7L, LESAC)	9.24	1.77	0.172	10.28	17.02	2.95	0.112	19.12	25.54	4.60	0.069	28.12	26.71	6.05	0.069	31.08
(7L7L-2, LESAC, FlipNetA)	9.44	1.64	0.112	11.02	16.80	1.95	0.112	18.27	23.03	3.98	0.069	23.74	23.26	6.77	0.553	26.50
(7L5L, RANSAC)	9.31	2.09	0.112	10.15	15.02	2.89	0.069	16.86	29.37	5.97	0.021	34.48	31.33	3.82	0.021	34.88
(7L5L, LESAC, nndist)	8.41	1.25	0.657	9.00	14.32	2.07	0.112	15.35	28.31	5.53	0.021	32.40	28.83	5.57	0.021	32.42
(7L5L, LESAC, FlipNetB)	8.56	1.47	-	9.56	13.52	2.29	0.657	15.73	21.86	4.10	0.446	23.18	23.21	5.51	0.343	27.32

TABLE 4.4 – Comparaison quantitative des méthodes de recalage global **sans affinement**. Le symbole (↓) indique que les faibles valeurs sont les meilleures. Les résultats sont regroupés en deux blocs de métriques, le premier bloc contenant les métriques nndist, TRE-1, TRE-4, TRE-5, et deuxième bloc contenant les métriques TRE-2, TRE-3, TRE-6 et TRE-7. Les colonnes présentent respectivement la moyenne (Avg), l'écart-type (Std), la p-valeur de la moyenne comparée à la meilleure moyenne, le troisième quartile (q3) entre les blocs de validation croisée, des métriques et pour chaque méthode.

4.2.6.3 Résultats qualitatifs

Les figures 4.15 et 4.16 montrent chacune un résultat avec les méthodes principales de recalage global comparées après affinement avec BCPD. Il s'agit de : (1)-(Manuelle), (2)-(PCA, nndist), (3)-(DIP, TEASER), (4)-(7L7L, RANSAC), et (7L5L, LESAC, FlipNetA). L'hypothèse rejetée par la sélection FlipNet est également présentée **(6) : (7L5L, LESAC, FlipNetA rejection)**, afin de se rendre compte déjà qu'il s'agit bien de l'hypothèse obtenue par retournement des pôles, puis visualiser sur ces exemples à quel point cette hypothèse peut être ressemblante à celle sélectionnée par FlipNet. Dans chaque figure, la première ligne montre des visualisations en damier représentant en première diagonale le rein dans l'image fixe CT et en seconde diagonale le rein dans l'image US 3D transformée dans le système de coordonnées du CT par la transformation estimée. Sur ces représentations en damier sont superposés les segmentations vérité terrain dans chaque modalité et le centre du pelvis rénal (point de repère numéro 1). La seconde ligne quant à elle présente des vues 3D des segmentations vérité terrain dans le système de coordonnées du CT après recalage. Ces exemples sont les mêmes que ceux utilisés pour visualiser les résultats des méthodes utilisant les surfaces, de la section 4.1. Les symboles ✓ et ✗ indiquent si les orientations "pôle inférieur-pôle supérieur" et transversale du rein dans l'image US 3D recalée, sont correctes.

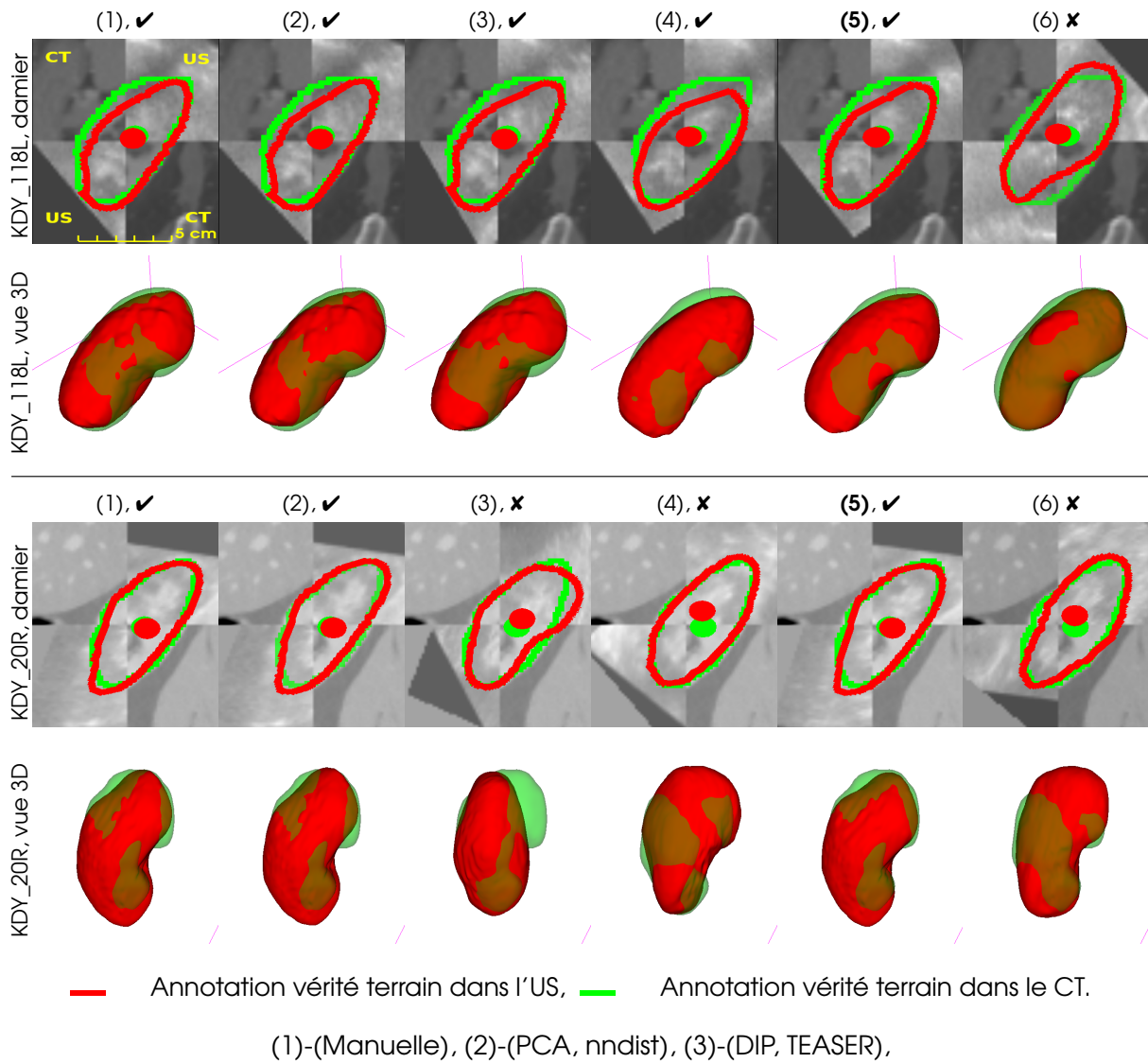
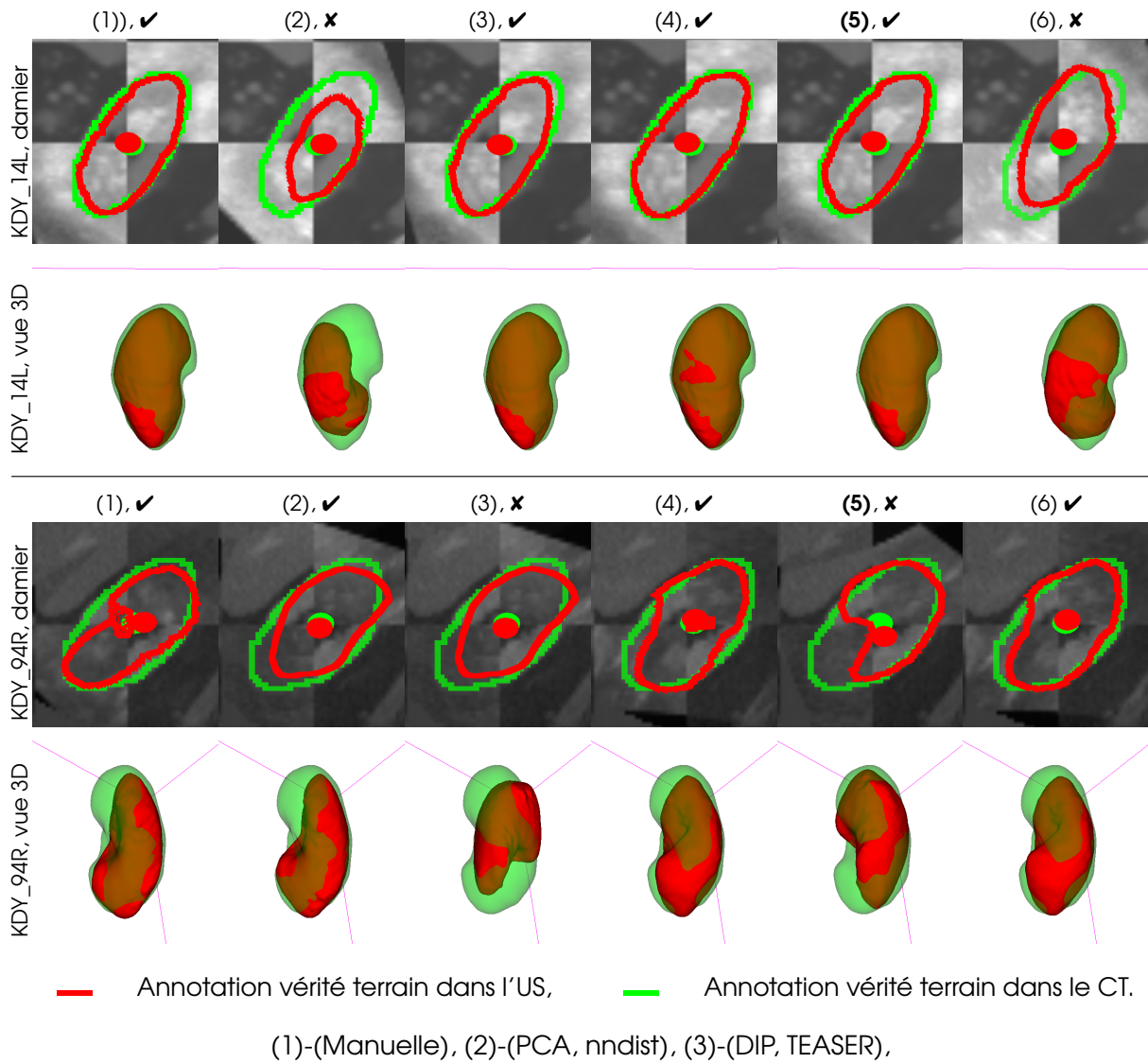


FIGURE 4.15 – Résultats qualitatifs illustrant les performances de recalage pour les exemples KDY_118L et KDY_20R, à l'aide de visualisations en damier (deux morceaux du CT sur la première diagonale, deux morceaux de l'image US recalée sur la deuxième diagonale) et de vues 3D. Chaque colonne correspond à une méthode de recalage global suivie d'un affinement BCPD. Sur les damiers, quatre annotations sont observables : Contour vert : segmentation vérité terrain du rein dans le CT. Point vert : position vérité terrain du point de repère numéro 1 dans le CT. Contour rouge : segmentation vérité terrain du rein dans l'image US déplacée dans le système de coordonnées du CT selon le recalage estimé. Point rouge : position vérité terrain du point de repère numéro 1 dans l'image US, déplacé dans le système de coordonnées du CT selon le recalage estimé. Les symboles ✓ et ✗ indiquent si les orientations "pôle inférieur-pôle supérieur" et transversale sont correctes.



(1)-(Manuelle), (2)-(PCA, nndist), (3)-(DIP, TEASER),

(4)-(7L7L, RANSAC), (5)- **(7L5L, LESAC, FlipNetA)**, (6)-(7L5L, LESAC, FlipNetA rejection)

FIGURE 4.16 – Résultats qualitatifs illustrant les performances de recalage pour les exemples KDY_14L et KDY_94R, à l'aide de visualisations en damier (deux morceaux du CT sur la première diagonale, deux morceaux de l'image US recalée sur la deuxième diagonale) et de vues 3D. La description est la même que celle de la Figure 4.15.

4.3 Conclusion

Une nouvelle approche pour de recalage global (vu comme initialisation d'un recalage précis) des reins dans les modalités CT et US 3D a été introduite. Elle a démontré une robustesse et des performances supérieures à celles des méthodes concurrentes de l'état de l'art. Les défis à relever étaient liés au faible contraste et au bruit associés aux images échographiques, mais aussi à

l’ambiguïté de recalage causée par la symétrie pôle supérieur - pôle inférieur dans la forme du rein. L’approche procède en trois étapes : localisation et étiquetage invariant par symétrie de points de repère dans les deux modalités par deux CNNs de segmentation multilabels, calcul des deux hypothèses de recalage les plus pertinentes par une approche d’échantillonnage exhaustif, et sélection de l’hypothèse correcte par un classifieur prédisant une sorte de masque de votes à l’échelle du voxel en faveur d’une hypothèse ou non. Bien que dans les expériences réalisées, l’approche proposée ait été utilisée pour initialiser un recalage précis utilisant les surfaces, elle-même ne nécessite pas que l’organe soit segmenté *a priori*. Par conséquent, elle peut être combinée à un recalage iconique précis. En perspectives, il serait intéressant de combiner cette approche avec une méthode d’affinement du recalage utilisant les modèles neuronaux profonds et n’inférant pas sur les surfaces, telle que VoxelMorph [75] ou TransMorph [76], écartant ainsi la nécessité de segmenter l’organe de toutes les étapes du pipeline de recalage. Il serait également intéressant d’évaluer l’impact du choix et du nombre de points de repère, sur les performances de recalage ; ainsi que de tester le système dans le cadre d’une validation clinique.

5. CONCLUSIONS ET PERSPECTIVES

La recherche menée dans cette thèse visait à surmonter les défis du recalage automatique d'images échographiques 3D et tomodensitométriques du rein, afin de permettre des applications précises en chirurgie percutanée guidée en temps réel par l'échographie. En intégrant aux systèmes de recalage, des modèles à base de réseaux de neurones profonds, de nouvelles approches permettant d'automatiser complètement le recalage multimodal dans un contexte clinique, ont été développées et validées.

5.1 Synthèse des contributions principales

Ce travail a été structuré autour de deux axes principaux :

Premièrement, introduire un jeu de données unique (TRUSTED), spécifiquement conçu pour le recalage d'images échographiques et scanner du rein. Ce jeu de données constitue une base avec des annotations rigoureuses pour l'entraînement et la validation des modèles de recalage, mais aussi de segmentation. Il permet de pallier le manque de données appariées et annotées, dans le domaine de l'imagerie abdominale multimodale. Les expériences de validation ont confirmé la qualité de ce jeu de données, tant pour les modèles de segmentation que pour les modèles recalage. Il permet ainsi d'assurer une généralisation des résultats d'évaluation des modèles et de soutenir la recherche sur le recalage automatique dans un contexte chirurgical.

Deuxièmement, présenter une méthode originale de recalage global automatique du rein, à base de réseaux de neurones profonds. L'approche, reposant sur la détection et l'étiquetage de points de repère anatomiques, combinés à un "classifieur" d'orientation, s'est montrée efficace dans la désambiguïsation des hypothèses de recalage provenant de la symétrie dans la forme du rein. Les résultats expérimentaux ont validé la robustesse et la précision de cette méthode, en comparaison aux méthodes performantes de l'état de l'art.

Les résultats ainsi obtenus des différentes analyses, ont démontré que le recalage multimodal des images échographiques et scanner du rein, de manière complètement automatique, est bel et bien réalisable, avec une précision,

certes à améliorer, mais déjà acceptable dans la plupart des cas. Cette automatisation offre de nouvelles possibilités pour l'utilisation de l'échographie en tant qu'outil principal de guidage dans des procédures peu invasives. Les méthodes développées dans cette thèse offrent une voie prometteuse vers des systèmes intelligents qui augmentent l'autonomie des chirurgiens tout en minimisant la dépendance à d'autres modalités d'imagerie plus coûteuses ou invasives.

5.2 Perspectives de recherche

Malgré les avancées vers une automatisation complète des systèmes de recalage multimodal intégrant des modèles neuronaux, plusieurs limitations demeurent, ouvrant ainsi des pistes de recherche pour approfondir les contributions de cette thèse.

Tout d'abord, le développement de méthodes précises de recalage d'images échographiques avec d'autres modalités nécessite des annotations denses et fiables (segmentations), un défi majeur en raison du coût élevé de la construction de tels ensembles de données. Bien que le jeu de données TRUSTED représente un progrès, il reste limité en quantité et en variabilité pour le développement de modèles réellement utilisables en contexte clinique. Afin de surmonter cette limitation, les méthodes d'apprentissage faiblement supervisé pourraient réduire le besoin d'annotations manuelles, notamment lors de l'étape d'affinement du recalage. Par ailleurs, l'utilisation d'environnements de simulation peut permettre de générer des images synthétiques, en s'appuyant sur des modèles anatomiques réalistes et diversifiés. En simulant différents scénarios cliniques et morphologies du rein humain, il serait possible de constituer des ensembles d'entraînement de grande taille, diversifiés, équilibrés et représentatifs des variations possibles. Dans le même ordre d'idée, l'utilisation de modèles génératifs d'images serait envisageable, afin de produire des données d'entraînement pour les modèles neuronaux. Un autre axe de recherche à explorer dans ce sens est la collaboration avec d'autres centres cliniques, dans un cadre d'apprentissage fédéré ("Federated-Learning") afin de diversifier les données (en termes d'instruments, de variabilité anatomique et de processus d'acquisitions). L'augmentation de la diversité et de la quantité des données permettrait de former des modèles qui généralisent mieux sur de nouveaux exemples cliniques.

Ensuite, l'automatisation complète du recalage multimodal, appliqué en contexte clinique, possède encore des limitations. Bien que le modèle proposé au Chapitre 4 soit capable de réaliser, dans le cadre de procédures rénales, un recalage automatique relativement précis (en comparaison avec l'état de

l'art), deux problématiques n'ont pas été abordées. Premièrement, celle de la définition de critère(s) utilisable pour décider automatiquement de l'acceptation ou pas d'une transformation de recalage issue d'un système automatique. Ce qui aiderait dans la prise de décision pour des applications en robotique, par exemple. Deuxièmement, la problématique de la quantification d'une valeur de confiance accordée à la réponse (accepté/rejeté) du système. De telles mesures de confiance aideraient le clinicien à prendre la décision finale sur l'acceptation ou le rejet du recalage proposé par le système automatique. En relation avec la définition de critères de validation automatique, il serait intéressant de construire de tels critères en s'appuyant sur les retours des cliniciens, c'est-à-dire en comprenant comment ils procèdent lorsqu'ils ont à prendre de telles décisions; puis en essayant de mimer leurs comportements. S'agissant de la mesure de confiance, une piste de recherche consiste à intégrer des modèles de prédiction de confiance au sein des algorithmes de recalage. De tels modèles pourraient évaluer des variations spatiales ou des incohérences dans les structures recalées, afin de définir des scores de confiance.

Enfin, les performances obtenues notamment au Chapitre 4 pourraient être améliorées en exploitant davantage les informations d'intensité des images. En effet, les méthodes ne faisant pas du tout intervenir de segmentation (à l'entraînement ou à l'inférence) n'ont pas encore été pleinement explorées, en raison de la difficulté de définir une mesure de similarité robuste entre des modalités très différentes. L'application de techniques comme l'apprentissage de représentation partagée, où un réseau de neurones apprend à projeter les images de différentes modalités dans un espace abstrait commun, pourrait offrir des solutions plus robustes au problème de la similarité entre deux modalités très différentes.

Bibliographie

- [1] K. KATSANOS, L. MAILLI, M. KROKIDIS, A. McGRATH, T. SABHARWAL et A. ADAM. « Systematic review and meta-analysis of thermal ablation versus surgical nephrectomy for small renal tumours ». In : *Cardiovascular and interventional radiology* 37 (2014), pages 427-437 (cf. page 22).
- [2] M. KUGLER, A. HOSTETTLER, L. SOLER, Y. REMOND et D. GEORGE. « A new algorithm for volume mesh refinement on merging geometries : application to liver and vascularisation ». In : *Journal of Computational and Applied Mathematics* 330 (2018), pages 429-440 (cf. page 22).
- [3] N. LAUZERAL, D. BORZACCHIELLO, M. KUGLER, D. GEORGE, Y. RÉMOND, A. HOSTETTLER et F. CHINESTA. « A model order reduction approach to create patient-specific mechanical models of human liver in computational medicine applications ». In : *Computer methods and programs in biomedicine* 170 (2019), pages 95-106 (cf. page 22).
- [4] N. LAUZERAL, D. BORZACCHIELLO, M. KUGLER, D. GEORGE, Y. RÉMOND, A. HOSTETTLER et F. CHINESTA. « Shape parametrization of bio-mechanical finite element models based on medical images ». In : *Computer Methods in Biomechanics and Biomedical Engineering : Imaging & Visualization* 7.5-6 (2019), pages 480-489 (cf. page 22).
- [5] M. KUGLER, A. HOSTETTLER, L. SOLER, D. BORZACCHIELLO, F. CHINESTA, D. GEORGE et Y. RÉMOND. « Numerical simulation and identification of macroscopic vascularised liver behaviour : Case of indentation tests ». In : *Bio-Medical Materials and Engineering* 28.s1 (2017), S107-S111 (cf. pages 22, 36).
- [6] D. GEORGE, M. BANIASSADI, Y. HOARAU, M. KUGLER et Y. RÉMOND. « Influence of the liver vascular distribution on its overall mechanical behavior : a first approach to multiscale fluid-structure homogenization ». In : *Journal of Cellular Immunotherapy* 4.1 (2018), pages 35-37 (cf. pages 22, 36).
- [7] A. HOSTETTLER, S. A. NICOLAU, L. SOLER, Y. RÉMOND et J. MARESCAUX. « A real-time predictive simulation of abdominal organ positions induced by free breathing ». In : *Biomedical Simulation : 4th International Symposium, ISBMS 2008, London, UK, July 7-8, 2008 Proceedings* 4. Springer. 2008, pages 89-97 (cf. pages 22, 37).

- [8] L. SOLER, S. NICOLAU, A. HOSTETTLER, J.-B. FASQUEL, V. AGNUS, A. CHARNOZ, J. MOREAU, B. DALLEMAGNE, D. MUTTER et J. MARESCAUX. « Computer-Assisted Digestive Surgery ». In : *Computational surgery and dual training*. Springer, 2009, pages 139-153 (cf. page 22).
- [9] T. COLLINS, P. CHAUVET, C. DEBIZE, D. PIZARRO, A. BARTOLI, M. CANIS et N. BOURDEL. « A system for augmented reality guided laparoscopic tumour resection with quantitative ex-vivo user evaluation ». In : *International Workshop on Computer-Assisted and Robotic Endoscopy*. Springer. 2016, pages 114-126 (cf. page 22).
- [10] S. CHAMBON, A. MORENO, A. P. SANTHANAM, J. P. ROLLAND, E. ANGELINI et I. BLOCH. « CT-PET landmark-based lung registration using a dynamic breathing model ». In : *14th International Conference on Image Analysis and Processing (ICIAP 2007)*. IEEE. 2007, pages 691-696 (cf. page 22).
- [11] A. BARTOLI. « Groupwise geometric and photometric direct image registration ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.12 (2008), pages 2098-2108 (cf. page 22).
- [12] M. SDIKA, L. ALSTON, L. MAHIEU-WILLIAME, J. GUYOTAT, D. ROUSSEAU et B. MONTCEL. « Robust real time motion compensation for intraoperative video processing during neurosurgery ». In : *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2016, pages 1046-1049 (cf. page 22).
- [13] Y. BENGIO, I. GOODFELLOW et A. COURVILLE. *Deep learning*. Tome 1. MIT press Cambridge, MA, USA, 2017 (cf. page 22).
- [14] O. PETIT, N. THOME, C. RAMBOUR, L. THEMYS, T. COLLINS et L. SOLER. « U-net transformer : Self and cross attention for medical image segmentation ». In : *Machine Learning in Medical Imaging : 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer. 2021, pages 267-276 (cf. page 22).
- [15] N. THOME. « Beyond Full Supervision in Deep Learning ». In : *Multi-faceted Deep Learning : Models and Data* (2021), pages 73-101 (cf. page 22).
- [16] G. SAIBRO, M. DIANA, B. SAUER, J. MARESCAUX, A. HOSTETTLER et T. COLLINS. « Automatic detection of steatosis in ultrasound images with comparative visual labeling ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pages 408-418 (cf. page 22).

- [17] M. PAVONE, B. SEELIGER, E. TEODORICO, M. GOGLIA, C. TALIENTO, N. BIZZARRI, L. LECOINTRE, C. AKLADIOS, A. FORGIONE, G. SCAMBIA et al. « Ultrasound-guided robotic surgical procedures : a systematic review ». In : *Surgical Endoscopy* 38.5 (2024), pages 2359-2370 (cf. page 23).
- [18] G. BERLYNE. « Ultrasonics in renal biopsy an aid to determination of kidney position ». In : *The Lancet* 278.7205 (1961), pages 750-751 (cf. page 23).
- [19] H. HOLM, J. K. KRISTENSEN, S. N. RASMUSSEN, A. NORTHEVED et H. BARLEBO. « Ultrasound as a guide in percutaneous puncture technique ». In : *Ultrasonics* 10.2 (1972), pages 83-86 (cf. page 23).
- [20] B. B. GOLDBERG et H. M. POLLACK. « Ultrasonic aspiration-biopsy transducer ». In : *Radiology* 108.3 (1973), pages 667-671 (cf. page 23).
- [21] J. MACHI et E. D. STAREN. *Ultrasound for surgeons*. Lippincott Williams & Wilkins, 2005 (cf. page 23).
- [22] DOC-MARKET. *marketplace for medical equipment and instruments*. <https://www.doc-market.eu/VWNUXDT8-siemens-x300-acuson-ultrasound?spr=en>. Accessed : 2024-10-30. 2024 (cf. page 24).
- [23] S. HEALTHINEERS. *Les grands moments du diagnostic par ultrasons*. <https://www.siemens-healthineers.com/fr/ultrasound/news/history>. Accessed : 2024-10-30. 2024 (cf. page 24).
- [24] M. EXPO. <https://www.medicalexpo.de/prod/ge-healthcare/product-70717-429143.html>. Accessed : 2024-10-30. 2024 (cf. page 24).
- [25] J. M. VELASCO et K. HOOD. « Percutaneous ultrasound guidance techniques and procedures ». In : *Abdominal Ultrasound for Surgeons*. Springer, 2014, pages 89-107 (cf. page 24).
- [26] G. POGGI, N. TOSORATTI, B. MONTAGNA et C. PICCHI. « Microwave ablation of hepatocellular carcinoma ». In : *World journal of hepatology* 7.25 (2015), page 2578 (cf. page 25).
- [27] M. B. A. HAGHIGHAT, A. AGHAGOLZADEH et H. SEYEDARABI. « Multi-focus image fusion for visual sensor networks in DCT domain ». In : *Computers & Electrical Engineering* 37.5 (2011), pages 789-797 (cf. page 26).
- [28] F. IDE. *Fiches ide*. <https://www.fiches-ide.fr/anatomie-physiologie/rein-vue-anterieur/34-rein-vue-anterieur/>. Accessed : 2024-10-30. 2024 (cf. page 27).

- [29] H. SUNG, J. FERLAY, R. L. SIEGEL, M. LAVERSANNE, I. SOERJOMATARAM, A. JEMAL et F. BRAY. « Global cancer statistics 2020 : GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries ». In : *CA : a cancer journal for clinicians* 71.3 (2021), pages 209-249 (cf. page 27).
- [30] L. BUKAVINA, K. BENSALAH, F. BRAY, M. CARLO, B. CHALLACOMBE, J. A. KARAM, W. KASSOUF, T. MITCHELL, R. MONTIRONI, T. O'BRIEN et al. « Epidemiology of renal cell carcinoma : 2022 update ». In : *European urology* 82.5 (2022), pages 529-542 (cf. pages 27, 28).
- [31] M. H. MOZAFFARI et W.-S. LEE. « Freehand 3-D ultrasound imaging : a systematic review ». In : *Ultrasound in medicine & biology* 43.10 (2017), pages 2099-2124 (cf. page 29).
- [32] N. MASOUMI, C. J. BELASSO, M. O. AHMAD, H. BENALI, Y. XIAO et H. RIVAZ. « Multimodal 3D ultrasound and CT in image-guided spinal surgery : public database and new registration algorithms ». In : *International Journal of Computer Assisted Radiology and Surgery* 16 (2021), pages 555-565 (cf. pages 29, 30).
- [33] A. LEROY. « Méthodes de recalage scanner/échographie. Application a la navigation chirurgicale des ponctions rénales percutanées ». Thèse de doctorat. Institut National Polytechnique de Grenoble-INPG, 2004 (cf. pages 29, 30).
- [34] Q. HUANG et Z. ZENG. « A review on real-time 3D ultrasound imaging technology ». In : *BioMed research international* 2017.1 (2017), page 6027029 (cf. page 29).
- [35] F. BUREAU, J. ROBIN, A. LE BER, W. LAMBERT, M. FINK et A. AUBRY. « Three-dimensional ultrasound matrix imaging ». In : *Nature Communications* 14.1 (2023), page 6793 (cf. page 29).
- [36] S. MAZAHERI, P. S. B. SULAIMAN, R. WIRZA, M. Z. DIMON, F. KHALID et R. M. TAYEBI. « A Review of Ultrasound and Computed Tomography Registration Approaches ». In : *2014 International Conference on Computer Assisted System in Health*. IEEE. 2014, pages 6-11 (cf. page 30).
- [37] J. RAMALHINHO. « Registration of ultrasound and computed tomography for guidance of laparoscopic liver surgery ». Thèse de doctorat. UCL (University College London), 2020 (cf. page 30).
- [38] V. MARKOVA, M. RONCHETTI, W. WEIN, O. ZETTINIG et R. PREVOST. « Global multi-modal 2D/3D registration via local descriptors learning ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pages 269-279 (cf. pages 30, 34, 46, 59, 92).

- [39] Y. SUN, A. MOELKER, W. J. NIESSEN et T. van WALSUM. « Towards robust CT-ultrasound registration using deep learning methods ». In : *Understanding and Interpreting Machine Learning in Medical Image Computing Applications : First International Workshops, MLCN 2018, DLF 2018, and iMIMIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16-20, 2018, Proceedings 1*. Springer. 2018, pages 43-51 (cf. page 30).
- [40] D. LEE, W. H. NAM, J. Y. LEE et J. B. RA. « Non-rigid registration between 3D ultrasound and CT images of the liver based on intensity and gradient information ». In : *Physics in Medicine & Biology* 56.1 (2010), page 117 (cf. page 30).
- [41] V. KARNIK, A. FENSTER, J. BAX, D. COOL, L. GARDI, I. GYACSKOV, C. ROMAGNOLI et A. WARD. « Assessment of image registration accuracy in three-dimensional transrectal ultrasound-guided prostate biopsy ». In : *Medical physics* 37.2 (2010), pages 802-813 (cf. page 30).
- [42] S. VAN DER MEER, S. M. CAMPS, W. J. VAN ELMPT, M. PODESTA, P. G. SANCHES, B. G. VANNESTE, D. FONTANAROSA et F. VERHAEGEN. « Simulation of pseudo-CT images based on deformable image registration of ultrasound images : a proof of concept for transabdominal ultrasound imaging of the prostate during radiotherapy ». In : *Medical Physics* 43.4 (2016), pages 1913-1920 (cf. page 30).
- [43] Z. M. C. BAUM, S. U. SAEED, Z. MIN, Y. HU et D. C. BARRATT. *MR to Ultrasound Registration for Prostate Challenge - Dataset*. Version 1.0.0. Zenodo, avril 2023 (cf. pages 30, 31, 73).
- [44] F. PAPARO, R. PICCAZZO, L. CEVASCO, A. PICCARDO, F. PINNA, F. BELLI, L. BACIGALUPO, E. BISCALDI, G. DE CARO et G. A. ROLLANDI. « Advantages of percutaneous abdominal biopsy under PET-CT/ultrasound fusion imaging guidance : a pictorial essay ». In : *Abdominal imaging* 39 (2014), pages 1102-1113 (cf. page 31).
- [45] B. FEI, O. A. ABIODUN-OJO, A. A. AKINTAYO, O. AKIN-AKINTAYO, F. TADE, P. T. NIEH, V. A. MASTER, M. ALEMOZAFFAR, A. O. OSUNKOYA, M. M. GOODMAN et al. « Feasibility and initial results : fluciclovine positron emission tomography/ultrasound fusion targeted biopsy of recurrent prostate cancer ». In : *The Journal of urology* 202.2 (2019), pages 413-421 (cf. page 31).
- [46] C. HUANG, Y. XIE, Y. LAN, Y. HAO, F. CHEN, Y. CHENG et Y. PENG. « A new framework for the integrative analytics of intravascular ultrasound and optical coherence tomography images ». In : *IEEE Access* 6 (2018), pages 36408-36419 (cf. page 31).

- [47] L. RÄBER, J. H. HEO, M. D. RADU, H. M. GARCIA-GARCIA, G. G. STEFANINI, A. MOSCHOVITIS, J. DIJKSTRA, H. KELBAEK, S. WINDECKER, P. W. SERRUYS et al. « Offline fusion of co-registered intravascular ultrasound and frequency domain optical coherence tomography images for the analysis of human atherosclerotic plaques ». In : *EuroIntervention* 8.1 (2012), pages 98-108 (cf. page 31).
- [48] D. S. MOLONY, L. H. TIMMINS, E. RASOUL-ARZUMLY, H. SAMADY et D. P. GIDDENS. « Evaluation of a framework for the co-registration of intravascular ultrasound and optical coherence tomography coronary artery pullbacks ». In : *Journal of biomechanics* 49.16 (2016), pages 4048-4056 (cf. page 31).
- [49] X. FANG, S. XU, B. J. WOOD et P. YAN. « Deep learning-based liver segmentation for fusion-guided intervention ». In : *International journal of computer assisted radiology and surgery* 15 (2020), pages 963-972 (cf. page 32).
- [50] J. LIU, G. SINGH, S. AL'AREF, B. LEE, O. OLERU, J. K. MIN, S. DUNHAM, M. R. SABUNCU et B. MOSADEGH. « Image registration in medical robotics and intelligent systems : fundamentals and applications ». In : *Advanced Intelligent Systems* 1.6 (2019), page 1900048 (cf. page 32).
- [51] M. SAHU, H. ISHIDA, L. CONNOLLY, H. FAN, A. DEGUET, P. KAZANZIDES, F. X. CREIGHTON, R. H. TAYLOR et A. MUNAWAR. « ENTR1 : Enhanced Navigational Toolkit for Robotic Interventions ». In : *IEEE Transactions on Medical Robotics and Bionics* (2024) (cf. page 32).
- [52] X. BAI, H. WANG, Y. QIN, J. HAN et N. YU. « MatchMorph : A real-time pre-and intra-operative deformable image registration framework for MRI-guided surgery ». In : *Computers in Biology and Medicine* 180 (2024), page 108948 (cf. page 32).
- [53] A. WITTEK, K. MILLER, R. KIKINIS et S. K. WARFIELD. « Patient-specific model of brain deformation : Application to medical image registration ». In : *Journal of biomechanics* 40.4 (2007), pages 919-929 (cf. page 32).
- [54] A. MILOJEVIC, D. PETER, N. B. HUBER, L. AZEVEDO, A. LATYSHEV, I. SAILER, M. GROSS, B. THOMASZEWSKI, B. SOLENTHALER et B. GÖZCÜ. « AutoSkull : Learning-Based Skull Estimation for Automated Pipelines ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024, pages 109-118 (cf. page 33).
- [55] S. GUPTA, P. GUPTA et V. S. VERMA. « Study on anatomical and functional medical image registration methods ». In : *Neurocomputing* 452 (2021), pages 534-548 (cf. page 33).

- [56] A. CHARNOZ, V. AGNUS, G. MALANDAIN, C. FOREST, M. TAJINE et L. SOLER. « Liver registration for the follow-up of hepatic tumors ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2005, pages 155-162 (cf. page 33).
- [57] A. HOSTETTLER, D. GEORGE, Y. RÉMOND, S. NICOLAU, L. SOLER et J. MARESCAUX. « Bulk modulus and volume variation measurement of the liver and the kidneys in vivo using abdominal kinetics during free breathing ». In : *Computer methods and programs in biomedicine* 100 2 (2010), pages 149-57 (cf. page 34).
- [58] A. KARIMI et A. SHOJAEI. « Measurement of the mechanical properties of the human kidney ». In : *Irbm* 38.5 (2017), pages 292-297 (cf. page 34).
- [59] A. LEROY, P. MOZER, Y. PAYAN et J. TROCCAZ. « Rigid registration of freehand 3D ultrasound and CT-scan kidney images ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2004, pages 837-844 (cf. page 34).
- [60] R. DELAUNAY, R. ZHANG, F. C. PEDROSA, N. FEIZI, D. SACCO, R. V. PATEL et J. JAGADEESAN. « Transformer-based local feature matching for multimodal image registration ». In : *Medical Imaging 2024 : Image Processing*. Tome 12926. SPIE. 2024, pages 138-143 (cf. pages 34, 59, 60).
- [61] P. JACKSON, R. SIMON et C. LINTE. « Surgical tracking, registration, and navigation characterization for image-guided renal interventions ». In : *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pages 5081-5084 (cf. page 34).
- [62] S. XING, J. C. ROMERO, P. ROY, D. W. COOL, D. TESSIER, E. C. CHEN, T. M. PETERS et A. FENSTER. « 3D US-CT/MRI registration for percutaneous focal liver tumor ablations ». In : *International Journal of Computer Assisted Radiology and Surgery* 18.7 (2023), pages 1159-1166 (cf. pages 34, 52, 53, 111).
- [63] A. CIFOR, L. RISSER, M. P. HEINRICH, D. CHUNG et J. A. SCHNABEL. « Rigid registration of untracked freehand 2D ultrasound sweeps to 3D CT of liver tumours ». In : *Abdominal Imaging. Computation and Clinical Applications : 5th International Workshop, Held in Conjunction with MICCAI 2013, Nagoya, Japan, September 22, 2013. Proceedings* 5. Springer. 2013, pages 155-164 (cf. pages 34, 35, 50).
- [64] F. BING, J. VAPPOU, E. BRETON, I. ENESCU, J. GARNON et A. GANGI. « Accuracy of a CT-Ultrasound fusion imaging guidance system used for hepatic percutaneous procedures ». In : *Journal of Vascular and Interventional Radiology* 30.7 (2019), pages 1013-1020 (cf. page 34).

- [65] W. WEIN, S. BRUNKE, A. KHAMENE, M. R. CALLSTROM et N. NAVAB. « c ». In : *Medical image analysis* 12.5 (2008), pages 577-585 (cf. pages 34, 35, 54, 68, 78, 92, 93).
- [66] G. IONESCU, S. LAVALLÉE et J. DEMONGEOT. « Automated registration of ultrasound with CT images : application to computer assisted prostate radiotherapy and orthopedics ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 1999, pages 768-777 (cf. page 34).
- [67] Z. FANG, H. DELINGETTE et N. AYACHE. « Anatomical Landmark Detection for Initializing US and MR Image Registration ». In : *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer. 2023, pages 165-174 (cf. pages 34, 46, 60, 61, 124).
- [68] M. MÜLLER, L. E. S. HELLJESSEN, R. PREVOST, I. VIOLA, K. NYLUND, O. H. GILJA, N. NAVAB et W. WEIN. « Deriving Anatomical Context from 4D Ultrasound. » In : *VCBM*. 2014, pages 173-180 (cf. pages 35, 54).
- [69] J. BANERJEE, Y. SUN, C. KLINK, R. GAHRMANN, W. J. NIESSEN, A. MOELKER et T. van WALSUM. « Multiple-correlation similarity for block-matching based fast CT to ultrasound registration in liver interventions ». In : *Medical image analysis* 53 (2019), pages 132-141 (cf. pages 35, 54, 55).
- [70] A. STRITTMATTER, L. R. SCHAD et F. G. ZÖLLNER. « Deep learning-based affine medical image registration for multimodal minimal-invasive image-guided interventions—A comparative study on generalizability ». In : *Zeitschrift für Medizinische Physik* 34.2 (2024), pages 291-317 (cf. page 35).
- [71] Y. HU, M. MODAT, E. GIBSON, W. LI, N. GHAVAMI, E. BONMATI, G. WANG, S. BANDULA, C. M. MOORE, M. EMBERTON et al. « Weakly-supervised convolutional neural networks for multimodal image registration ». In : *Medical image analysis* 49 (2018), pages 1-13 (cf. pages 35, 64).
- [72] X. SONG, X. XU, S. XU, B. TURKBEY, T. SANFORD, B. J. WOOD et P. YAN. « Distance map supervised landmark localization for MR-TRUS registration ». In : *Medical Imaging 2023 : Image Processing*. Tome 12464. SPIE. 2023, pages 721-726 (cf. page 35).
- [73] K. EPPENHOF. « Supervised deformable image registration using deep neural networks ». English. Proefschrift. Phd Thesis 1 (Research TU/e / Graduation TU/e). Biomedical Engineering, février 2020. ISBN : 978-90-386-4970-2 (cf. page 36).

- [74] H. DELINGETTE, S. COTIN et N. AYACHE. « A hybrid elastic model allowing real-time cutting, deformations and force-feedback for surgery training and simulation ». In : *Proceedings Computer Animation 1999*. IEEE. 1999, pages 70-81 (cf. page 36).
- [75] G. BALAKRISHNAN, A. ZHAO, M. R. SABUNCU, J. GUTTAG et A. V. DALCA. « Voxel-morph : a learning framework for deformable medical image registration ». In : *IEEE transactions on medical imaging* 38.8 (2019), pages 1788-1800 (cf. pages 36, 45, 64, 66, 92, 93, 143).
- [76] J. CHEN, E. C. FREY, Y. HE, W. P. SEGARS, Y. LI et Y. DU. « Transmorph : Transformer for unsupervised medical image registration ». In : *Medical image analysis* 82 (2022), page 102615 (cf. pages 36, 45, 65, 67, 92, 93, 143).
- [77] J. ZOU, B. GAO, Y. SONG et J. QIN. « A review of deep learning-based deformable medical image registration ». In : *Frontiers in Oncology* 12 (2022), page 1047215 (cf. page 37).
- [78] B. R. THOMSON, J. N. SMIT, O. V. IVASHCHENKO, N. F. KOK, K. F. KUHLMANN, T. J. RUERS et M. FUSAGLIA. « MR-to-US registration using multiclass segmentation of hepatic vasculature with a reduced 3D U-Net ». In : *International conference on medical image computing and computer-assisted intervention*. Springer. 2020, pages 275-284 (cf. pages 37, 53).
- [79] Q. ZENG, S. MOHAMMED, E. H. PANG, C. SCHNEIDER, M. HONARVAR, J. LOBO, C. HU, J. JAGO, G. NG, R. ROHLING et al. « Learning-based US-MR liver image registration with spatial priors ». In : *International conference on medical image computing and computer-assisted intervention*. Springer. 2022, pages 174-184 (cf. page 37).
- [80] R. E. ONG, S. D. HERRELL III, M. I. MIGA et R. L. GALLOWAY JR. « A kidney deformation model for use in non-rigid registration during image-guided surgery ». In : *Medical Imaging 2008 : Visualization, Image-Guided Procedures, and Modeling*. Tome 6918. SPIE. 2008, pages 317-325 (cf. page 37).
- [81] P. JACKSON, K. MERRELL, R. SIMON et C. LINTE. « Investigating uncertainty in augmented reality enhanced renal navigation using in vitro patient-specific tissue-mimicking phantoms ». In : *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2022, pages 2997-3002 (cf. page 37).
- [82] M. YANG, H. DING, L. ZHU et G. WANG. « Ultrasound fusion image error correction using subject-specific liver motion model and automatic image registration ». In : *Computers in biology and medicine* 79 (2016), pages 99-109 (cf. page 37).

- [83] Y. HU, E. GIBSON, H. U. AHMED, C. M. MOORE, M. EMBERTON et D. C. BARRATT. « Population-based prediction of subject-specific prostate deformation for MR-to-ultrasound image registration ». In : *Medical image analysis* 26.1 (2015), pages 332-344 (cf. page 37).
- [84] B. LI, G. E. CHRISTENSEN, J. DILL, E. A. HOFFMAN et J. M. REINHARDT. « 3D intersubject warping and registration of pulmonary CT images for a human lung model ». In : *Medical Imaging 2002 : Physiology and Function from Multidimensional Images*. Tome 4683. SPIE. 2002, pages 324-335 (cf. page 37).
- [85] Y. WANG, Q. ZHENG et P. A. HENG. « Online robust projective dictionary learning : shape modeling for MR-TRUS registration ». In : *IEEE Transactions on Medical Imaging* 37.4 (2017), pages 1067-1078 (cf. page 38).
- [86] A. MOHAMED, C. DAVATZIKOS et R. TAYLOR. « A combined statistical and biomechanical model for estimation of intra-operative prostate deformation ». In : *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002 : 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part II* 5. Springer. 2002, pages 452-460 (cf. page 38).
- [87] A. SOTIRAS, C. DAVATZIKOS et N. PARAGIOS. « Deformable Medical Image Registration : A Survey ». In : *IEEE Transactions on Medical Imaging* 32.7 (2013), pages 1153-1190 (cf. page 40).
- [88] C. STUDHOLME, D. HILL et D. HAWKES. « An overlap invariant entropy measure of 3D medical image alignment ». In : *Pattern Recognition* 32.1 (1999), pages 71-86. ISSN : 0031-3203 (cf. page 40).
- [89] D. LOECKX, P. SLAGMOLEN, F. MAES, D. VANDERMEULEN et P. SUETENS. « Nonrigid image registration using conditional mutual information ». en. In : *Inf. Process. Med. Imaging* 20 (2007), pages 725-737 (cf. page 40).
- [90] C. STUDHOLME, C. DRAPACA, B. IORDANOVA et V. CARDENAS. « Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change ». en. In : *IEEE Trans. Med. Imaging* 25.5 (mai 2006), pages 626-639 (cf. page 40).
- [91] H.-M. CHAN, A. CHUNG, S. YU, A. NORBASH et W. WELLS. « Multi-modal image registration by minimizing Kullback-Leibler distance between expected and observed joint class histograms ». In : *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*. Tome 2. 2003, pages 11-570 (cf. page 40).

- [92] R. LIAO, C. GUETTER, C. XU, Y. SUN, A. KHAMENE et F. SAUER. « Learning-Based 2D/3D Rigid Registration Using Jensen-Shannon Divergence for Image-Guided Surgery ». In : *Medical Imaging and Augmented Reality*. Sous la direction de G.-Z. YANG, T. JIANG, D. SHEN, L. GU et J. YANG. Berlin, Heidelberg : Springer Berlin Heidelberg, 2006, pages 228-235. ISBN : 978-3-540-37221-9 (cf. page 40).
- [93] G. WU, M. KIM, Q. WANG, Y. GAO, S. LIAO et D. SHEN. « Unsupervised deep feature learning for deformable registration of MR brain images ». In : *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pages 649-656 (cf. page 40).
- [94] X. CHENG, L. ZHANG et Y. ZHENG. « Deep similarity learning for multimodal medical images ». In : *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* 6.3 (2018), pages 248-252 (cf. page 40).
- [95] G. HASKINS, J. KRUECKER, U. KRUGER, S. XU, P. A. PINTO, B. J. WOOD et P. YAN. « Learning deep similarity metric for 3D MR–TRUS image registration ». In : *Int. J. Comput. Assist. Radiol. Surg.* 14.3 (2019), pages 417-425 (cf. pages 40, 67, 68).
- [96] M. SIMONOVSKY, B. GUTIÉRREZ-BECKER, D. MATEUS, N. NAVAB et N. KOMODAKIS. « A deep metric for multimodal registration ». In : *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pages 10-18 (cf. page 40).
- [97] F. ISENSEE, P. F. JAEGER, S. A. KOHL, J. PETERSEN et K. H. MAIER-HEIN. « nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation ». In : *Nature methods* 18.2 (2021), pages 203-211 (cf. pages 41, 53, 84, 125).
- [98] O. PETIT, N. THOME, C. RAMBOUR, L. THEMUR, T. COLLINS et L. SOLER. « U-Net Transformer : Self and Cross Attention for Medical Image Segmentation ». In : *Machine Learning in Medical Imaging*. Sous la direction de C. LIAN, X. CAO, I. REKIK, X. XU et P. YAN. Cham : Springer International Publishing, 2021, pages 267-276 (cf. page 41).
- [99] G. K. L. TAM, Z. CHENG, Y. K. LAI, F. C. LANGBEIN, Y. LIU, D. MARSHALL, R. R. MARTIN, X. SUN et P. L. ROSIN. « Registration of 3D Point Clouds and Meshes : A Survey from Rigid to Nonrigid ». In : *IEEE Transactions on Visualization and Computer Graphics* 19.7 (juillet 2013), pages 1199-1217 (cf. page 41).

- [100] X. HUANG, G. MEI, J. ZHANG et R. ABBAS. « A comprehensive survey on point cloud registration ». In : *arXiv preprint arXiv :2103.02690* (2021). Review paper, version 2 (cf. page 41).
- [101] F. WANG et Z. ZHAO. « A survey of iterative closest point algorithm ». In : *2017 Chinese Automation Congress (CAC)*. Jinan, China, 2017, pages 4395-4399 (cf. page 41).
- [102] S. RUSINKIEWICZ et M. LEVOY. « Efficient variants of the ICP algorithm ». In : *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*. 2001, pages 145-152 (cf. page 41).
- [103] A. MYRONENKO et X. SONG. « Point Set Registration : Coherent Point Drift ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (2010), pages 2262-2275 (cf. page 41).
- [104] D. CHETVERIKOV, D. SVIRKO, D. STEPANOV et P. KRSEK. « The Trimmed Iterative Closest Point algorithm ». In : *2002 International Conference on Pattern Recognition*. Tome 3. 2002, 545-548 vol.3 (cf. page 41).
- [105] B. K. P. HORN, H. M. HILDEN et S. NEGAHDARIPOUR. « Closed-form solution of absolute orientation using orthonormal matrices ». In : *J. Opt. Soc. Am. A* 5.7 (juillet 1988), pages 1127-1135 (cf. page 42).
- [106] M. A. FISCHLER et R. C. BOLLES. « Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography ». In : *Communications of the ACM* 24.6 (1981), pages 381-395 (cf. pages 43, 49, 109, 110, 125).
- [107] Y. FU, Y. LEI, T. WANG, P. PATEL, A. B. JANI, H. MAO, W. J. CURRAN, T. LIU et X. YANG. « Biomechanically constrained non-rigid MR-TRUS prostate registration using deep learning based 3D point cloud matching ». In : *Medical Image Analysis* 67 (2021), page 101845. ISSN : 1361-8415 (cf. page 45).
- [108] W. J. M. van de VEN, Y. HU, J. O. BARENTSZ, N. KARSSEMEIJER, D. BARRATT et H. J. HUISMAN. « Biomechanical modeling constrained surface-based image registration for prostate MR guided TRUS biopsy ». en. In : *Med. Phys.* 42.5 (mai 2015), pages 2470-2481 (cf. page 45).
- [109] J. H. SONG. *Methods for evaluating image registration*. The University of Iowa, 2017 (cf. page 46).
- [110] R. B. RUSU, N. BLODOW et M. BEETZ. « Fast point feature histograms (FPFH) for 3D registration ». In : *2009 IEEE international conference on robotics and automation*. 2009, pages 3212-3217 (cf. pages 48, 109).

- [111] H. YANG, J. SHI et L. CARLONE. « TEASER : Fast and Certifiable Point Cloud Registration ». In : *IEEE Trans. Robotics* (2020) (cf. pages 49, 109-111).
- [112] A. H. FORUZAN et H. R. MOTLAGH. « Multimodality liver registration of Open-MR and CT scans ». In : *International journal of computer assisted radiology and surgery* 10 (2015), pages 1253-1267 (cf. pages 49, 50).
- [113] H. Z. U. REHMAN et S. LEE. « Automatic image alignment using principal component analysis ». In : *IEEE Access* 6 (2018), pages 72063-72072 (cf. page 49).
- [114] J. RAMALHINHO, H. F. TREGIDGO, K. GURUSAMY, D. J. HAWKES, B. DAVIDSON et M. J. CLARKSON. « Registration of untracked 2D laparoscopic ultrasound to CT images of the liver using multi-labelled content-based image retrieval ». In : *IEEE Transactions on Medical Imaging* 40.3 (2020), pages 1042-1054 (cf. pages 50, 51).
- [115] K. S. ARUN, T. S. HUANG et S. D. BLOSTEIN. « Least-Squares Fitting of Two 3-D Point Sets ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9.5 (1987), pages 698-700 (cf. pages 52, 92).
- [116] A. MYRONENKO et X. SONG. « Point set registration : Coherent point drift ». In : *IEEE transactions on pattern analysis and machine intelligence* 32.12 (2010), pages 2262-2275 (cf. page 52).
- [117] O. HIROSE. « A Bayesian Formulation of Coherent Point Drift ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.7 (2021), pages 2269-2286 (cf. pages 52, 92, 111).
- [118] S. DU, N. ZHENG, S. YING et J. LIU. « Affine iterative closest point algorithm for point set registration ». In : *Pattern Recognition Letters* 31.9 (2010), pages 791-799 (cf. page 52).
- [119] S. CHENG, I. MARRAS, S. ZAFEIRIOU et M. PANTIC. « Statistical non-rigid ICP algorithm and its application to 3D face alignment ». In : *Image and Vision Computing* 58 (2017), pages 3-12 (cf. page 52).
- [120] A. LEROY, P. MOZER, Y. PAYAN, F. RICHARD, E. CHARTIER-KASTLER et J. TROCCAZ. « Percutaneous renal puncture : requirements and preliminary results ». In : *arXiv preprint physics/0610209* (2006) (cf. page 52).
- [121] Y. SONG, J. TOTZ, S. THOMPSON, S. JOHNSEN, D. BARRATT, C. SCHNEIDER, K. GURUSAMY, B. DAVIDSON, S. OURSELIN, D. HAWKES et al. « Locally rigid, vessel-based registration for laparoscopic liver surgery ». In : *International journal of computer assisted radiology and surgery* 10 (2015), pages 1951-1961 (cf. page 52).

- [122] A. BAIG, M. A. CHAUDHRY et A. MAHMOOD. « Local normalized cross correlation for geo-registration ». In : *Proceedings of 2012 9th International Bhurban Conference on Applied Sciences & Technology (IBCAST)*. IEEE. 2012, pages 70-74 (cf. page 54).
- [123] A. KHAMENE, P. BLOCH, W. WEIN, M. SVATOS et F. SAUER. « Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy ». In : *Medical image analysis* 10.1 (2006), pages 96-112 (cf. pages 54, 92).
- [124] C. W. HURKMANS, P. REMEIJER, J. V. LEBESQUE et B. J. MIJNHEER. « Set-up verification using portal imaging ; review of current clinical practice ». In : *Radiotherapy and oncology* 58.2 (2001), pages 105-120 (cf. page 54).
- [125] T. LANGE, N. PAPENBERG, S. HELDMANN, J. MODERSITZKI, B. FISCHER, H. LAMECKER et P. M. SCHLAG. « 3D ultrasound-CT registration of the liver using combined landmark-intensity information ». In : *International journal of computer assisted radiology and surgery* 4 (2009), pages 79-88 (cf. page 54).
- [126] Y. ZHANG, Y. Zou et P. X. LIU. « Point Cloud Registration in Laparoscopic Liver Surgery Using Keypoint Correspondence Registration Network ». In : *IEEE Transactions on Medical Imaging* (2024) (cf. pages 56, 57).
- [127] L. SOLER, A. HOSTETTLER, V. AGNUS, A. CHARNOZ, J.-B. FASQUEL, J. MOREAU, A.-B. OSSWALD, M. BOUHADJAR et J. MARESCAUX. « 3D image reconstruction for comparison of algorithm database ». In : URL : <https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01> (2010) (cf. page 57).
- [128] Z. YANG, R. SIMON et C. A. LINTE. « Learning feature descriptors for pre-and intra-operative point cloud matching for laparoscopic liver registration ». In : *International journal of computer assisted radiology and surgery* 18.6 (2023), pages 1025-1032 (cf. pages 56, 57).
- [129] F. POIESI et D. BOSCAINI. « Distinctive 3D local deep descriptors ». In : *2020 25th International conference on pattern recognition (ICPR)*. 2021, pages 5720-5727 (cf. pages 57, 58, 109).
- [130] H. FAN, H. Su et L. J. GUIBAS. « A point set generation network for 3d object reconstruction from a single image ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pages 605-613 (cf. page 58).

- [131] F. POIESI et D. BOSCAINI. « Learning general and distinctive 3D local deep descriptors for point cloud registration ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.3 (2022), pages 3979-3985 (cf. page 58).
- [132] J. SUN, Z. SHEN, Y. WANG, H. BAO et X. ZHOU. « LoFTR : Detector-free local feature matching with transformers ». In : *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pages 8922-8931 (cf. page 60).
- [133] C. PAYER, D. ŠTERN, H. BISCHOF et M. URSCHLER. « Regressing heatmaps for multiple landmark localization using CNNs ». In : *International conference on medical image computing and computer-assisted intervention*. 2016, pages 230-238 (cf. pages 60, 124).
- [134] Y. FU, Y. LEI, T. WANG, P. PATEL, A. B. JANI, H. MAO, W. J. CURRAN, T. LIU et X. YANG. « Biomechanically constrained non-rigid MR-TRUS prostate registration using deep learning based 3D point cloud matching ». In : *Medical Image Analysis* 67 (2021), page 101845. issn : 1361-8415 (cf. pages 61, 62).
- [135] Z. M. BAUM, Y. HU et D. C. BARRATT. « Real-time multimodal image registration with partial intraoperative point-set data ». In : *Medical image analysis* 74 (2021), page 102231 (cf. pages 62, 63).
- [136] Y. HU, M. MODAT, E. GIBSON, N. GHAVAMI, E. BONMATI, C. M. MOORE, M. EMBERTON, J. A. NOBLE, D. C. BARRATT et T. VERCAUTEREN. « Label-driven weakly-supervised learning for multimodal deformable image registration ». In : *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. 2018, pages 1070-1074 (cf. pages 64, 65).
- [137] M. JADERBERG, K. SIMONYAN, A. ZISSERMAN et al. « Spatial transformer networks ». In : *Advances in neural information processing systems* 28 (2015) (cf. page 65).
- [138] Z. LIU, Y. LIN, Y. CAO, H. HU, Y. WEI, Z. ZHANG, S. LIN et B. GUO. « Swin transformer : Hierarchical vision transformer using shifted windows ». In : *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pages 10012-10022 (cf. page 66).
- [139] V. FEOKTISTOV. *Differential evolution*. Springer, 2006 (cf. page 68).
- [140] S. ANAM. « Rainfall prediction using backpropagation algorithm optimized by Broyden-Fletcher-Goldfarb-Shanno algorithm ». In : *IOP Conference Series : Materials Science and Engineering*. Tome 567. 1. IOP Publishing. 2019, page 012008 (cf. page 68).

- [141] M. RONCHETTI, W. WEIN, N. NAVAB, O. ZETTINIG et R. PREVOST. « Disa : Differentiable similarity approximation for universal multimodal registration ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pages 761-770 (cf. page 68).
- [142] Y. FU, Y. LEI, T. WANG, W. J. CURRAN, T. LIU et X. YANG. « Deep learning in medical image registration : a review ». In : *Physics in Medicine & Biology* 65.20 (2020), 20TR01 (cf. page 72).
- [143] X. LIU, L. SONG, S. LIU et Y. ZHANG. « A review of deep-learning-based medical image segmentation methods ». In : *Sustainability* 13.3 (2021), page 1224 (cf. page 72).
- [144] X. CHEN, X. WANG, K. ZHANG, K.-M. FUNG, T. C. THAI, K. MOORE, R. S. MANNEL, H. LIU, B. ZHENG et Y. QIU. « Recent advances and clinical applications of deep learning in medical image analysis ». In : *Medical Image Analysis* 79 (2022), page 102444. issn : 1361-8415 (cf. page 72).
- [145] A. S. PANAYIDES, A. AMINI, N. D. FILIPOVIC, A. SHARMA, S. A. TSAFTARIS, A. YOUNG, D. FORAN, N. DO, S. GOLEMATI, T. KURC et al. « AI in medical imaging informatics : current challenges and future directions ». In : *IEEE journal of biomedical and health informatics* 24.7 (2020), pages 1837-1857 (cf. page 72).
- [146] F. ANDREY, K. SIAVASH, S. C. ANTONIO, L. ANDRAS, S. F. SIDNEY, T. KEMAL, N. S. EMILY, K. TINA, Z. CHENXI, W. WILLIAM M., N. PAUL L., A. PURANG et T. CLARE M. « Open-source image registration for MRI-TRUS fusion-guided prostate interventions ». In : *International Journal of Computer Assisted Radiology and Surgery* 10 (2015), pages 925-934 (cf. pages 73, 74).
- [147] V. DE LUCA, T. BENZ, S. KONDO, L. KÖNIG, D. LÜBKE, S. ROTHLÜBBERS, O. SOMPHONE, S. ALLAIRE, M. A. L. BELL, D. Y. F. CHUNG et al. « The 2014 liver ultrasound tracking benchmark ». In : *Physics in Medicine & Biology* 60.14 (2015), page 5571 (cf. page 73).
- [148] M. RISCINTI, M. MACIAS, C. SOLARI et J. KILPATRICK. « The POCUS Atlas - A Novel Crowdsourced Ultrasound Archive ». In : *Western Journal of Emergency Medicine : Integrating Emergency Care with Population Health* (janvier 2018) (cf. page 73).
- [149] Y. XU, B. ZHENG, X. LIU, T. WU, J. JU, S. WANG, Y. LIAN, H. ZHANG, T. LIANG, Y. SANG et al. « Improving artificial intelligence pipeline for liver malignancy diagnosis using ultrasound images and video frames ». In : *Briefings in Bioinformatics* 24.1 (2023), bbac569 (cf. page 74).

- [150] H. SIEBERT, L. HANSEN et M. P. HEINRICH. « Learning a Metric for Multimodal Medical Image Registration without Supervision Based on Cycle Constraints ». In : *Sensors* 22.3 (2022). issn : 1424-8220 (cf. page 74).
- [151] K. CLARK, B. VENDT, K. SMITH, J. FREYMAN, J. KIRBY, P. KOPPEL, S. MOORE, S. PHILLIPS, D. MAFFITT, M. PRINGLE et al. « The Cancer Imaging Archive (TCIA) : maintaining and operating a public information repository ». In : *Journal of digital imaging* 26 (2013), pages 1045-1057 (cf. page 74).
- [152] Y. JI, H. BAI, J. YANG, C. GE, Y. ZHU, R. ZHANG, Z. LI, L. ZHANG, W. MA, X. WAN et al. « AMOS : A Large-Scale Abdominal Multi-Organ Benchmark for Versatile Medical Image Segmentation ». In : *arXiv preprint arXiv :2206.08023* (2022) (cf. page 74).
- [153] A. E. KAVUR, N. S. GEZER, M. BARIS, S. ASLAN, P.-H. CONZE, V. GROZA, D. D. PHAM, S. CHATTERJEE, P. ERNST, S. OZKAN, B. BAYDAR, D. LACHINOV, S. HAN, J. PAULI, F. ISENSEE, M. PERKONIGG, R. SATHISH, R. RAJAN, D. SHEET, G. DOVLETOV, O. SPECK, A. NURNBERGER, K. H. MAIER-HEIN, G. {BOZDAGI AKAR}, G. UNAL, O. DICLE et M. A. SELVER. « CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation ». In : *Medical Image Analysis* 69 (2021), page 101950. issn : 1361-8415 (cf. pages 74, 75).
- [154] O. A. J. del TORO, O. GOKSEL, B. MENZE, H. MÜLLER, G. LANGS, M.-A. WEBER, I. EGGEL, K. GRUENBERG, M. HOLZER, G. KOTSIOS-KONTOKOTSIOS et al. « VISCERAL-VISual Concept Extraction challenge in RAdioLogY : ISBI 2014 challenge organization ». In : *Proceedings of the VISCERAL Challenge at ISBI* 1194 (2014), pages 6-15 (cf. page 74).
- [155] O. JIMENEZ-DEL-TORO, H. MULLER, M. KRENN, K. GRUENBERG, A. A. TAHA, M. WINTERSTEIN, I. EGGEL, A. FONCUBIERTA-RODRIGUEZ, O. GOKSEL, A. JAKAB, G. KONTOKOTSIOS, G. LANGS, B. H. MENZE, T. SALAS FERNANDEZ, R. SCHAER, A. WALLEYO, M.-A. WEBER, Y. DICENTE CID, T. GASS, M. HEINRICH, F. JIA, F. KAHL, R. KECHICHIAN, D. MAI, A. B. SPANIER, G. VINCENT, C. WANG, D. WYETH et A. HANBURY. « Cloud-Based Evaluation of Anatomical Structure Segmentation and Landmark Detection Algorithms : VISCERAL Anatomy Benchmarks ». In : *IEEE Transactions on Medical Imaging* 35.11 (2016), pages 2459-2475 (cf. page 74).
- [156] M. GRAMMATIKOPOULOU, E. FLOUTY, A. KADKHODAMOHAMMADI, G. QUELLEC, A. CHOW, J. NEHME, I. LUENGO et D. STOYANOV. « CaDIS : Cataract dataset for surgical RGB-image segmentation ». In : *Medical Image Analysis* 71 (2021), page 102053 (cf. page 75).

- [157] A. BUSTOS, A. PERTUSA, J.-M. SALINAS et M. de la IGLESIA-VAYÁ. « Padchest : A large chest x-ray image dataset with multi-label annotated reports ». In : *Medical image analysis* 66 (2020), page 101797 (cf. page 75).
- [158] X. LUO, W. LIAO, J. XIAO, J. CHEN, T. SONG, X. ZHANG, K. LI, D. N. METAXAS, G. WANG et S. ZHANG. « WORD : A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from CT image ». In : *Medical Image Analysis* 82 (2022), page 102642 (cf. page 75).
- [159] P. BILIC, P. CHRIST, H. B. LI, E. VORONTSOV, A. BEN-COHEN, G. KAISSIS, A. SZESKIN, C. JACOBS, G. E. H. MAMANI, G. CHARTRAND et al. « The liver tumor segmentation benchmark (lits) ». In : *Medical Image Analysis* 84 (2023), page 102680 (cf. page 75).
- [160] M. AUBREVILLE, N. STATHONIKOS, C. A. BERTRAM, R. KLOPFLEISCH, N. TER HOEVE, F. CIOMPI, F. WILM, C. MARZAHN, T. A. DONOVAN, A. MAIER et al. « Mitosis domain generalization in histopathology images—The MIDOG challenge ». In : *Medical Image Analysis* 84 (2023), page 102699 (cf. page 75).
- [161] A. GUPTA, S. GEHLOT, S. GOSWAMI, S. MOTWANI, R. GUPTA, Á. G. FAURA, D. ŠTEPEC, T. MARTINČIČ, R. AZAD, D. MERHOF et al. « SegPC-2021 : A challenge & dataset on segmentation of Multiple Myeloma plasma cells from microscopic images ». In : *Medical Image Analysis* 83 (2023), page 102677 (cf. page 75).
- [162] Q. DA, X. HUANG, Z. LI, Y. ZUO, C. ZHANG, J. LIU, W. CHEN, J. LI, D. XU, Z. HU et al. « DigestPath : A benchmark dataset with challenge review for the pathological detection and segmentation of digestive-system ». In : *Medical Image Analysis* 80 (2022), page 102485 (cf. page 75).
- [163] F. ANDREY, B. REINHARD, K.-C. JAYASHREE, F. JULIEN, J. C. FILLION-ROBIN, P. SONIA, B. CHRISTIAN, J. DOMINIQUE, F. FIONA, S. MILAN, B. JOHN, A. STEPHEN, J. V. MILLER, P. STEVE et K. RON. « 3D Slicer as an image computing platform for the Quantitative Imaging Network ». English (US). In : *Magnetic Resonance Imaging* 30.9 (novembre 2012). Funding Information : We would like to thank all current and past users and developers of 3D Slicer for their contribution to this software. The authors have been supported in part by the following NIH grants. BWH : U01CA151261, P41EB015898, P41RR13218, U54EB005149 and 1R01CA111288 ; University of Iowa : U01-CA140206 ; GE : P41RR13218 and U54EB005149 ; MGH : 1U01CA154601-01 and 4R00LM009889-03 . We are grateful to the various agencies and programs that funded support and development of 3D Slicer over the years., pages 1323-1341. issn : 0730-725X (cf. page 76).

- [164] G. YANG, G. LI, T. PAN, Y. KONG, J. WU, H. SHU, L. LUO, J.-L. DILLENSEGER, J.-L. COATRIEUX, L. TANG et al. « Automatic segmentation of kidney and renal tumor in ct images based on 3d fully convolutional neural network with pyramid pooling module ». In : *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE. 2018, pages 3790-3795 (cf. page 77).
- [165] C. E. DANIEL Alexander J. and Buchanan, T. ALLCOCK, D. SCERRI, E. F. COX, B. L. PRESTWICH et S. T. FRANCIS. « Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network ». In : *Magnetic Resonance in Medicine* 86.2 (2021), pages 1125-1136 (cf. page 77).
- [166] S. WARFIELD, K. H. ZOU et W. M. WELLS. « Simultaneous truth and performance level estimation (STAPLE) : an algorithm for the validation of image segmentation ». In : *IEEE Transactions on Medical Imaging* 23 (2004), pages 903-921 (cf. pages 77, 78).
- [167] R. E. ONG, C. L. GLISSON, S. D. HERRELL, M. I. MIGA et R. GALLOWAY. « A deformation model for non-rigid registration of the kidney ». In : *Medical Imaging 2009 : Visualization, Image-Guided Procedures, and Modeling*. Tome 7261. SPIE. 2009, pages 1022-1030 (cf. pages 78, 93).
- [168] P. SRIRAMAKRISHNAN, T. KALAISELVI, S. PADMAPRIYA, N. SHANTHI, S. RAMKUMAR et N. KALAICHELVI. « An medical image file formats and digital image conversion ». In : *Int. J. Eng. Adv. Technol* 9.1S3 (2019), pages 74-78 (cf. page 80).
- [169] T. LEWINER, H. LOPES, A. W. VIEIRA et G. TAVARES. « Efficient implementation of marching cubes' cases with topological guarantees ». In : *Journal of graphics tools* 8.2 (2003), pages 1-15 (cf. pages 81, 86, 92).
- [170] M.-P. DUBUISSON et A. K. JAIN. « A modified Hausdorff distance for object matching ». In : *Proceedings of 12th International Conference on Pattern Recognition*. Tome 1. 1994, 566-568 vol.1 (cf. page 82).
- [171] Ö. ÇIÇEK, A. ABDULKADIR, S. S. LIENKAMP, T. BROX et O. RONNEBERGER. « 3D U-Net : Learning Dense Volumetric Segmentation from Sparse Annotation ». In : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Sous la direction de S. OURSELIN, L. JOSKOWICZ, M. R. SABUNCU, G. UNAL et W. WELLS. Cham : Springer International Publishing, 2016, pages 424-432. ISBN : 978-3-319-46723-8 (cf. page 84).
- [172] F. MILLETARI, N. NAVAB et S.-A. AHMADI. « V-Net : Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation ». In : *2016 Fourth*

- International Conference on 3D Vision (3DV)*. 2016, pages 565-571 (cf. page 84).
- [173] Z. SHEN, C. LIN et S. ZHENG. « COTR : Convolution in Transformer Network for End to End Polyp Detection ». In : *2021 7th International Conference on Computer and Communications (ICCC)* (2021), pages 1757-1761 (cf. page 84).
- [174] J. C. REINHOLD, B. E. DEWEY, A. CARASS et J. L. PRINCE. « Evaluating the impact of intensity normalization on MR image synthesis ». In : *Medical Imaging 2019 : Image Processing*. Tome 10949. SPIE. 2019, pages 890-898 (cf. page 85).
- [175] N. JACOBSEN, A. DEISTUNG, D. TIMMANN, S. L. GOERICKE, J. R. REICHENBACH et D. GÜLLMAR. « Analysis of intensity normalization for optimal segmentation performance of a fully convolutional neural network ». In : *Zeitschrift für Medizinische Physik* 29.2 (2019), pages 128-138 (cf. page 85).
- [176] M. J. CARDOSO, W. LI, R. BROWN, N. MA, E. KERFOOT, Y. WANG, B. MURREY, A. MYRONENKO, C. ZHAO, D. YANG, V. NATH, Y. HE, Z. XU, A. HATAMIZADEH, A. MYRONENKO, W. ZHU, Y. LIU, M. ZHENG, Y. TANG, I. YANG, M. ZEPHYR, B. HASHEMIAN, S. ALLE, M. Z. DARESTANI, C. BUDD, M. MODAT, T. VERCAUTEREN, G. WANG, Y. LI, Y. HU, Y. FU, B. GORMAN, H. JOHNSON, B. GENEREUX, B. S. ERDAL, V. GUPTA, A. DIAZ-PINTO, A. DOURSON, L. MAIER-HEIN, P. F. JAEGER, M. BAUMGARTNER, J. KALPATHY-CRAMER, M. FLORES, J. KIRBY, L. A. D. COOPER, H. R. ROTH, D. XU, D. BERICAT, R. FLOCA, S. K. ZHOU, H. SHUAIB, K. FARAHANI, K. H. MAIER-HEIN, S. AYLWARD, P. DOGRA, S. OURSELIN et A. FENG. *MONAI : An open-source framework for deep learning in healthcare*. 2022 (cf. page 85).
- [177] S. DU, G. XU, S. ZHANG, X. ZHANG, Y. GAO et B. CHEN. « Robust rigid registration algorithm based on pointwise correspondence and correntropy ». In : *Pattern Recognition Letters* 132 (2020), pages 91-98 (cf. page 92).
- [178] T. HORSTMANN, O. ZETTINIG, W. WEIN et R. PREVOST. « Orientation Estimation of Abdominal Ultrasound Images with Multi-Hypotheses Networks ». In : *Medical Imaging with Deep Learning*. 2022 (cf. page 92).
- [179] I. FIGUEROA-GARCIA, J.-M. PEYRAT, G. HAMARNEH et R. ABUGHARBIEH. « Biomechanical kidney model for predicting tumor displacement in the presence of external pressure load ». In : *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2014, pages 810-813 (cf. page 93).
- [180] W. SCHROEDER, K. MARTIN et B. LORENSEN. *The Visualization Toolkit (4th ed.)*. Kitware, 2006. ISBN : 978-1-930934-19-1 (cf. page 93).

- [181] A. LEROY, P. MOZER, Y. PAYAN et J. TROCCAZ. « Intensity-based registration of freehand 3D ultrasound and CT-scan images of the kidney ». In : *International journal of computer assisted radiology and surgery* 2.1 (2007), pages 31-41 (cf. page 96).
- [182] D. L. PHAM, C. XU et J. L. PRINCE. « Current methods in medical image segmentation ». In : *Annual review of biomedical engineering* 2.1 (2000), pages 315-337 (cf. page 106).
- [183] F. ISENSEE, T. WALD, C. ULRICH, M. BAUMGARTNER, S. ROY, K. MAIER-HEIN et P. F. JAEGER. « nnU-Net Revisited : A Call for Rigorous Validation in 3D Medical Image Segmentation ». In : *arXiv preprint arXiv :2404.09556* (2024) (cf. page 107).
- [184] W. E. LORENSEN et H. E. CLINE. « Marching cubes : A high resolution 3D surface construction algorithm ». In : *Seminal graphics : pioneering efforts that shaped the field*. 1998, pages 347-353 (cf. page 107).
- [185] K. S. ARUN, T. S. HUANG et S. D. BLOSTEIN. « Least-squares fitting of two 3-D point sets ». In : *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), pages 698-700 (cf. page 108).
- [186] W. NDZIMBONG, C. FOURNIOL, L. THEMYR, N. THOME, Y. KEEZA, B. SAUER, P.-T. PIECHAUD, A. MEJEAN, J. MARESCAUX, D. GEORGE et al. « TRUSTED : The Paired 3D Transabdominal Ultrasound and CT Human Data for Kidney Segmentation and Registration Research ». In : *arXiv preprint arXiv :2310.12646* (2023) (cf. page 113).
- [187] X. WANG, X. YANG, H. DOU, S. LI, P.-A. HENG et D. NI. « Joint segmentation and landmark localization of fetal femur in ultrasound volumes ». In : *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. 2019, pages 1-5 (cf. page 124).
- [188] K. K. LIANG. « Efficient conversion from rotating matrix to rotation axis and angle by extending Rodrigues' formula ». In : *arXiv preprint arXiv :1810.02999* (2018) (cf. page 129).
- [189] M. TAN et Q. V. LE. « EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks ». In : *ICML*. 2019 (cf. page 131).
- [190] R. CARUANA. « Multitask Learning ». In : *Machine Learning* 28 (1997), pages 41-75 (cf. page 132).