



**HAL**  
open science

# Navigation en temps-réel de biopsie prostatique assistée par intelligence artificielle

Tamara Dupuy

► **To cite this version:**

Tamara Dupuy. Navigation en temps-réel de biopsie prostatique assistée par intelligence artificielle. Santé. Université Grenoble Alpes [2020-..], 2023. Français. NNT : 2023GRALS016 . tel-04185929

**HAL Id: tel-04185929**

**<https://theses.hal.science/tel-04185929>**

Submitted on 23 Aug 2023

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

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

THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement

Spécialité : BIS - Biotechnologie, instrumentation, signal et imagerie  
pour la biologie, la médecine et l'environnement

Unité de recherche : Translational Innovation in Medicine and Complexity

### Navigation en temps-réel de biopsie prostatique assistée par intelligence artificielle

### AI Assistance for Real-Time Prostate Biopsy Navigation

Présentée par :

**Tamara DUPUY**

#### Direction de thèse :

**Sandrine VOROS**

Directrice de Recherche, INSERM, TIMC

**Jocelyne TROCCAZ**

Directrice de Recherche, CNRS, TIMC

**Clément BEITONE**

Maître de Conférences, Université Grenoble Alpes, TIMC

Directrice de thèse

Co-directrice de thèse

Co-encadrant de thèse

#### Rapporteurs :

**Alain LALANDE**

MCU-PH, Université de Bourgogne

**Antoine SIMON**

Maître de Conférences, Université de Rennes, LTSI

#### Thèse soutenue publiquement le 13 juin 2023, devant le jury composé de :

**Gaëlle FIARD**

PU-PH, Université Grenoble Alpes

**Alain LALANDE**

MCU-PH, Université de Bourgogne

**Antoine SIMON**

Maître de Conférences, Université de Rennes, LTSI

**Carole LARTIZIEN**

Directrice de Recherche, CNRS, CREATIS

**Ingerid REINERTSEN**

Associate Professor, Norwegian University, SINTEF

**Sandrine VOROS**

Directrice de Recherche, INSERM, TIMC

**Jocelyne TROCCAZ**

Directrice de Recherche, CNRS, TIMC

Présidente

Rapporteur

Rapporteur

Examinatrice

Examinatrice

Directrice

Co-directrice

#### Invités :

**Clément BEITONE**

Maître de Conférences, Université Grenoble Alpes, TIMC





---

## Remerciements

---

Ce manuscrit conclut un travail mené après quatre belles années de thèse sur un sujet que j'ai trouvé particulièrement innovant, riche et stimulant. Je suis très reconnaissante d'avoir vécu une telle aventure, qui m'a permise un véritable épanouissement personnel, d'autant plus favorisé par les nombreuses personnes qui m'ont accompagnées et que je souhaite remercier ici.

Tout d'abord, j'aimerais remercier chaleureusement mon équipe encadrante : mes deux directrices de thèse Jocelyne Troccaz et Sandrine Voros, mon encadrant Clément Beitone. Un trio "de choc".

Si leurs compétences scientifiques et techniques ne font plus aucun doute à travers leurs carrières ponctuées de réussites en tout genre, je tiens à souligner que ce sont également des personnes d'exception qui ont égayé mon travail au quotidien en plus de contribuer à la bonne réussite de cette thèse. Merci pour votre entente joyeuse et contagieuse, votre répartie aussi mordante qu'hilarante, vos valeurs nobles et engagées, votre appétence pour la rigueur scientifique, vos conseils si précieux et surtout votre soutien indéfectible avec toujours beaucoup de bienveillance. Je suis très chanceuse de vous avoir eus à mes côtés et je vous serai toujours reconnaissante d'avoir cru en moi, plus que ce que j'étais capable de le faire moi-même.

Un merci tout particulier à Clément, qui a été bien plus qu'un encadrant et co-bureau. Merci pour chaque moment partagé, pour m'avoir écoutée et comprise, et surtout pour m'avoir guidée et conseillée. Comme je le dis souvent, je n'aurais sans doute pas pu réussir cette thèse sans toi.

Je tiens à remercier tous les membres du jury (rapporteurs et examinatrices) pour avoir accepté d'évaluer ce travail de thèse et pour le temps et l'intérêt qu'ils ont porté à mes travaux. Merci pour les retours de vos rapports détaillés et constructifs, vos remarques avisées et pour les discussions intéressantes que nous avons échangées tous ensemble. Je suis très reconnaissante d'avoir pu interagir avec des scientifiques de votre expertise et de retenir cette journée de soutenance comme particulièrement riche et agréable.

Je remercie également tous les membres du TIMC, et plus particulièrement des équipes très complices de GMCAO et de BIOMECA, que j'ai pu côtoyer chaque jour. Je garde de très bons souvenirs de ces nombreuses pauses café sujettes aux dérives et débats, de mes intrusions intempe-

---

tives dans vos bureaux, des midi-annexes illégales malgré mes retards fréquents, des compétitions de goûters, des sorties escalade, ou encore des soirées arrosées autour de bretzel et d'empanadas. J'ai pu rencontrer des personnes qui sont aujourd'hui de véritables amis.

Je ne peux évoquer cette thèse sans remercier toutes les personnes sans qui ce travail n'aurait pas pu être possible : que ce soit le laboratoire, l'école doctorale, les financeurs, ou les nombreux collaborateurs. Je remercie en particulier le service d'urologie du CHU de Grenoble pour leur aide et conseils précieux concernant les enjeux et usages cliniques et pour l'accès aux données cliniques. Merci à l'industriel Koelis qui a également suivi l'évolution de cette thèse et qui je l'espère continuera à valoriser ce travail. Merci aux différentes étudiantes que j'ai pu rencontrer et encadrer durant leur stage, cela a été un réel plaisir de travailler avec vous en plus d'être une expérience enrichissante.

Je terminerai en adressant mon infinie gratitude et toute ma tendresse à mon compagnon, mes amis, ma famille et ma belle-famille. Ils sont ma source d'apaisement et de bien-être, ainsi que ma dose de rire et de bonheur au quotidien. Merci pour la richesse de vos personnalités et merci de faire partie de ma vie.

*À mes amis,  
bientôt confrères et consœurs,  
rencontrés pendant ces belles années à Grenoble*



---

# Table des matières

---

<b>Acronymes</b>	<b>viii</b>
<b>I Introduction</b>	<b>1</b>
I.1 Contexte clinique . . . . .	2
I.1.1 Cancer de la prostate : un problème de santé publique majeure . . . . .	2
I.1.2 Dépistage, diagnostic et thérapie . . . . .	2
I.1.3 Difficultés liées aux procédures de biopsies . . . . .	4
I.2 Contexte de ces travaux . . . . .	7
I.2.1 Nos objectifs . . . . .	7
I.2.2 Réalisation de ces travaux . . . . .	9
I.2.3 Organisation du manuscrit . . . . .	10
<b>II État-de-l’art</b>	<b>13</b>
II.1 Rappels des concepts du recalage d’images . . . . .	15
II.1.1 Méthodes de recalage « classique » . . . . .	16
II.1.2 Méthodes de recalage « par apprentissage profond » . . . . .	19
II.2 Évolution des avancées . . . . .	21
II.2.1 Recalage 3D/3D avec fusion IRM pour le repérage et/ou la cartographie . . . . .	21
II.2.2 Recalage 2D/3D pour le guidage en temps-réel . . . . .	25
II.3 Travaux de recalage « 2D/3D » par des approches IA . . . . .	31
II.3.1 Une seule coupe en entrée . . . . .	31
II.3.2 Coupe et volume en entrée . . . . .	34
II.3.3 Séquence de coupes en entrée . . . . .	36
II.4 Analyse et synthèse de l’état-de-l’art . . . . .	40
II.4.1 Types d’entrée . . . . .	40
II.4.2 Types de sortie . . . . .	43
II.4.3 Types de structure . . . . .	45

II.4.4	Types d'optimisation . . . . .	48
II.4.5	Qualité des données et d'évaluation . . . . .	50
II.5	Notre positionnement . . . . .	53
<b>III</b>	<b>Mise en place d'un protocole de validation expérimentale</b>	<b>55</b>
III.1	Génération de données . . . . .	57
III.1.1	Définitions et notations des données nécessaires à notre étude . . . . .	57
III.1.2	La base de données clinique disponible . . . . .	60
III.1.3	Simulation de trajectoire . . . . .	62
III.1.4	Entrées optionnelles : représentation de la trajectoire réalisée . . . . .	76
III.2	Protocole de validation . . . . .	79
III.2.1	Métriques . . . . .	79
III.2.2	Évaluation séquentielle sur une trajectoire complète . . . . .	82
<b>IV</b>	<b>Recalage 2D/3D avec des informations positionnelles</b>	<b>85</b>
IV.1	Objectifs et contributions . . . . .	87
IV.2	Matériel & Méthodes . . . . .	88
IV.2.1	Approche proposée . . . . .	88
IV.2.2	Apprentissage . . . . .	90
IV.3	Expériences et configurations . . . . .	91
IV.3.1	Étude d'ablation . . . . .	91
IV.3.2	Étude d'interprétabilité . . . . .	92
IV.3.3	Comparaison des encodages . . . . .	92
IV.3.4	Jeux de données . . . . .	93
IV.4	Résultats . . . . .	95
IV.4.1	Étude d'ablation . . . . .	95
IV.4.2	Étude d'interprétabilité : biais des zones noires . . . . .	100
IV.4.3	Variation des caractéristiques du jeu de données . . . . .	102
IV.5	Discussion . . . . .	104
IV.5.1	Discussion sur les contributions méthodologiques . . . . .	104
IV.5.2	Discussion sur les contributions expérimentales . . . . .	106
IV.5.3	Conclusions sur ce premier modèle . . . . .	107
<b>V</b>	<b>Recalage 2D/3D avec des informations spatiales</b>	<b>111</b>
V.1	Objectifs et contributions . . . . .	113
V.2	Matériel & Méthodes . . . . .	114
V.2.1	Approche proposée . . . . .	114
V.2.2	Apprentissage . . . . .	117
V.3	Expériences et configurations . . . . .	118
V.3.1	Différentes formes d'informations spatiales . . . . .	118
V.3.2	Différentes architectures . . . . .	119
V.3.3	Jeux de données . . . . .	120
V.3.4	Comparaison avec l'état-de-l'art . . . . .	121
V.3.5	Paramètres d'apprentissage . . . . .	121

V.4	Résultats . . . . .	122
V.4.1	Différentes formes d'informations spatiales . . . . .	122
V.4.2	Différentes architectures . . . . .	125
V.4.3	Comparaison avec l'état-de-l'art . . . . .	126
V.5	Discussion . . . . .	131
V.5.1	Discussion sur les contributions méthodologiques . . . . .	131
V.5.2	Discussion sur les contributions expérimentales . . . . .	134
V.5.3	Conclusions sur ce deuxième modèle . . . . .	135
<b>VI</b>	<b>Recalage 2D/3D avec des informations structurales</b>	<b>137</b>
VI.1	Objectifs et contributions . . . . .	139
VI.2	Matériel & Méthodes . . . . .	140
VI.2.1	Approche proposée . . . . .	140
VI.2.2	Apprentissage . . . . .	142
VI.3	Expériences et configurations . . . . .	144
VI.3.1	Étude d'ablation . . . . .	144
VI.3.2	Différentes architectures testées . . . . .	145
VI.3.3	Jeux de données . . . . .	146
VI.3.4	Paramètres d'apprentissage . . . . .	148
VI.3.5	Métriques de segmentation . . . . .	148
VI.4	Résultats . . . . .	149
VI.4.1	Analyse quantitative . . . . .	149
VI.4.2	Analyse temporelle . . . . .	151
VI.4.3	Analyse qualitative . . . . .	154
VI.4.4	Pondération et apprentissage . . . . .	154
VI.5	Discussion . . . . .	157
VI.5.1	Discussion sur les contributions méthodologiques . . . . .	157
VI.5.2	Conclusions sur ce troisième modèle . . . . .	159
<b>VII</b>	<b>Conclusion</b>	<b>161</b>
VII.1	Rappel des contributions . . . . .	162
VII.1.1	Contributions méthodologiques . . . . .	163
VII.1.2	Contributions expérimentales . . . . .	164
VII.2	Perspectives . . . . .	166
VII.2.1	Vers une nouvelle formulation du problème . . . . .	166
VII.2.2	Vers une application clinique . . . . .	168
VII.3	Publication et communication . . . . .	169
<b>Annexe A</b>	<b>Architecture détaillée des réseaux proposés</b>	<b>178</b>
<b>Annexe B</b>	<b>Comparaison des encodages des transformations</b>	<b>183</b>
<b>Bibliographie</b>		<b>IX</b>



---

## Table des acronymes

---

### Général

GT	Ground truth
GPU	Graphics Processing Unit
ITK	Insight Toolkit
DoF	Degree of Freedom
NIH	National Institute of Health
IMU	Inertial Measurement Uni
PSA	Prostate Specific Antigen
ROI	Region of Interest

### Métriques

CE	Cross-Entropy
BCE	Binary Cross-Entropy
DSC	Dice Similarity Coefficient
MSE	Mean Square Error
MSD	Mean Surface Distance
HD	Hausdorff Distance
TRE	Target Registration Error
NCC	Normalized Cross Correlation
PCC	Pearson Cross Correlation
SSIM	Structural Similarity Index Measure
R2	R-score
CD	Chamfer Distance
BE	Bending Energy
MI	Mutual Information
L2	L2-Norm
GE	Geodesic Error

### Imageries

US	Ultrasound
TRUS	Transrectal Ultrasound
IRM	Magnetic Resonance Imaging
CT	Computed tomography
HIFU	High Intensity Focalised Ultrasounds
PET	Positron emission tomography

### Type d'apprentissage

SL	Supervised Learning
WSL	Weakly Supervised Learning
SSL	Self Supervised Learning
MTL	Multi-Task Learning
DDF	Dense Displacement Field

### Structure de réseau

CNN	Convolution Neural Network
STN	Spatial Transform Network
LSTM	Long Short-Term Memory
ReLU	Rectified Linear Unit
FC	Fully Connected
MLP	Multi Layers Perceptron
PC	Point Cloud

### Méthodes

FEM	Finite Element Method
PCA	Principal Component Analysis
RANSAC	Random Sample Consensus
ICP	Iterative Closest Point
SGD	Stochastic Gradient Descent



---

## Introduction

---

*Ce* chapitre présente le contexte clinique dans lequel s'inscrit ce travail de thèse.

Dans un premier temps, les aspects épidémiologiques actuels du cancer de la prostate sont présentés avec les différentes démarches de diagnostic et de traitements pour la prise en charge de cette maladie. En particulier, nous décrivons le déroulement d'un examen de biopsies prostatiques et l'importance d'un ciblage de qualité durant celui-ci. Les contraintes cliniques lors de la réalisation de cet examen sont mises en avant afin d'introduire, dans un deuxième temps, les solutions déjà existantes et les défis toujours d'actualité. Enfin, nous présentons les principales études et contributions menées tout au long de cette thèse pour répondre aux problématiques traitées.

### Sommaire

---

<b>I.1</b>	<b>Contexte clinique</b>	<b>2</b>
I.1.1	Cancer de la prostate : un problème de santé publique majeure	2
I.1.2	Dépistage, diagnostic et thérapie	2
I.1.3	Difficultés liées aux procédures de biopsies	4
<b>I.2</b>	<b>Contexte de ces travaux</b>	<b>7</b>
I.2.1	Nos objectifs	7
I.2.2	Réalisation de ces travaux	9
I.2.3	Organisation du manuscrit	10

---

## I.1 Contexte clinique

### I.1.1 Cancer de la prostate : un problème de santé publique majeure

La prostate est une glande de l'appareil génital masculin, située sous la vessie et en avant du rectum. De forme quasi-sphérique d'environ 4 cm de diamètre, sa fonction principale est de sécréter une partie du liquide séminal, l'un des constituants du sperme, et de le stocker dans les vésicules séminales. L'anatomie de la prostate peut être définie en plusieurs zones [1] centrale, antérieure, de transition et périphérique (voir Figure I.1). On définit également « la base » comme zone de contact supérieur avec la vessie, et « l'apex » pour la zone inférieure. C'est principalement la zone périphérique (postérieure), facilement palpable au toucher rectal, qui est le siège de développement des cancers de la prostate (environ 70%) [2].

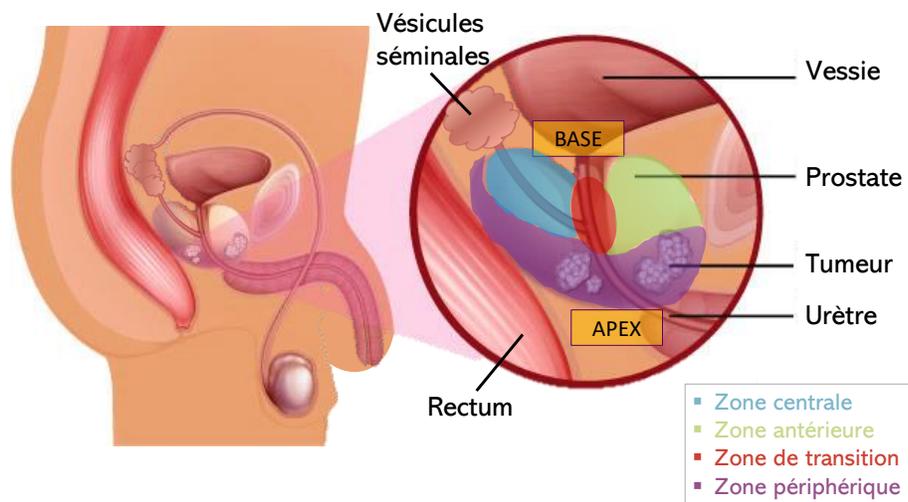


FIGURE I.1 – Anatomie de la prostate.

Avec plus de 1.4 million de nouveaux cas de cancers de la prostate diagnostiqués et environ 375 000 décès recensés dans le monde en 2020, le cancer de la prostate est classé au 2ème rang des cancers les plus fréquents chez l'homme (après le cancer du poumon) et est la 5ème cause de décès par cancer chez l'homme [3] (voir Figure I.2). En France, chaque année, plus de 65 000 nouveaux cas de cancers de la prostate sont recensés, et environ 9 000 hommes en décèdent. Sans facteur de risque réellement identifié, ce cancer survient généralement chez des hommes âgés de 65 ans et plus, avec un âge médian au moment du décès de 83 ans. Depuis plusieurs années, on assiste à une diminution régulière de sa mortalité [4] qui s'explique notamment par l'amélioration des traitements et par l'accès au dépistage précoce de ce cancer.

### I.1.2 Dépistage, diagnostic et thérapie

La maladie ayant quelquefois une progression très lente, sans manifester le moindre symptôme, on préconise un dépistage individuel à partir de 50 ans. En premier lieu, on recommande une prise de sang pour connaître le taux de PSA (antigène prostatique spécifique), qui est un marqueur de

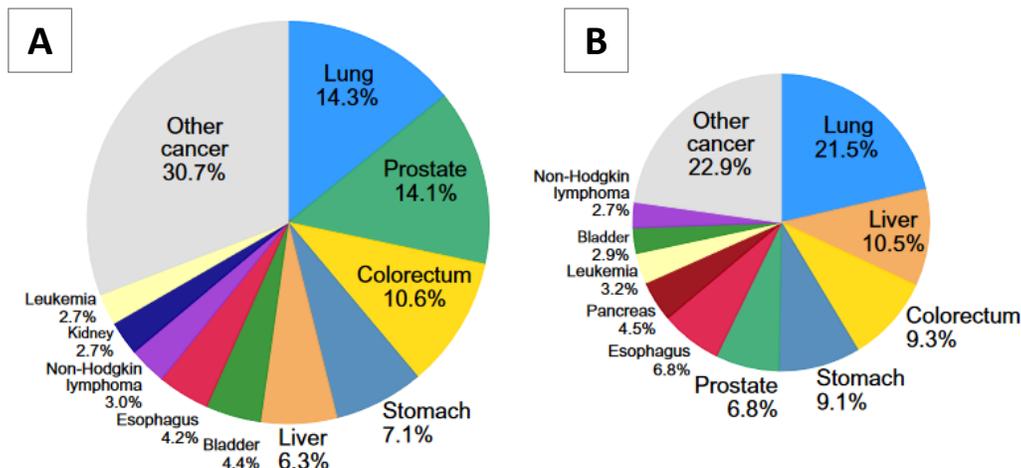


FIGURE I.2 – Répartition des nouveaux cas (A) et des décès (B) pour les 10 cancers les plus fréquents en 2020 pour les hommes. Source : GLOBOCAN 2020 [3].

pathologie prostatique [5]. En cas de suspicion de cancer, le médecin peut effectuer un toucher rectal visant à palper la prostate et détecter d’éventuelles irrégularités de forme ou de consistance. Cependant, cet examen est souvent insuffisant pour garantir une absence de tumeur à un stade précoce.

Lorsqu’une anomalie a été détectée, seul un examen de biopsie peut confirmer avec précision le diagnostic de cancer. Cet examen consiste en plusieurs prélèvements de tissus sur lesquels sont réalisés des tests histologiques pour révéler la présence éventuelle de cellules cancéreuses et en déterminer la gravité (Score de Gleason). Le diagnostic peut être encore affiné par le biais de différentes imageries médicales [5, 6], par exemple : échographie Doppler, scanner (“Computed-Tomography” ou “Positron Emission Tomography”), imagerie par résonance magnétique (IRM) ou une scintigraphie osseuse. Certaines sont illustrées sur la Figure I.3. L’IRM permet notamment d’identifier et d’évaluer avec précision l’étendue du cancer, tandis que la scintigraphie osseuse et le PET-scan sont utilisés pour détecter d’éventuelles métastases sur les os ou d’autres organes. En cas de suspicion de cancer dans une zone spécifique de la prostate, des biopsies supplémentaires pourront alors y être réalisées.

Le traitement à privilégier dépend des caractéristiques de la maladie, de son stade et son évolution, de l’âge, des potentielles contre-indications et de l’état général du patient. Les traitements les plus répandus sont l’ablation chirurgicale, une irradiation par radiothérapie externe ou interne (curiethérapie), une thérapie hormonale ou encore une chimiothérapie [5, 6]. Les risques d’impuissance et d’incontinence sont les principales complications communes à toutes les thérapies proposées [7], en plus des risques liés à une chirurgie traditionnelle. Ces complications sont assez fréquentes et sont à évaluer en fonction du rapport bénéfices/risques pour le patient. Les traitements focaux se présentent de plus en plus comme une alternative prometteuse pour traiter uniquement les tissus cancéreux et limiter les risques liés aux traitements radicaux [8]. Les plus utilisés actuellement sont l’HIFU (“High Intensity Focalised Ultrasounds”), la cryothérapie, la photothérapie dynamique, l’électroporation irréversible, et la thermothérapie par laser. Certaines de ces techniques sont encore en évaluation clinique.

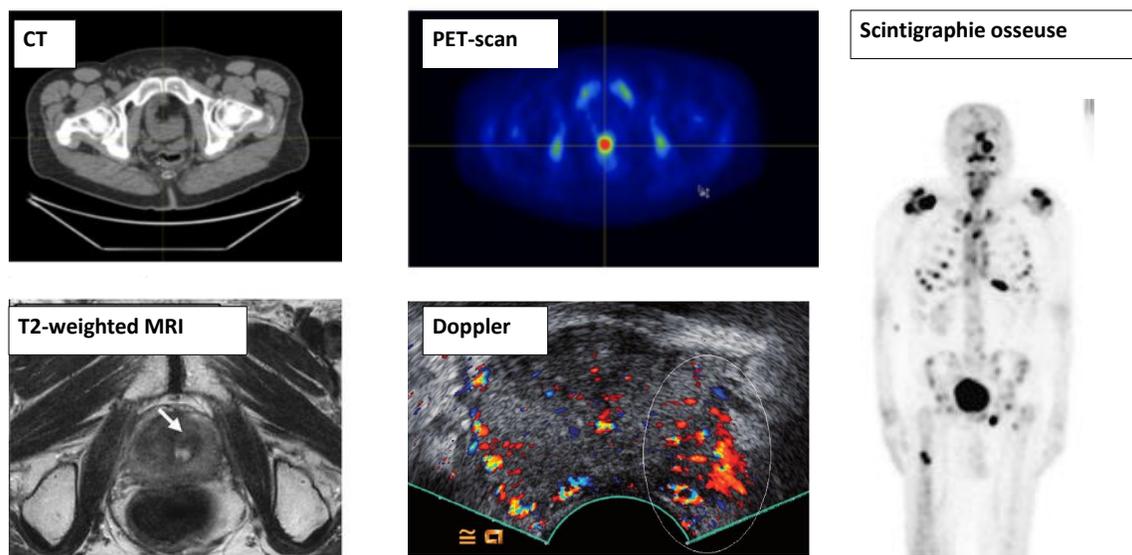


FIGURE I.3 – Différents types d'imagerie utilisée dans le cadre de la détection du cancer de la prostate : Scintigraphie [9] ; Doppler [10] ; PET et CT [11].

La capacité à détecter et à localiser avec précision la zone cancéreuse dans la glande est un enjeu majeur pour le diagnostic et l'évaluation du stade d'un éventuel cancer afin de prendre les bonnes décisions thérapeutiques et ainsi éviter le sur/sous traitement. La localisation est d'autant plus cruciale pour les thérapies focales, afin d'appliquer le traitement efficacement et localement. Cette localisation ne provient uniquement de la localisation et de l'identification des prélèvements effectués durant l'examen de biopsie. Cependant, dans la pratique clinique courante, il n'existe pas de systèmes qui permettent de visualiser, précisément et en temps réel, la position des biopsies. Le bon repérage et ciblage durant cet examen restent un défi actuel majeur.

### I.1.3 Difficultés liées aux procédures de biopsies

Depuis leur introduction en 1989 [12, 13], les biopsies sous guidage échographique ultrasonore (US) sont devenues l'examen de référence pour diagnostiquer le cancer de la prostate. Elles se sont largement imposées face aux procédures de biopsies par guidage direct IRM, présentant de nombreuses limitations logistiques (disponibilité, délai, coût) ou pratiques (accessibilité au patient, confort, durée, compatibilité électromagnétique) [14].

Une biopsie est un examen qui consiste à venir prélever plusieurs échantillons de tissus, et à les faire analyser par un anatomopathologiste. Les biopsies prostatiques se réalisent, le plus souvent, par voie transrectale (TRUS), sous anesthésie locale et sous protection antibiotique. Une sonde échographique est placée dans le rectum du patient, sur laquelle est fixé un guide servant à introduire l'aiguille, nécessaire pour les prélèvements de tissus. Depuis quelques années, la voie trans-périnéale est également de plus en plus utilisée. Celle-ci se réalise au bloc opératoire, sous anesthésie générale, et au moyen d'une aiguille fine traversant le périnée. Bien qu'elles permettent une diminution des

risques d'infection [2, 15] et une meilleure accessibilité à certaines régions de la glande, ces biopsies trans-périnéales sont cependant plus complexes à mettre en œuvre et restent minoritaires en clinique. Nous ne les aborderons pas davantage dans ce manuscrit.

Les biopsies se font majoritairement de manière « standardisée / systématisée » avec des prélèvements de tissus (carottes) répartis de manière homogène sur la glande selon un schéma systématique standard illustré en Figure I.4. Ce schéma répartit 12 biopsies de la base à l'apex de la prostate, échantillonnant la partie médiane et latérale, sur chacune des moitiés droite et gauche de la glande. Ce schéma a évolué durant les dernières décennies, passant de 6 à 12 carottes, afin de réduire le nombre de conclusions altérées de diagnostic « faux négatif » [2, 15, 16]. Les biopsies peuvent être aussi ciblées dans le cas où une zone suspecte a été localisée, au préalable, par une imagerie préopératoire (généralement IRM). Enfin, l'ensemble de ces prélèvements est labellisé, selon le secteur visé, avant d'être envoyés en analyse anatomo-pathologique. On espère alors obtenir un diagnostic fiable permettant d'évaluer la présence, l'étendue et l'agressivité d'un éventuel cancer et de proposer une stratégie thérapeutique adaptée.

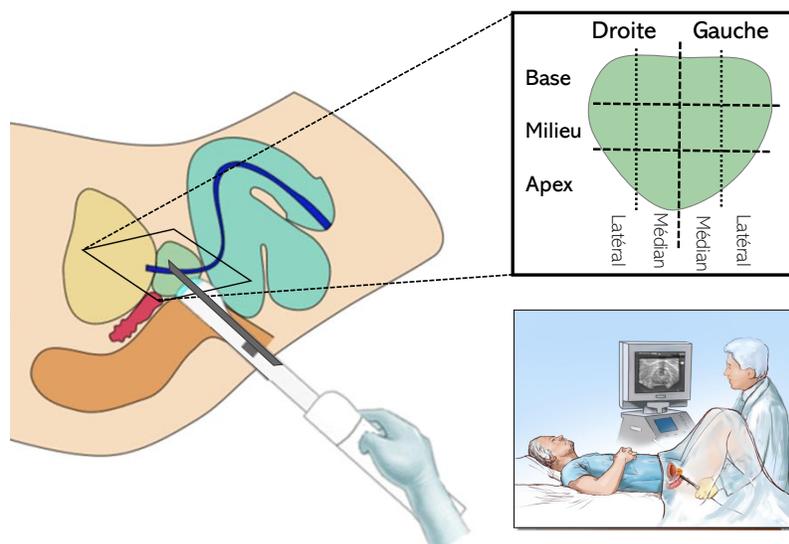


FIGURE I.4 – Procédure de biopsie de prostate par voie transrectale et représentation des 12 zones d'échantillonnage « standardisées/systématisées ».

Exceptée l'information apportée par la visualisation de l'imagerie de guidage peropératoire échographique et son expérience, le praticien n'a aucun autre retour ou confirmation sur la position de la sonde par rapport à la position du prélèvement ciblé. Durant la procédure, deux éléments rendent le repérage et la localisation des biopsies difficiles : la modalité bidimensionnelle échographique couramment utilisée, les potentiels bougés ou déformations anatomiques.

D'une part, les images échographiques sont de qualité souvent variable (bruitée, speckle, artefacts) avec mauvaise sensibilité des ultrasons pour distinguer les tissus cancéreux des tissus sains. De plus, l'aspect bidimensionnel (2D) de cette modalité limite la visualisation de la scène tridimensionnelle (3D) dans laquelle évolue le praticien. La reconstruction mentale 3D pour visualiser l'organe et le déplacement à réaliser est d'autant plus complexe que la prostate est un organe quasi-sphérique

et symétrique.

La prostate est, d'autre part, un organe mobile et déformable [17, 18], qui se déforme en fonction des mouvements induits par le patient (respiration, bougés) ou par le praticien (pression et déplacements de la sonde échographique, et insertion de l'aiguille). Ces bougés et déformations modifient l'anatomie de la prostate, et donc les positions des zones que l'on cherche à atteindre.

Toutes ces difficultés combinées, perturbent la bonne représentation et la cartographie mentale du praticien durant l'examen. Sans une parfaite prise en compte des déplacements et des déformations de la prostate, le repérage et le ciblage des biopsies sont alors altérés. Ceci se traduit par un risque accru de manquer les zones cibles (environ 30% de faux négatifs estimés [2, 15, 16]), et par une mauvaise association des résultats histologiques avec la localisation des prélèvements testés pouvant compromettre le diagnostic ainsi que les décisions thérapeutiques. Ces difficultés se traduisent également par un allongement de la durée de l'examen et une augmentation de l'inconfort du patient.

Une étude récente [19] a tenté d'illustrer les conséquences liées aux difficultés durant la procédure de biopsie sur la position des biopsies prélevées. La Figure I.5 présente les résultats de différentes cartographies de biopsies, réalisées par 5 urologues seniors experts (expérience supérieure à 100 examens) devant respecter un schéma systématisé en sextant, sur un fantôme de prostate. La variabilité de distribution des biopsies ainsi que la surprise des experts au moment des résultats, reflètent bien la difficulté à réaliser avec précision des biopsies prostatiques.

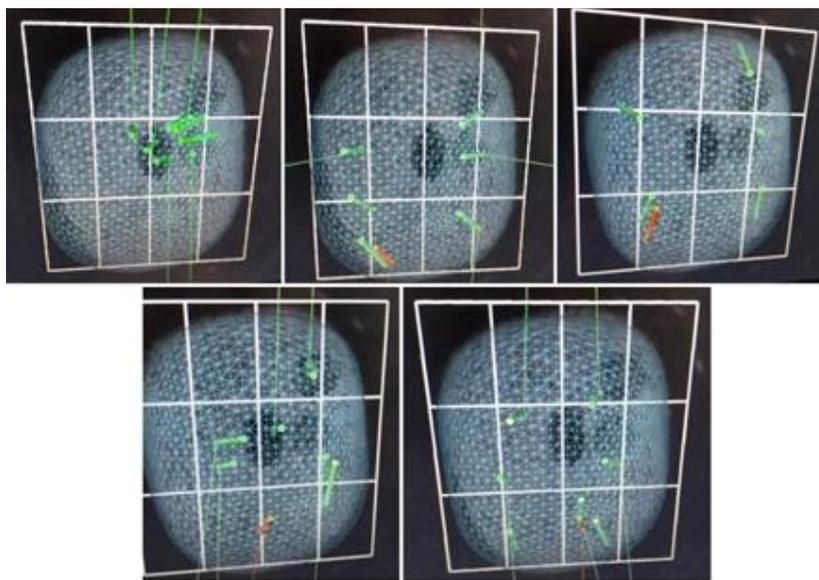


FIGURE I.5 – Illustration de la diversité de cartographies des biopsies, réalisées par les 5 urologues experts.

## I.2 Contexte de ces travaux

### I.2.1 Nos objectifs

Le principal objectif de cette thèse est de concevoir un dispositif d'assistance à la navigation, permettant à l'urologue de mieux se repérer et ainsi mieux cibler les biopsies. Nous définissons la navigation comme la suite des gestes réalisés par l'urologue, muni de la sonde, pour atteindre la position du prélèvement souhaité. Un tel dispositif devra permettre à l'urologue de visualiser en continu la position du tir qui serait effectué pour la position courante de la sonde, par rapport à une référence anatomique fixe et connue, le tout en tenant compte des bougés et déformations de la cible. Les enjeux cliniques sont multiples : favoriser le repérage des biopsies, faciliter les procédures, améliorer la qualité du geste, raccourcir la courbe d'apprentissage des praticiens, et permettre le recours à des traitements focaux.

Pour cela, nous nous sommes basés sur des méthodes de traitement d'images, et plus précisément sur du recalage 2D(US)/3D(US), permettant de localiser la coupe échographique courante (de l'imagerie peropératoire) par rapport à un volume de référence (pris en début d'examen), en alignant au mieux le contenu anatomique de la prostate. Toutes les images ultrasonores de ce manuscrit seront traitées dans le mode-B. Le principe de l'approche est illustré par la Figure I.6.

La condition « temps réel » est définie en lien avec la fréquence d'acquisition de l'imagerie échographique (entre 10-20 Hz), soit un temps de réponse inférieur à 50 ms. Les besoins en termes de « précision » de localisation des biopsies sont définis à partir d'un consensus médical, estimant la taille de la plus petite tumeur traitable à 0.5 cc (volume d'une sphère d'environ 5 mm de rayon) [20,21]. Cela fixe donc des valeurs de seuil importantes pour évaluer la qualité d'un tel dispositif d'assistance lors d'un échantillonnage de biopsies.

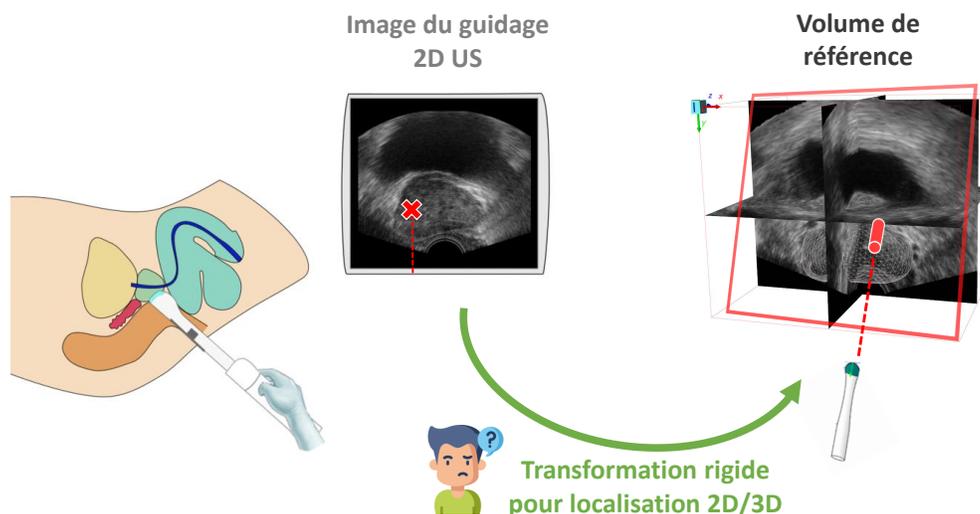


FIGURE I.6 – Dispositif d'assistance à la navigation permettant à l'urologue de visualiser la position de la biopsie en cours de réalisation.

Différentes dénominations sont à définir et à retenir pour tout le reste de ce manuscrit afin de différencier les types de mouvements occasionnés durant une procédure de biopsie de routine. Ceux-ci sont définis ci-dessous et illustrés dans la Figure I.7.

- Tout d’abord, le repère « sonde »  $R_{sonde}$  est associé à la position mobile de la sonde échographique. Il est possible de repérer la sonde dans l’espace (localisation absolue) et de suivre ses déplacements dans le temps (suivi relatif). Des localisateurs externes de type optique (Polaris, etc.) ou électromagnétique (“EM-tracking”) permettent de localiser la position absolue de la sonde par rapport aux coordonnées du référentiel physique ( $R_{phys}$ ). D’autres capteurs de type inertiel (IMU de l’anglais “Inertial Measurement Unit”), plus économique, simple et compact, peuvent mesurer le déplacement relatif de la sonde par rapport à une position initiale.
- Le repère « échographique »  $R_{US}$  est associé à l’image acquise par la sonde, durant le guidage échographique. Nous nous concentrons sur des sondes échographiques "end-fire" adaptées pour des biopsies par voie transrectale, où l’aiguille passe par un guide solidaire de cette sonde (voir Fig.I.4). La sonde, de taille fixe et connue, peut être calibrée de façon à connaître la position de  $R_{US}$  par rapport à  $R_{sonde}$ . La position de l’aiguille, et donc de la biopsie, pourraient donc être connues s’il n’y avait pas d’autres déplacements supplémentaires en jeu.
- Le repère  $R_{prostate}$ , associé à la position de la prostate, est celui dont on cherche à définir la position. Que ce soit la sonde qui se déplace autour de la prostate, ou la prostate qui bouge par rapport à une position fixe de la sonde à cause de mouvements induits par le patient (respiration, bougés), la relation entre le repère de la prostate et celui de la sonde est difficile à établir et à maintenir. De plus, la prostate est un organe qui se déforme en fonction de la pression de la sonde et/ou de l’aiguille. Ces déformations (non-linéaires) au sein du même repère de la prostate, complexifient d’autant plus les positions anatomiques de la glande.

Depuis plusieurs années, des dispositifs d’assistance existent pour améliorer la connaissance des positions des biopsies voire pour guider le clinicien vers des zones spécifiques. Pour l’essentiel, les solutions combinent des dispositifs de mesure de position de la sonde échographique et des méthodes de traitement d’image. Ces solutions ont été classées en deux approches de recalage qui diffèrent selon leur principe de localisation : les méthodes basées « suivi de sonde » se concentrent uniquement sur la localisation de la sonde mais ne prennent pas en compte les mouvements et/ou déformation de l’anatomie durant la procédure. Les méthodes basées « suivi d’organe » se concentrent bien sur la localisation de l’organe au sein de l’image acquise, cependant l’acquisition de l’imagerie et les temps de calcul ne permettent pas une localisation en temps-réel. De plus en plus de méthodes visent à combiner les deux types de localisation pour trouver un meilleur compromis entre rapidité et précision. Ces méthodes seront décrites plus en détails dans le chapitre suivant.

Afin de répondre aux limites de l’existant, nous proposons dans cette thèse :

- de développer des outils d’assistance innovants mettant en œuvre des techniques de recalage basées sur de l’intelligence artificielle (notamment par apprentissage profond ou « deep learning »), fortement adaptées au besoin du temps-réel et à la quantité d’informations présentes lors d’un traitement d’images.
- de considérer des informations liées à la trajectoire réalisée, pouvant faciliter la tâche de recalage durant la navigation. Diverses façons d’exploiter des informations de trajectoire (images,

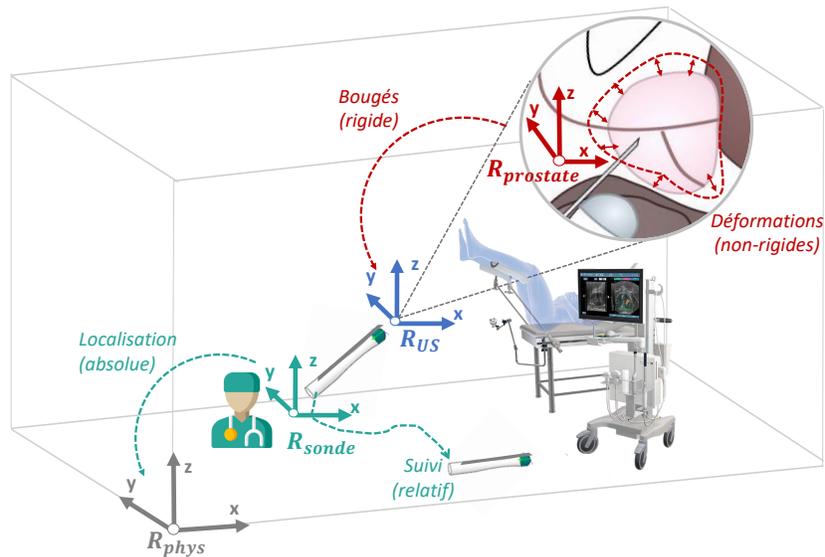


FIGURE I.7 – Visualisation et dénomination des différents mouvements induits par le praticien (déplacements de la sonde échographique), par patient (respiration, bougés) ou par insertion de l’aiguille (pression, déformations).

capteurs, etc.) et de les intégrer dans une architecture profonde sont proposés dans cette thèse.

- de respecter les conditions cliniques de réalisme et de faisabilité de notre application. Cela s’illustre notamment au travers des données utilisées et par la validation des résultats.

## I.2.2 Réalisation de ces travaux

Cette thèse a été réalisée au sein de l’équipe GMCAO (Gestes Médico-Chirurgicaux Assistés par Ordinateur), du laboratoire TIMC (UMR CNRS 5525 – Université Grenoble Alpes). Spécialisée dans les thématiques de traitement d’images médicales et robotique médicale, l’équipe développe des méthodes d’assistance au geste médico-chirurgical depuis 1985 en collaboration avec des nombreux partenaires industriels et d’équipes cliniques.

Les travaux proposés ont été conduits en partenariat étroit avec le service d’urologie du CHU de Grenoble Alpes (Dr Gaëlle Fiard notamment) afin que les solutions développées soient pensées en respectant au plus près les contraintes et usages cliniques. Ce travail s’appuie également sur la collaboration avec l’industriel KOELIS, spécialisé dans la chirurgie guidée par l’image en urologie, qui a valorisé plusieurs travaux de l’équipe GMCAO au travers notamment des plateformes de cartographie échographiques (Urostation<sup>®</sup> et Trinity<sup>®</sup>).

Les financements des travaux liés à cette thèse sont partagés entre plusieurs acteurs :

- L’institut d’Intelligence Artificielle MIAI (3IA Grenoble-Alpes), créé en France en 2019, via le programme “Investissement d’Avenir” de l’ANR (sous la référence ANR-19-P3IA-0003) dans le contexte de la chaire « CAMI Assistant ».
- Le Labex CAMI (sous la référence ANR-11-LABX-0004)

- La région Auvergne Rhône-Alpes (projet ProNavIA) pour le développement de systèmes de navigation des biopsies de la prostate basée sur l'imagerie et l'IA.
- BPI France (projet PSPC DIANA), pour le développement des nouveaux systèmes d'assistance à la prise en charge des cancers de la prostate.

### I.2.3 Organisation du manuscrit

Ce manuscrit résume les travaux réalisés, portant sur l'étude de méthodes d'apprentissage profond dans le but d'assister en temps réel la navigation de biopsies prostatiques. Il s'articule autour de 7 chapitres suivant l'organisation suivante :

- Le Chapitre II présente une vue d'ensemble des travaux de la littérature s'approchant de nos objectifs de travail de thèse. Après quelques rappels des concepts liés à la théorie du recalage en imagerie médicale, le chapitre se concentre sur l'évolution des différentes méthodes d'assistance au repérage et au guidage des biopsies prostatiques, puis au « recalage 2D/3D » en général. Les travaux sont classés selon différentes catégories, avant de conclure quant à notre positionnement et nos contributions proposées dans ce contexte scientifique.
- Le Chapitre III détaille notre protocole de génération de données et de validation expérimentale. Après avoir listé l'ensemble des données nécessaires à notre étude et les éléments des bases de données cliniques dont nous disposons, le chapitre décrit l'évolution de trois différents procédés de simulation avec leurs avantages et inconvénients respectifs. Un protocole d'évaluation séquentielle, à partir d'évaluation de recalages successifs le long d'une trajectoire complète, est proposé et mis en perspective face aux réalismes de déroulement clinique.
- Le Chapitre IV présente une première méthode de recalage 2D/3D, utilisant des a priori « positionnels » liés à la trajectoire réalisée. Différentes expériences sont réalisées selon l'ajout et la combinaison des données d'entrée, le type d'encodage des transformations, les fonctions de coût utilisées et les caractéristiques des jeux de données.
- Le Chapitre V propose une seconde version de méthode de recalage, qui exploite les informations contextuelles « spatiales » liées à l'espace de référence par rapport auquel on cherche à localiser les images peropératoires. Plusieurs types d'entrée spatiales sont testées et combinées à notre première méthode afin de trouver la configuration qui maximise au mieux les performances de recalage 2D/3D. Une comparaison avec d'autres méthodes similaires de la littérature est également réalisée.
- Le Chapitre VI présente une troisième évolution de nos travaux, en exploitant des informations « structurelles » qui permettent de considérer des a priori sur les contours de la prostate au sein des images. L'association de ces données d'entrée avec des méthodes d'apprentissage multi-tâche est notamment étudiée pour aller au-delà des méthodes précédentes. Un intérêt particulier est porté sur le type d'architecture adapté aux tâches conjointes et sur les techniques d'optimisation à privilégier.
- Le Chapitre VII résume l'évolution des différentes méthodes proposées, avec leurs contributions et limitations respectives. Des perspectives d'évolution sont évoquées, pour aller plus loin dans la recherche d'information contextuelle liée à la navigation et en vue d'une application clinique. Ce chapitre conclut ce manuscrit.

### Points clé

- ❖ Le cancer de la prostate est un problème de santé publique. Seul un examen de biopsie permet d'établir définitivement le diagnostic.
- ❖ Durant cet examen, le bon ciblage des biopsies reste un défi majeur de par la modalité bidimensionnelle échographiques utilisée, ainsi que les déformations et bougés de la prostate.
- ❖ Un mauvais ciblage des biopsies peut conduire à un mauvais diagnostic et de mauvaises décisions thérapeutiques.
- ❖ Le principal objectif de cette thèse est d'établir un dispositif d'assistance à la navigation, permettant à l'urologue de se repérer en continu et avec précision, et ainsi mieux repérer et cibler des biopsies.
- ❖ Les principales nouveautés de notre étude sont de considérer des informations liées à la trajectoire réalisée et de développer des approches basées sur l'intelligence artificielle, tout en respectant les conditions de réalisation clinique.



---

### État-de-l'art

---

*Ce* chapitre propose une vue d'ensemble des travaux de la littérature afin de positionner nos objectifs et notre problématique dans le contexte scientifique actuel.

Des dispositifs d'assistance existent pour améliorer la connaissance des positions des biopsies voire pour guider le clinicien vers des zones spécifiques. Ils se basent essentiellement sur des méthodes de traitement d'images et notamment de recalage d'images. Après quelques rappels des concepts liés à la théorie du recalage en imagerie médicale, nous montrerons l'évolution des différentes méthodes d'assistance au repérage et au guidage des biopsies prostatiques. Puis, nous étendrons nos recherches à tous les travaux concernant le « recalage 2D/3D », en général, présentant des approches méthodologiques intéressantes. Nous en proposerons une analyse et une synthèse, à travers une classification des différentes approches. Enfin, nous conclurons quant à notre positionnement et nos contributions proposées au travers de cette thèse.

---

## Sommaire

---

<b>II.1</b>	<b>Rappels des concepts du recalage d'images . . . . .</b>	<b>15</b>
II.1.1	Méthodes de recalage « classique » . . . . .	16
II.1.1.1	Les attributs caractéristiques . . . . .	16
II.1.1.2	Le critère de similarité . . . . .	16
II.1.1.3	Type de transformation géométrique . . . . .	17
II.1.1.4	Stratégie d'optimisation . . . . .	17
II.1.1.5	Évaluation du recalage . . . . .	18
II.1.2	Méthodes de recalage « par apprentissage profond » . . . . .	19
<b>II.2</b>	<b>Évolution des avancées . . . . .</b>	<b>21</b>
II.2.1	Recalage 3D/3D avec fusion IRM pour le repérage et/ou la cartographie . . . . .	21
II.2.1.1	Méthodes classiques pionnières . . . . .	22
II.2.1.2	Méthodes récentes . . . . .	22
II.2.2	Recalage 2D/3D pour le guidage en temps-réel . . . . .	25
II.2.2.1	Méthodes classiques pionnières . . . . .	25
II.2.2.2	Méthodes récentes . . . . .	26
<b>II.3</b>	<b>Travaux de recalage « 2D/3D » par des approches IA . . . . .</b>	<b>31</b>
II.3.1	Une seule coupe en entrée . . . . .	31
II.3.2	Coupe et volume en entrée . . . . .	34
II.3.3	Séquence de coupes en entrée . . . . .	36
<b>II.4</b>	<b>Analyse et synthèse de l'état-de-l'art . . . . .</b>	<b>40</b>
II.4.1	Types d'entrée . . . . .	40
II.4.1.1	Entrées positionnelles . . . . .	40
II.4.1.2	Entrées spatiales . . . . .	41
II.4.1.3	Entrées séquentielles . . . . .	42
II.4.1.4	Entrées structurelles . . . . .	42
II.4.2	Types de sortie . . . . .	43
II.4.2.1	Paramétrage des transformations . . . . .	43
II.4.2.2	Coordonnées . . . . .	44
II.4.2.3	Classification . . . . .	44
II.4.2.4	Incertitude . . . . .	44
II.4.3	Types de structure . . . . .	45
II.4.3.1	Structures hiérarchiques . . . . .	46
II.4.3.2	Structures dynamiques . . . . .	46
II.4.4	Types d'optimisation . . . . .	48
II.4.4.1	Auto-pénalisation . . . . .	48
II.4.4.2	Multi-tâches . . . . .	49
II.4.5	Qualité des données et d'évaluation . . . . .	50
II.4.5.1	Génération des données d'entrée . . . . .	50
II.4.5.2	Génération des vérités terrain . . . . .	51
II.4.5.3	Validation . . . . .	51
II.4.5.4	Déroulement . . . . .	52
<b>II.5</b>	<b>Notre positionnement . . . . .</b>	<b>53</b>

---

## II.1 Rappels des concepts du recalage d'images

Le recalage d'images consiste à aligner dans un même référentiel physique deux images, appelées image source  $I_{flottante}$  et image de référence  $I_{ref}$  (voir Figure II.1). Ces images peuvent représenter le même ou différents sujets, être issues de différentes modalités d'imagerie, prises à différents moments ou depuis différents points de vue. En imagerie médicale, ce recalage est utilisé pour différentes finalités cliniques selon les situations en jeu (mono ou multimodales, inter ou intra-patients, etc.). La Table II.1 en résume certaines de ces utilisations.

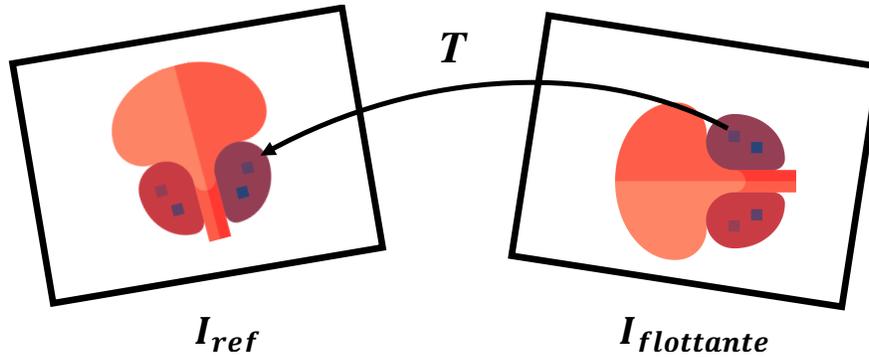


FIGURE II.1 – Alignement spatial durant un recalage d'images. Par convention, la transformation  $T$  est définie de l'image flottante (espace de départ) à l'image de référence (espace de destination).

TABLE II.1 – Différents exemples d'objectifs cliniques utilisant le recalage en imagerie médicale.

	Intra-patient	Inter-patients
Mono-modal	<ul style="list-style-type: none"> <li>• Suivi et surveillance de l'évolution d'une pathologie</li> <li>• Contrôle post-opératoire</li> </ul>	<ul style="list-style-type: none"> <li>• Construction d'atlas statistiques</li> <li>• Étude de variabilité au sein d'une population</li> </ul>
Multi-modal	<ul style="list-style-type: none"> <li>• Assistance au diagnostic</li> <li>• Planification préopératoire</li> <li>• Transfert d'informations complémentaire</li> </ul>	<ul style="list-style-type: none"> <li>• Superposition de modèle global (atlas, etc.)</li> </ul>

Les travaux concernant le recalage d'images sont très nombreux dans la littérature du traitement d'images médicales [22, 23]. On le retrouve notamment pour des applications de guidage d'intervention, de segmentation, de reconstruction d'images, ou encore de suivi de mouvement [24]. Dans cette partie, nous rappelons rapidement les concepts du recalage d'images. Nous présenterons d'abord le principe des méthodes classiques (c'est-à-dire utilisant des techniques classiques de traitement d'images), puis le principe des méthodes plus récentes apparues avec la démocratisation de l'apprentissage profond (ou « deep learning »).

### II.1.1 Méthodes de recalage « classique »

Le recalage est un problème d'optimisation dont l'objectif est d'estimer la transformation géométrique optimale  $T$ , alignant au mieux ces deux images. Les méthodes de recalage sont nombreuses et dépendent de plusieurs éléments : le choix des « caractéristiques » utilisées au sein des images, de la mesure de similarité  $S$  choisie pour exprimer le niveau de correspondance entre les images, et de la nature de la transformation à établir. Les parties suivantes détaillent les différents choix possibles pour chacune de ces étapes. Leur classification est résumée dans la Figure II.2.

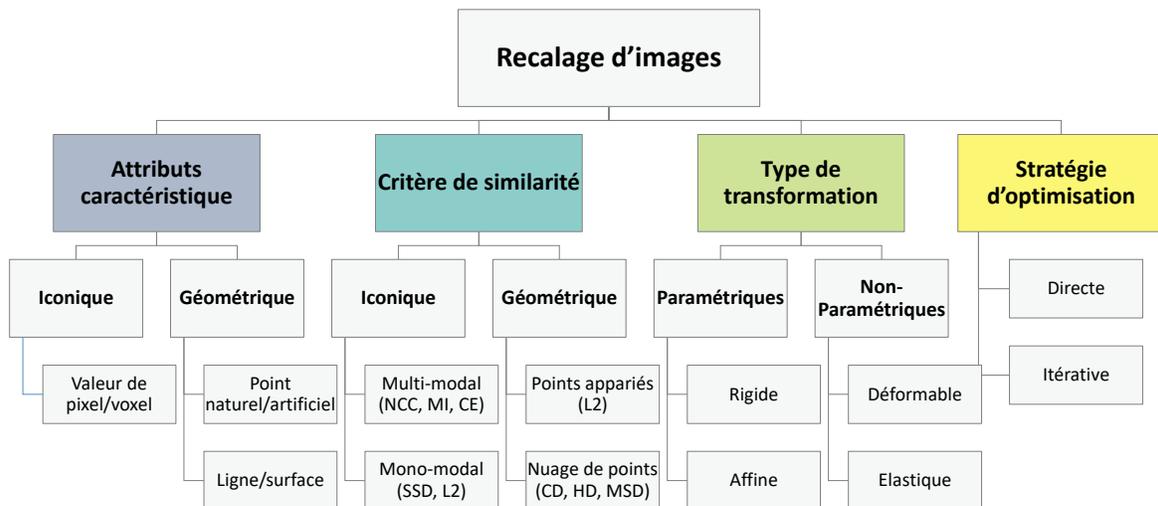


FIGURE II.2 – Classification générale des méthodes de recalage en fonction du type de caractéristiques, du critère de similarité, du type de transformation, et de la stratégie d'optimisation.

#### II.1.1.1 Les attributs caractéristiques

Les caractéristiques correspondent aux informations utiles et représentatives des deux images, qui seront utilisées pour optimiser la mesure de similarité. Il peut s'agir :

- D'attributs iconiques reflétant le contenu de l'image (valeurs de pixels/voxels).
- D'attributs géométriques reflétant des repères anatomiques ou structurels (points, courbes, surfaces, etc.). Ces attributs géométriques proviennent généralement d'éléments intrinsèques (naturels) ou extrinsèques (artificiellement ajoutés). L'extraction des attributs géométriques peut s'effectuer grâce à l'intervention manuelle d'experts ou grâce à des outils de détection automatique, et sont généralement réduits à un sous-ensemble plus facilement exploitable.

#### II.1.1.2 Le critère de similarité

Une fois défini le type d'information à utiliser pour guider le recalage, il s'agit de lui associer un critère de similarité permettant de quantifier la notion de ressemblance entre les deux images.

Les métriques iconiques comparent les intensités de pixel/voxels des images. Dans un cadre monomodal, on fait l'hypothèse que les deux images présentent des niveaux de contrastes similaires, où des mesures de différence classiques sont applicables. Dans un cadre multimodal, en revanche, l'hypothèse précédente n'est plus vérifiée. Les mesures utilisées se basent alors sur la distribution des intensités des deux images. Malgré leur pertinence, les mesures iconiques impliquent, en général, des temps de calcul longs où il est nécessaire de parcourir l'ensemble des voxels de l'image pour effectuer le calcul. De plus, elles ne permettent pas de prendre en compte l'organisation structurale de la zone d'intérêt.

Les métriques géométriques quantifient la distance entre les repères choisis. Pour des points déjà appariés et identifiés entre les deux images, une mesure de distance euclidienne est couramment utilisée. Pour des ensembles de points de type courbe ou surface, d'autres métriques de distance moyennées sur la totalité de l'ensemble sont alors utilisées. Même si ces méthodes sont plus rapides que les méthodes iconiques, elles sont cependant dépendantes de la qualité de la localisation des attributs caractéristiques, ainsi que leur bonne mise en correspondance.

Enfin, les méthodes hybrides combinent et pondèrent ces deux types de métriques afin d'exploiter les avantages liés à chacune.

### II.1.1.3 Type de transformation géométrique

Il s'agit désormais de définir un modèle de transformation. Ce choix conditionne la manière dont l'image flottante (de départ) sera géométriquement modifiée et alignée à la deuxième. Deux types de représentation sont possibles pour décrire les transformations géométriques : paramétriques ou non paramétriques

Les transformations paramétriques s'appliquent de manière unique et globale sur la totalité de l'image. Parmi celles-ci, on trouve les transformations rigides, les plus simples, qui sont constituées uniquement des paramètres de rotation et de translation. Les transformations affines, quant à elles, considèrent des paramètres supplémentaires d'étirement et de cisaillement. Quel que soit le type de transformations, elles peuvent être représentées sous plusieurs formes : coordonnées homogènes, matricielles, quaternions, angles d'Euler, etc.

Les transformations non-paramétriques sont non-linéaires et peuvent être appliquées de manière locale et indépendante à chaque voxel ou groupes de voxels. Ces transformations se décrivent sous la forme de champs de déplacement dense et vectoriel. Certaines peuvent même s'associer à des modèles mathématiques (physiques, statistiques, biomécaniques, etc.), on parle alors de transformations élastiques.

### II.1.1.4 Stratégie d'optimisation

Une fois définis les caractéristiques des images, le critère de similarité, et le type de transformation, il s'agit désormais de définir la stratégie d'optimisation la plus adaptée qui finalise ce processus de recalage. L'optimisation permet d'estimer la transformation optimale  $T$ , entre les deux images, en maximisant ou minimisant la mesure de similarité  $S$  sur les caractéristiques identifiées des images (voir eq.II.1b). Selon le nombre de paramètres à estimer et la complexité du problème, des stratégies directes ou itératives peuvent alors être établies. Lorsque c'est possible, une étape d'initialisation est également utilisée pour pré-aligner les deux images. Cette étape permet de réduire l'espace de

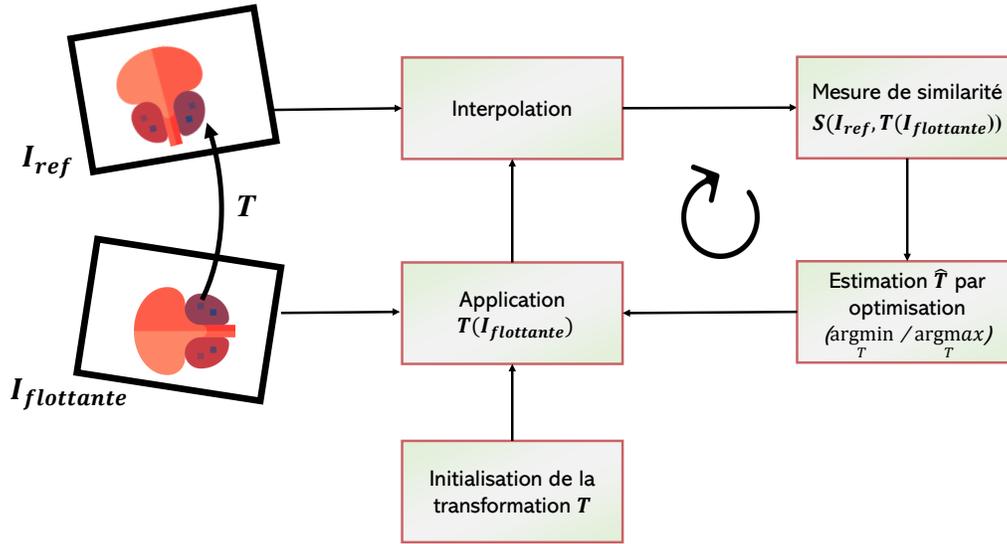


FIGURE II.3 – Processus itératif de recalage iconique d’images : à chaque itération, une nouvelle transformation  $\hat{T}$  est estimée et appliquée à l’image flottante (départ). Le rôle de l’interpolateur est de rééchantillonner l’image ainsi obtenue, pour adapter les pixels au système de coordonnées de destination (référence). La mesure de similarité est alors de nouveau calculée puis optimisée.

recherche et d’accélérer la convergence de l’optimisation tout en s’affranchissant des problématiques de minima locaux. Un exemple de processus itératif iconique est décrit en Figure II.3.

$$\hat{T} = \arg \max_T S(I_{ref}, T(I_{flottante})) \quad (\text{II.1a})$$

$$\text{ou } \arg \max_T S(I_{ref}, T(I_{flottante})) \quad (\text{II.1b})$$

### II.1.1.5 Évaluation du recalage

Enfin, l’évaluation de la qualité d’un recalage est une étape cruciale. Selon l’accessibilité, ou non, à une vérité terrain (noté GT de l’anglais “Ground-Truth”), différentes métriques sont utilisées.

Dans le cas de vérités terrain accessibles, il est possible de confronter les résultats obtenus par la méthode de recalage aux résultats des recalages déjà connus. On peut choisir de se concentrer sur les paramètres de transformation ou sur des caractéristiques géométriques identifiées. Des mesures statistiques, comme le coefficient de détermination ( $R^2$ ) ou encore la corrélation (PCC), sont très souvent employés pour analyser la qualité de la régression entre les paramètres de recalage estimés et théoriques. La TRE (“Target Registration Error”) permet, quant à elle, de mesurer la différence d’alignement entre plusieurs points repères de l’image (aussi appelés « fiduciels »). Ceux-ci peuvent être définis de manière manuelle ou automatique, et doivent être indépendants du processus de recalage [25]. Sa simplicité et sa rapidité de calcul font d’elle une métrique particulièrement utilisée dans la littérature du recalage [24, 26]. En revanche, pour une mesure robuste, elle nécessite d’avoir suffisamment de points, dont les localisations sont fiables et reproductibles entre le même ou plusieurs experts.

La définition de vérités terrain pouvant être un processus long et fastidieux, imprécis et parfois même irréalisable, d'autres mesures sont aussi utilisées pour mesurer la qualité du recalage. Des mesures de similarité du contenu iconique de l'image, comme la NCC ("Normalized Cross-Correlation") ou SSIM ("Structural Similarity Index Measure") sont très répandues pour comparer deux images avant et après recalage. D'autres encore, utilisent des critères de cohérence inverse [27], où la prédiction d'une transformation d'un point A vers un point B doit être inversible, et dont la transformée inverse doit être égale à celle trouvée lors du calcul de la transformation de ce même point B vers A. Cela permet de vérifier la robustesse de la prédiction de recalage en sens inverse. Enfin, une évaluation qualitative par des experts reste l'évaluation la plus répandue lorsqu'une mesure quantitative vérité terrain ne peut être établie rigoureusement.

Les problématiques liées à la difficulté d'évaluer le recalage restent encore un enjeu actuel, dont nous aurons l'occasion de discuter dans les chapitres III et VII.

Malgré leur utilisation très répandue, ces méthodes de recalage souffrent néanmoins de nombreuses limites : (i) des problèmes de convergence vers des minima locaux, (ii) temps de calcul importants, (iii) nécessité d'une bonne initialisation, (iv) choix difficile pour définir la métrique de similarité la plus pertinente selon la problématique (multimodale, multi-dimensionnelle) ou encore (v) la qualité de l'extraction de points caractéristiques dans le cadre du recalage géométrique. Les méthodes classiques utilisées dans le cadre du repérage et du guidage des biopsies prostatiques seront détaillées dans la section suivante (voir section II.2).

### II.1.2 Méthodes de recalage « par apprentissage profond »

Depuis ces dernières années, le « deep learning » est de plus en plus utilisé dans le cadre du traitement d'images [28], et en particulier pour des applications de recalage [24, 26, 29, 30].

L'apprentissage profond est un sous-domaine de l'intelligence artificielle, qui utilise des réseaux de neurones à plusieurs couches, pour créer des modèles algorithmiques capables d'apprendre à effectuer des tâches à partir de données (d'apprentissage), et d'en généraliser l'exécution sur de nouvelles données. Durant sa phase d'apprentissage, le réseau cherche à minimiser une fonction de coût qui évalue la qualité du recalage courant, comme le ferait une technique de recalage classique sur une mesure de similarité. En revanche, la technique d'optimisation va, au travers de la rétropropagation, mettre à jour les poids du réseau et non pas la transformation en elle-même. À travers cette action indirecte, les sorties du réseau (prédictions) sont mises à jour et améliorées successivement pendant tout le processus d'entraînement. Une fois le réseau entraîné, il pourra être appliqué de manière directe sur de nouvelles données, pour réaliser la prédiction de recalage en une seule étape.

Dans un premier temps, les réseaux cherchaient à estimer automatiquement les mesures de similarité qui pouvaient s'avérer difficiles à construire manuellement, notamment dans un cadre multimodal. Au fur et à mesure, des travaux ont cherché à estimer directement les paramètres de transformation à appliquer pour recalibrer les images. Les approches ont été majoritairement supervisées dans un premier temps, mais face à la difficulté de générer et/ou de labelliser de telles données, on voit apparaître de nouvelles méthodes semi-supervisées voire non-supervisées. Ainsi, la plupart des revues sur les techniques d'apprentissage profond pour du recalage listent les différentes méthodes au travers de ces 5 familles d'apprentissage : mesure de similarité (SIM), supervisées (SL),

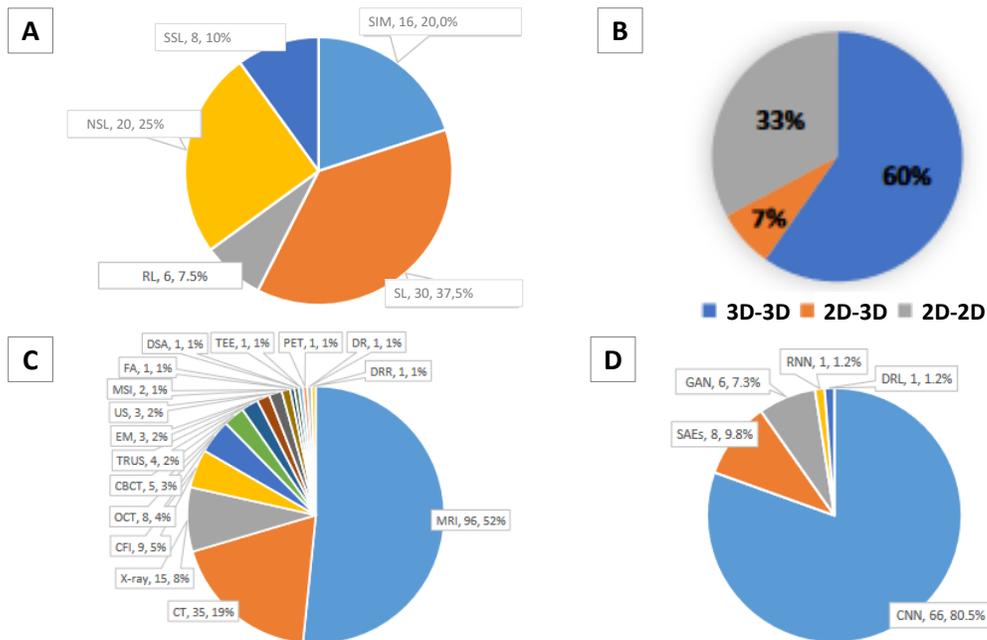


FIGURE II.4 – Fréquence d'apparition des méthodes de recalage d'images médicales par apprentissage profond dans la littérature selon : (A) la famille d'apprentissage, (B) la dimensionalité traitée, (C) la modalité des images et (D) l'architecture du réseau. Les graphes sont issus des revues [24, 26].

non-supervisées (NSL), auto-supervisées (SSL), et par renforcement (RL). La Figure II.4 illustre et catalogue les différentes méthodes de recalage d'images traitées ainsi que leur fréquence d'apparition dans la littérature selon leur modalité, dimension, méthode d'apprentissage et réseau utilisé. On remarque que la plupart des méthodes se basent sur de l'apprentissage supervisé en utilisant des réseaux de convolution.

L'apprentissage profond a souvent réussi à dépasser les limites du recalage classique concernant les problèmes de convergence/stagnation, de temps de calcul, et d'extraction de caractéristiques. Cependant, les méthodes d'apprentissage profond sont également sujettes à des limites, notamment au besoin d'une grande base de données (pas toujours accessible, ni labellisée ou non-biaisée), au besoin d'une implémentation sur GPU, et au manque d'interprétabilité lié à la difficulté à expliquer les décisions d'un réseau de neurone profond (côté « boîte noire »).

Tandis que la grande majorité des études de la littérature traite des problématiques de recalage 3D/3D IRM, on peut faire le constat que très peu d'études (7%) se concentrent sur des problématiques similaires à la nôtre : recalage 2D/3D pour de l'échographie. La partie suivante présentera notamment les méthodes basées sur l'apprentissage profond permettant d'améliorer le repérage ou le guidage des biopsies prostatiques.

## II.2 Évolution des avancées pour le repérage et le guidage des biopsies de prostate

Assister l'urologue à mieux se repérer et ainsi mieux se déplacer durant un examen de biopsies de prostate a occasionné de nombreuses recherches depuis deux décennies. Dans un premier temps, celles-ci se sont concentrées sur l'apport d'informations préopératoires pour améliorer le repérage et la cartographie des cibles via notamment des fusions multimodales 3D/3D lors de la navigation. Plus récemment, l'accent a été mis sur l'assistance continue au guidage, via des méthodes de recalage d'images 2D/3D, pour se rapprocher de l'objectif du temps-réel. Ces différentes méthodes font interagir différentes imageries, comme illustré dans la Figure II.5. Nous détaillons ci-dessous ces approches.

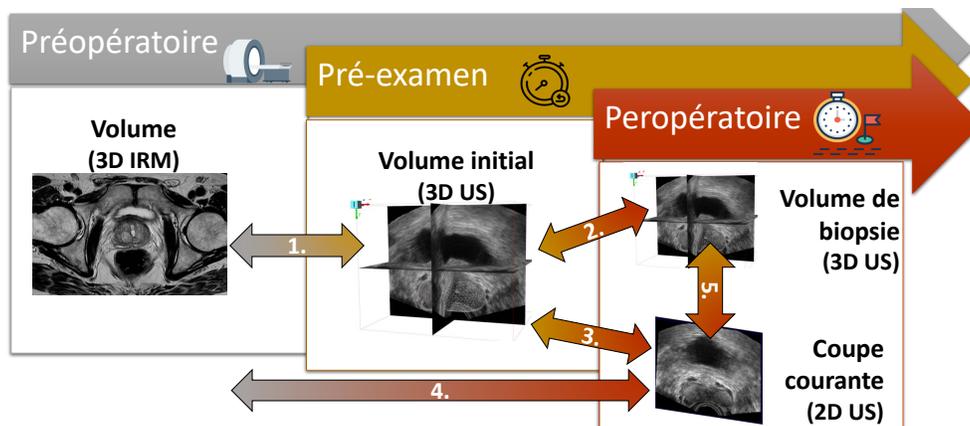


FIGURE II.5 – Représentation des différentes interactions de recalage selon les imageries et étapes durant la procédure de biopsie de prostate.

### II.2.1 Recalage 3D/3D avec fusion IRM pour le repérage et/ou la cartographie

Bien que l'IRM, de par sa complexité et son inaccessibilité en temps-réel, ne puisse être utilisée lors du guidage de procédure de biopsies de prostate, son intérêt reste néanmoins entier pour la recherche et la détection de lésions prostatiques en préopératoire. En effet, cette modalité présente une meilleure qualité d'image et une sensibilité particulière pour la détection précise de zones suspectes, souvent non visibles à l'échographie seule. L'utilisation conjointe de l'imagerie IRM préopératoire et de l'imagerie échographique durant la procédure, permet de transférer la cartographie des zones cancéreuses préopératoire en peropératoire [2, 31–33]. L'approche des biopsies échoguidées 3D avec fusion IRM se concentre sur l'apport d'informations préopératoires via un recalage 3D/3D entre le volume IRM préopératoire et un volume échographique réalisé en début d'intervention ou au cours de la procédure (voir interactions 1 et 2 de la Figure II.5). Les paragraphes suivants décrivent l'évolution de différentes méthodes concernées depuis ces dernières décennies : avec les méthodes classiques puis par apprentissage profond. Un descriptif plus détaillé de certaines de ces méthodes est également fourni en Table II.2.

### II.2.1.1 Méthodes classiques pionnières

Dans un premier temps, la plupart des travaux de fusion multimodale se basent sur des méthodes classiques [17, 34, 35], nécessitant souvent des étapes intermédiaires US-US (3D/3D [17, 35] ou de plusieurs coupes empilées 2.5D/3D [34]). Certains de ces travaux ont donné lieu à la mise au point de dispositifs commercialisés pour l'assistance à la cartographie de biopsies de prostate [2, 31, 32] : comme Uronav<sup>®</sup> [34], Artemis<sup>®</sup> [35], Urostation<sup>®</sup> et Trinity<sup>®</sup> de Koelis [17], etc. Ils permettent notamment de vérifier la localisation de carottes en cours (dites virtuelles) ou déjà prélevées, via l'affichage de la cartographie pendant ou après la biopsie. Ces dispositifs ont été classés en deux approches de recalage différentes selon le principe de localisation utilisée : « suivi d'organe » [17] ou « suivi de sonde » [34, 35].

Cependant, toutes ces méthodes de fusion multimodale classique basent leur optimisation sur une mesure de similarité liée à l'intensité des deux images, difficile à définir de manière pertinente si l'on considère la grande différence de contraste entre les deux types d'imageries ainsi que les différences de plans et de propriétés d'isotropie.

### II.2.1.2 Méthodes récentes

Afin de dépasser ces difficultés, de plus en plus d'études se sont basées sur des méthodes par apprentissage profond [36–46] avec parfois la prédiction de mesure de similarité [42] ou l'apport d'informations de segmentation [38–41, 43, 44, 47]. Ces segmentations se retrouvent le plus souvent sous forme d'images binaires, ou de plus en plus sous forme de nuages de points [38, 48]. De même que pour les méthodes classiques, la segmentation est utilisée :

- pour faciliter l'initialisation de la tâche de recalage multimodal avec un a priori de forme comme « caractéristique » [38, 40, 43] ou
- pour disposer d'un « label » d'auto-apprentissage sur une tâche différente de celle du recalage initial [39–41, 43, 47]. Cela permet de s'affranchir d'une mesure de vérité terrain comme la similarité des images, pas toujours révélatrice de la qualité du recalage.

Parmi toutes les méthodes recensées, nous choisissons d'en décrire certaines d'entre elles, présentant des concepts innovants et récents :

Zeng et al. [40] proposent un recalage non rigide 3D multimodal (IRM/TRUS) pouvant améliorer le guidage durant la curiethérapie ou les examens de biopsie. Leur méthode consiste en une structure hiérarchique composée de 3 réseaux (voir Figure II.6) : (i) premièrement, deux CNN parallèles permettent la segmentation de la prostate sur les volumes IRM et TRUS respectivement, (ii) puis, ces deux segmentations sont concaténées et utilisées en données d'entrée d'un réseau (2D-Unet) pour prédire un premier recalage affine. Enfin, (iii) un dernier réseau (3D-Unet) est appliqué pour prédire le champ vectoriel de déformation. Leur optimisation est basée sur deux termes de correction : une correction basée sur la distance entre des points surfaciques (DSC ou "Dice Similarity Coefficient"), et un terme de régularisation. Ces termes sont calculés de manière auto-supervisée en utilisant les segmentations des images. Cela permet à la fois de s'affranchir du choix d'une mesure de similarité multimodale sur les images et de la disponibilité des vérités terrain durant l'apprentissage. Après plusieurs comparaisons, ils précisent que l'étape de segmentation permet d'affiner la précision du

modèle et obtiennent une TRE moyenne finale d'environ de  $2.53 \pm 1.39$  mm évaluée sur des fiduciels (points repères de l'image).

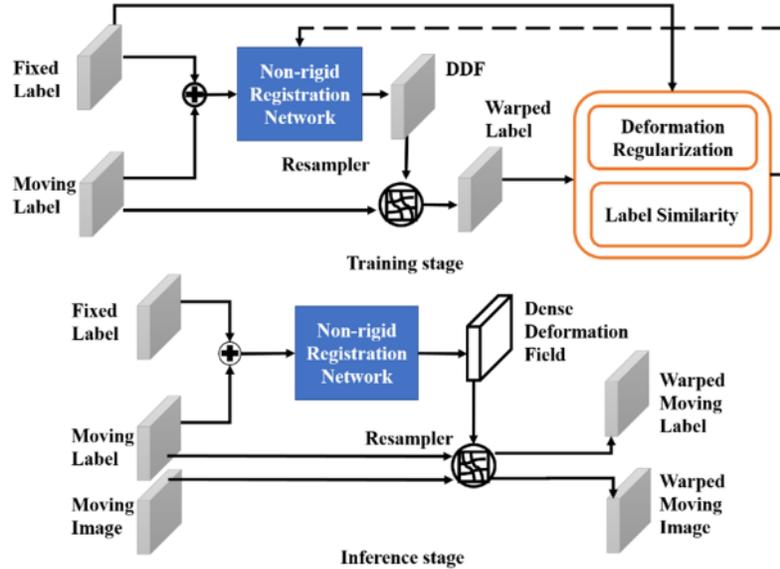


FIGURE II.6 – Fusion 3D multimodale en utilisant la segmentation comme caractéristique et label dans une stratégie d'apprentissage semi-supervisé. Empruntée à Zeng et al. [40].

Fu et al. [38] proposent un recalage non rigide sur des nuages de points pour améliorer la localisation de biopsies prostatiques à partir des images 3D(IRM)/3D(US). Pour cela, leur méthode consiste : (i) en deux CNN parallèles permettant la segmentation de la prostate et sa conversion en nuage de points, (ii) ainsi qu'un réseau de correspondance de points (PointNet) estimant le champ de déformation. Ils ajoutent également des termes de correction durant l'optimisation afin de mieux respecter les contraintes biomécaniques et physiques. En se comparant à d'autres méthodes de recalage sur les images segmentées, ils démontrent que les nuages de points sont une forme plus adaptée. Bien qu'ils obtiennent une TRE moyenne finale d'environ de  $1.57 \pm 0.77$  mm, le temps d'exécution est d'environ 5 min et ne correspond pas à une utilisation en temps réel. De plus, les vérités terrain sont établies à partir d'ICP (« Iterative Closest Point ») modifiées et des simulations FEM (« Finite Element Method ») qui ne sont pas spécifiques au patient et peuvent de ce fait être biaisées.

Elmahdy et al. [39] se concentrent sur la réalisation, simultanée et multi-tâches, de segmentation et de recalage, dans le contexte du guidage multimodal pour la radiothérapie de la prostate. Pour cela, ils comparent plusieurs architectures de réseau (encodage partagé, croisé) ainsi que plusieurs pondérations des tâches durant l'optimisation. Ils démontrent que la stratégie multi-tâches surpasse celle de la tâche unique, grâce au partage d'informations à la fois au niveau de l'architecture et du calcul d'optimisation. D'après leurs résultats, bien que l'apport de la segmentation soit plus important en tant qu'élément de l'optimisation, l'ajout de segmentation en données d'entrée peut également s'avérer pertinent selon les réseaux utilisés.

## II.2. ÉVOLUTION DES AVANCÉES

Malgré leur précision, ces fusions multimodales 3D nécessitent l'acquisition ou la reconstruction d'un volume (plusieurs secondes), tout en gardant une position statique de la sonde. Cela ne permet donc pas une assistance continue durant le guidage complet vers une prochaine cible. De plus, ces recalages peuvent rapidement devenir erronés à cause des déformations anatomiques en temps-réel durant l'examen : occasionnées par le patient (respiration, bougé) ou par le praticien (pression de la sonde, et insertion de l'aiguille). L'utilisation des images peropératoires 2D se rapproche de l'objectif d'assistance à la navigation en temps-réel.

TABLE II.2 – Comparaison des principales études de fusion 3D multimodales liées au repérage et guidage durant un examen de la prostate. Se référer à la table des acronymes.

Auteur Année	Recalage	Organe	Transfo	Architecture & Apprentissage	Renvois aux classes d'analyse	Résultats	Évaluation Données & GT
Hu 2018 [47]	3D US / 3D IRM	Prostate	Non-rigide	(i) 3D-CNN (ii) 3D-Unet $\Rightarrow$ DDF SSL : CE + L2	II.4.3.1-Hiérarchique II.4.4.1-Pénalisation	<b>TRE</b> : 4,2 mm <b>DSC</b> : 0.88	Fiduciels (76 patients) 111 paires d'images
Haskins 2018 [42]	3D US / 3D IRM	Prostate	Rigide	(i) CNN $\Rightarrow$ TRE (ii) Itératif Newton SIM : MSE(TRE)	$\emptyset$	<b>TRE</b> : 3,86 mm initialement à 16 mm	679 paires US-IRM (NIH database)
Zeng 2020 [40]	3D US / 3D IRM	Prostate (curie.)	Non-rigide	(i) 3D-CNN $\Rightarrow$ Seg (ii) 2D-Unet $\Rightarrow$ doF (iii) 3D-Unet $\Rightarrow$ DDF SSL : DSC+ MSD + BE	II.4.1.4-Segmentation II.4.3.1-Hiérarchique II.4.4.1-Pénalisation	<b>TRE</b> : 2,53 $\pm$ 1,39 mm <b>DSC</b> : 0,91 $\pm$ 0,02 <b>HD</b> : 4,41 mm	Fiduciels (36 patients) et segm. manuelles
Guo 2020 [37]	3D US / 3D IRM	Prostate	Rigide	3D ResNeXt $\Rightarrow$ doF SL : MSE(DoF)	II.4.3.1-Hiérarchique	<b>MSE</b> : 3.66mm initialement à 9.42 mm	679 paires US-IRM (NIH database)
Fu 2020 [38]	3D US / 3D IRM	Prostate	Non-rigide	(i) 3D-CNN $\Rightarrow$ Seg (ii) ICP (iii) PC-Unet $\Rightarrow$ DDF SL :CD+ MSE(DDF)	II.4.1.4-Segmentation II.4.2.1-Paramétrage II.4.4.1-Pénalisation II.4.5.2-Génération GT	<b>TRE</b> : 1,57 $\pm$ 0,77 mm <b>DSC</b> : 0.94 $\pm$ 0.02 <b>HD</b> : 2.96 $\pm$ 1.00 <b>TIME</b> : 3 à 5 min	Fiduciels (50 patients) GT (DDF) calculée à partir de simulation FEM non patient spécifique
Elmahdy 2021 [39]	3D CT / 3D CT	Prostate (radio.)	Non-rigide	Shared-Encoder Unet cross-stitch $\Rightarrow$ DDF MTL :DSC,NCC, BE	II.4.1.4-Segmentation II.4.4.2-Multi-tâches	<b>MSD</b> : 1.06 $\pm$ 0.3 mm	Segm. manuelles (20 patients)
Baum 2021 [48]	2Ds US / 3D IRM	Prostate	Non-rigide	(i)Global Point-Net (ii) MLP $\Rightarrow$ DDF SSL : CD+GMM	II.4.1.4-Segmentation II.4.2.2-Coordonnées II.4.4.1-Pénalisation	<b>CD</b> : 1.40 $\pm$ 0.20 <b>HD</b> : 6.38 $\pm$ 1.48 <b>TRE</b> : 4.88 $\pm$ 1.56	Fiduciels (108 paires, 76 patients)

## II.2.2 Recalage 2D/3D pour le guidage en temps-réel

Plus récemment, l'accent a été mis sur l'assistance dynamique à la navigation via des méthodes de recalage d'images 2D/3D, fonctionnant en temps-réel. Celles-ci consistent à réaliser un recalage entre l'image échographique 2D peropératoire et un volume de référence : IRM préopératoire (voir interaction 4, Figure II.5) ou US de début d'intervention (voir interactions 3 et 5 de la Figure II.5). Ces méthodes présentent des objectifs similaires aux nôtres et concernent la même application clinique. Nous avons pris soin de décrire en Table II.3 leurs performances ainsi que l'approche choisie pour générer des données d'entraînement et valider les performances. Les paragraphes suivants décrivent l'évolution de différentes méthodes concernées depuis ces dernières décennies : avec les méthodes classiques puis par apprentissage profond.

### II.2.2.1 Méthodes classiques pionnières

Dans un premier temps, ces méthodes sont majoritairement classiques et iconiques [20, 34, 45, 49, 50]. Seules quelques méthodes se disent hybrides [51, 52], en considérant également des métriques géométriques. Ces méthodes classiques nécessitent généralement une bonne initialisation de la transformation recherchée pour éviter les problèmes de convergence vers des minima locaux. La plupart initialisent leur recalage manuellement [34, 45], par l'intermédiaire de la localisation de la sonde (dans le référentiel physique  $R_{phys}$ ) [20, 34, 45, 49, 50], ou à partir de recalages 3D/3D précédents [17, 51] parfois combinés à un suivi de sonde [51]. L'idée d'apporter un recalage continu, sans intervention externe ni suivi de sonde, est de plus en plus envisagée en utilisant les résultats du recalage précédent de manière séquentielle [52]. Mais toutes ces méthodes présentent des performances qui ne sont pas compatibles avec les attentes cliniques (temps de calcul notamment) et dont les évaluations sont souvent limitées à très peu de données. Nous choisissons d'en décrire certaines d'entre elles ci-dessous :

Zhang et al. [45] proposent une assistance à la navigation applicable pour le guidage lors du traitement de curiethérapie ou de biopsie de prostate. S'inspirant d'une méthode plus ancienne [34], ils réalisent leur recalage 2D/3D multimodal en deux étapes : (i) un premier recalage 3D IRM/3D US est manuellement initialisé par un expert puis optimisé grâce à la méthode des gradients sur la métrique d'information mutuelle, puis (ii) un deuxième recalage 2D US/3D IRM est cette fois-ci initialisé grâce à la localisation de la sonde. Ils obtiennent une TRE moyenne finale d'environ  $2.52 \pm 0.46$  mm à partir de fiduciels définis pour 5 patients. En revanche, le temps de calcul est de plus de 35 secondes, et la méthode ne peut fonctionner qu'avec des initialisations provenant d'intervention externe ou de localisation de la sonde.

Gillies et al. [20] décrivent une méthode de recalage 2D/3D rigide qui optimise la NCC à travers l'algorithme de Powell. Ils visent à compenser les déformations en temps-réel à travers des recalages toutes les secondes, initialisés à partir de la position donnée par un bras robotisée portant la sonde localisée. Après plusieurs améliorations de leur méthode initiale [50] (un sous-échantillonnage, l'optimisation des calculs sur GPU, la détection d'une zone d'intérêt, une contrainte de direction de recherche et une priorisation des déplacements de rotation sur la translation), ils obtiennent une TRE moyenne d'environ 2 mm évaluée sur 7 patients [49].

Beitone et al. [52] présentent une méthode de recalage pour l'assistance continue à la localisation de biopsies de prostate, en recalant l'image peropératoire 2D à un volume de référence (acquis

au site précédent de biopsie). Ils comparent deux initialisations différentes pour ce recalage : (i) une initialisation « statique » utilisant le recalage 3D IRM/3D US non-rigide disponible au site de biopsie précédent [17] et combiné au suivi de sonde depuis (étude préliminaire [51]), ou (ii) une initialisation « séquentielle » considérant les résultats du recalage 2D/3D précédent comme initialisation du prochain. Leur méthode de recalage se compose de plusieurs étapes (voir Figure II.7), appliquées de manière itérative pour améliorer les résultats : (i) détection de points caractéristiques (géométriques et iconiques) 2D et 3D, (ii) correspondance et rejet des paires à partir de mesures iconiques sur les images, et (iii) calcul de la transformation à partir d'un l'algorithme ICP hybride combinant métriques géométriques et iconiques. Pour évaluer la performance de leur approche, ils simulent des transformations de déplacement de sonde ( $\pm 5^\circ$  and  $\pm 7.5\text{cm}$ ) au sein de 900 volumes issus de 80 patients. Par rapport à leur étude préliminaire [51], ils obtiennent une amélioration significative des résultats : avec une TRE finale d'environ  $3.80 \pm 0.4$  mm ainsi qu'une optimisation du temps de calcul, estimé entre 60-200 ms. Leur initialisation séquentielle, qui utilise les résultats du recalage précédent comme initialisation du recalage suivant, est prometteuse et permet de s'affranchir notamment d'un suivi ou de la localisation de sonde.

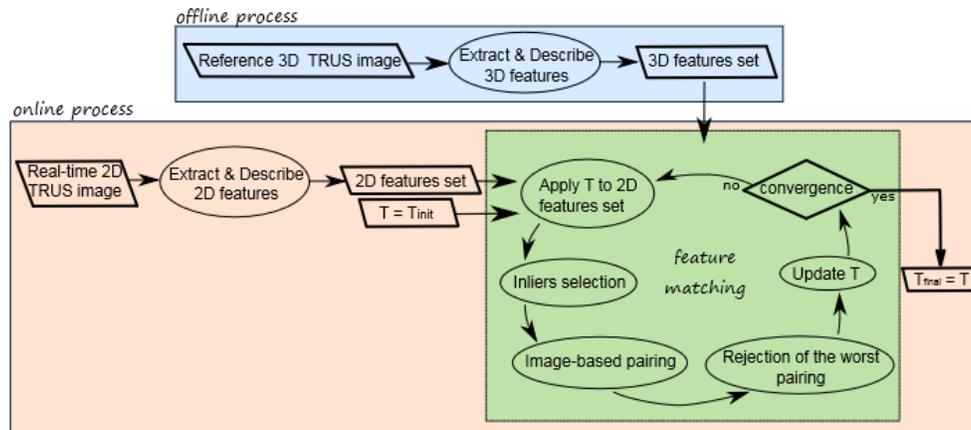


FIGURE II.7 – Illustration de la méthodologie de Beitone et al. [52] basée sur l'étude préliminaire [51, 53].

### II.2.2.2 Méthodes récentes

On trouve désormais, de plus en plus des méthodes d'apprentissage profond qui se sont penchées sur ce problème [54–57] et qui ont notamment démontré leur supériorité en se comparant à certaines des méthodes classiques citées précédemment. Cependant, bien que ces méthodes présentent des fréquences et des précisions de recalage satisfaisantes, les conditions expérimentales sont encore limitées, notamment en termes de réalisme clinique, que ce soit au niveau de leur génération de données ou au niveau de leur évaluation.

Bhardwaj et al. [54] proposent de combiner des méthodes de recalage classiques et par apprentissage supervisées afin de réaliser un recalage non rigide 2D(US)/3D(IRM) pour les biopsies de prostate. Après une première initialisation manuelle 3D(IRM)/3D(US) qui servira de « référence », leur méthode repose sur un processus hiérarchique (illustré en Figure II.8) : tout d'abord, une étape

de segmentation sur les deux imageries permet d'obtenir des points caractéristiques 2D et 3D sur lesquels sera réalisée, par la suite, une étape de recalage rigide classique initialisé par la transformation de « référence ». Enfin, une correction non-rigide est appliquée grâce à un réseau supervisé (2D-Unet) qui estime la carte de déplacement entre la coupe courante US et la coupe extraite du volume IRM à partir du recalage rigide précédemment établi. La fonction de coût est une métrique de distance entre les cartes de champs estimées et les vérités terrain, établies par la librairie ITK. Ils évaluent leur méthode sur des coupes enregistrées et localisées (position de la sonde seule) durant de vraies procédures cliniques, sur une base de données de 12 patients. Ils obtiennent des corrections de recalage jusqu'à  $2.99 \pm 1.51$  mm (TRE) par rapport à l'initialisation de référence qui se détériore au fil de la procédure ( $8.9 \pm 5.1$  mm). Sur cette même base de données, ils surpassent certaines des méthodes classiques citées précédents. Néanmoins, compte tenu des étapes de recalages classiques coûteuses en temps de calcul, ils obtiennent leurs résultats en plus de 100 ms. La validité de leur vérité terrain, obtenue par une méthode de la littérature, est de plus fortement limitée et discutable.

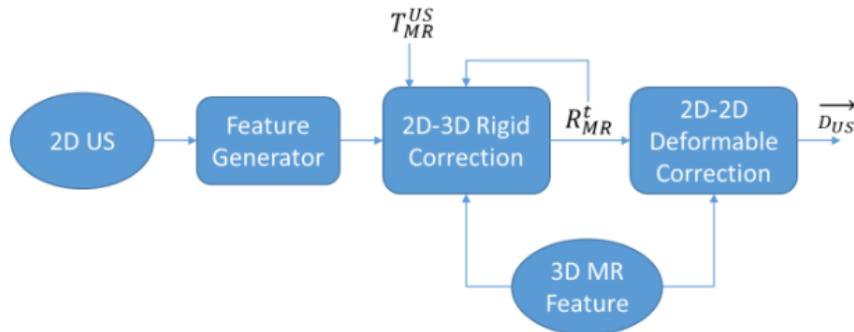


FIGURE II.8 – Illustration de la méthodologie de Bhardwaj et al. [54]

Bhardwaj et al. [55] proposent par la suite une amélioration de leur travaux précédents pour notamment se rapprocher de l'objectif temps réel. Pour cela, ils remplacent leur initialisation de recalage rigide classique 3D/3D par une correction de translation (plan) entre la coupe US courante segmentée et la coupe segmentée extraite du volume IRM à partir de la position de la sonde. De plus, ils modifient la fonction de coût de leur réseau de correction non-rigide en quantifiant directement la différence obtenue entre les segmentations transformées par les cartes de déplacements estimées, on parle d'auto-pénalisation. Leur méthodologie est également illustrée sur la Figure II.9. Avec une méthode d'évaluation similaire et sur quelques patients supplémentaires, ils obtiennent une TRE finale d'environ 2.51 mm avec un temps de calcul d'environ 70 ms. Leur méthode se limite cependant à l'estimation de petites transformations, notamment dans le plan, et nécessite obligatoirement l'utilisation de la localisation de la sonde qui n'est pas toujours disponible lors de réelles procédures cliniques.

Guo et al. [56] proposent de réaliser un recalage rigide 2D(US)/3D(US) pour assister la navigation de biopsie prostatique en temps-réel. Leur première contribution est d'intégrer dans deux branches parallèles la coupe courante US et une sous-partie du volume US, tout en préservant un équilibre entre les deux types de caractéristiques ainsi déduites. Ces sous-volumes d'entrée sont initialisés, centrés et orientés, à partir de la vérité terrain bruitée, pour limiter l'espace de recherche. La deuxième contribution est d'ajouter un terme d'auto-pénalisation à partir de mesures de simi-

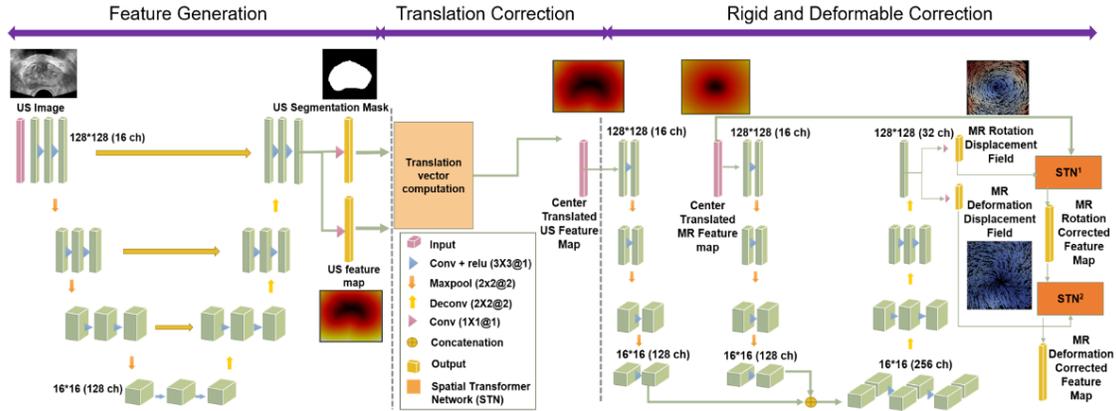


FIGURE II.9 – Illustration de la méthodologie de Bhardwaj et al. [55]

larité entre la coupe extraite du volume à la position prédite et la coupe 2D courante, en utilisant notamment un échantillonneur différentiable. La méthode est illustrée sur la Figure II.10. Ils démontrent notamment que cette architecture équilibrée permet de résoudre la difficulté de combiner des caractéristiques de natures différentes (2D ou 3D). Sur une large base de données composée des coupes localisées et utilisées lors de la reconstruction d'un volume, ils se comparent à plusieurs méthodes classiques précédemment citées et obtiennent des erreurs d'environ 2.7 mm. Cependant, tout comme leur phase d'apprentissage, l'évaluation repose également sur une initialisation des sous-volumes grâce à la vérité terrain bruitée, très avantageuse. Cela ne serait pas faisable sur des données cliniques.

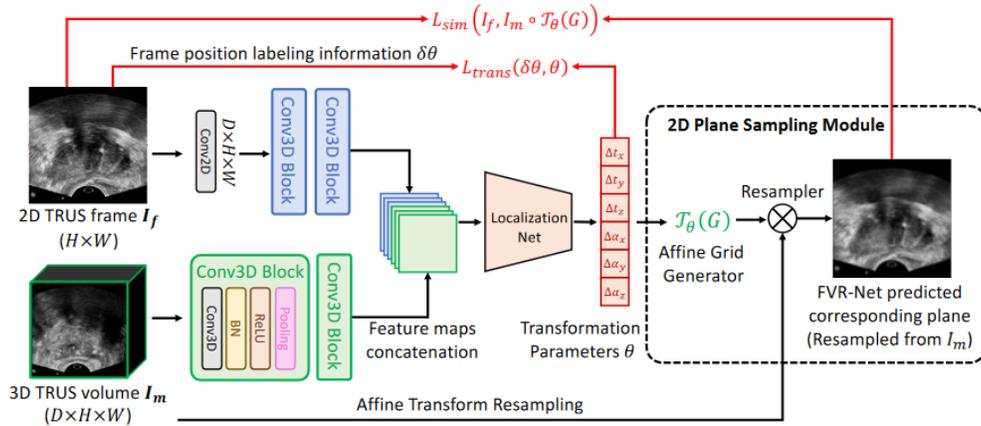


FIGURE II.10 – Illustration de la méthodologie de Guo et al. [56]

Guo et al. [57] proposent par la suite, une nouvelle version d'assistance à la navigation de biopsie, avec des recalages US 2D/3D évalués de manière plus réaliste sur des trajectoires et sans aucune intervention externe. Leur méthode se décompose en quatre étapes hiérarchiques (voir Fig.II.11) (i) Une première estimation de la position spatiale 3D de l'image courante est faite à partir d'une

méthode classique iconique. Puis, (ii) un réseau (2D/2D) estime les déplacements relatifs entre deux images successives. Ces transformations sont successivement combinées à la transformation globale initiale, présentant alors un risque d'accumulation d'erreur. (iii) Une étape de correction est alors appliquée, entre l'image extraite (du contexte 3D US) à la position prédite et l'image courante. (iv) Cette étape est répétée plusieurs fois, pour ajuster l'estimation jusqu'à atteindre le maximum de similarité pour la nouvelle coupe extraite. Pour cela, un mécanisme de contrôle de la qualité du recalage est utilisé et sélectionne la coupe présentant la plus grande similarité avec l'image courante. Les itérations de cette étape permettent de réduire considérablement l'erreur accumulée, mais introduisent un temps de traitement supplémentaire. La méthode a été validée sur des procédures cliniques de fusion 2D-US à 3D-MR sur 618 sujets. Ils obtiennent une erreur moyenne d'environ  $6.95 \pm 8.03$  mm (sur les coins de l'image) ou de 1,93 mm (sur 2 fiduciels établis par des experts), avec une fréquence d'environ 5 Hz. Cependant, la vérité terrain de leur localisation reste limitée à la localisation de la sonde sans aucune information sur la déformation de l'organe, et leur métrique n'est évaluée qu'à 12 moments pendant une procédure de trajectoire complète.

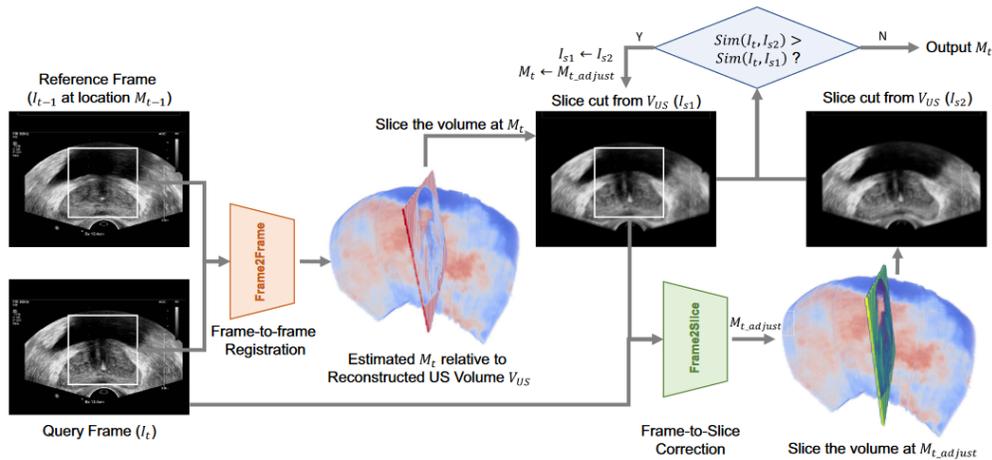


FIGURE II.11 – Illustration de la méthodologie de Guo et al. [57]

Parmi les études référencées autour des problématiques de recalage 2D(US)/3D(US) pour le guidage de procédure autour de la prostate (biopsie, curiethérapie, etc.), nous en avons retenu quatre principales [54–57] dont les objectifs, méthodologies et attentes clinique sont les plus similaires aux nôtres, à cette heure. Nous nous comparerons à elles dans la suite de ce manuscrit. Au vu du faible nombre de contributions de la littérature sur cette problématique précise, nous avons également étendu nos recherches aux méthodes de recalage 2D/3D toutes applications et modalités confondues. À travers le terme « recalage 2D/3D » nous ciblons non seulement les problèmes de localisation de coupes dans un volume mais également les problèmes de reconstruction de volume à partir de coupes. Bien que les performances obtenues soient difficilement comparables à celles attendues lors de recalage 2D/3D échographique pour des biopsies, nous décrirons dans la partie suivante certains travaux présentant des approches méthodologiques intéressantes.

## II.2. ÉVOLUTION DES AVANCÉES

TABLE II.3 – Comparaison de certaines études de recalage 2D liées au guidage en temps réel durant un examen autour de la prostate. Se référer à la table des acronymes. \*Ici, le symbole  $\delta$  correspond à l'écart relatif.

Auteur Année	Recalage	Organe	Transfo	Architecture & Apprentissage	Renvois aux classes d'analyse	Résultats	Évaluation Données & GT
Xu 2008 [34]	2.5D US / 3D IRM	Prostate	Rigide	(i) init. Manuelle (ii) init. Suivi sonde (iii) Tradi. iconique SGD(SSD)	II.4.1.1-Positionnelle II.4.3.1-Hiérarchique II.4.5.1-Génération II.4.5.4-Déroulement	<b>TRE</b> : $2,1 \pm 1,12$ <b>TIME</b> : 100 s	Fiduciels (fantômes)
Zhang 2015 [45]	2D US / 3D IRM	Prostate	Rigide	(i) init. Manuelle (ii) init. Suivi sonde (iii) Tradi. iconique SGD(MI)	II.4.1.1-Positionnelle II.4.3.1-Hiérarchique II.4.5.1-Génération II.4.5.4-Déroulement	<b>TRE</b> : $2,52 \pm 0,46$ <b>TIME</b> : 35 s	Fiduciels (5 patients)
Gillies 2017 [20]	2D US / 3D US	Prostate	Rigide	Tradi. iconique Powell(NCC)	II.4.1.1-Positionnelle II.4.5.1-Génération	<b>TRE</b> : $1,4 \pm 1,17$ mm plan $0,2 \pm 0,3$ mm hors plan $0,7 \pm 0,4$ mm roll $0,8 \pm 1,0^\circ$ <b>TIME</b> : $57 \pm 20$ ms	Fiduciels (7 patients) validation sur fantômes
Selmi 2018 [51]	2D US / 3D US	Prostate	Rigide	(i) init. 3D/3D*suivi (ii) Detection feature (iii) Match (iv) Tradi. hybride ICP( dist(f, I) )	II.4.1.1-Positionnelle II.4.5.1-Génération II.4.5.4-Déroulement	<b>TRE</b> : $3,91 \pm 3,22$ mm initialement à $8,1 \pm 4,3$ mm <b>TIME</b> : 4-10 s	Fiduciels (20 patients) sur 160 volumes  2D extraites de simulation sur volume de $\pm 5^\circ$ et $\pm 7,5$ mm
Beitone 2021 [52]	2D US / 3D US	Prostate	Rigide	(i) init. Prédiction (ii) Detection feature (iii) Match - Rejet (iv) Tradi. hybride ICP( dist(f, I) )	II.4.1.1-Positionnelle II.4.5.1-Métriques II.4.5.1-Génération II.4.5.4-Déroulement	<b>TRE</b> : $3,80 \pm 0,40$ mm initialement à 8.56 mm <b>TIME</b> : 60-200 ms	Maillage rigide sur 900 volumes issus de 80 patients  2D extraites de simulation sur volume de $\pm 5^\circ$ et $\pm 7,5$ mm
Bhardwaj 2020 [54]	2D US / 3D IRM	Prostate	Non-rigide	(i) init. manuelle 3D IRM/3D US (ii) 2D-Unet $\Rightarrow$ Seg (iii) Tradi. iconique SGD( MSE(seg) ) (iv) 2x 2D-Unet $\Rightarrow$ DDF  SL :DSC/ MSE(DDF)	II.4.1.1-Positionnelle II.4.1.4-Segmentation II.4.3.1-Hiérarchique II.4.5.2-Génération GT II.4.5.4-Déroulement	<b>TRE</b> : $2,99 \pm 1,51$ mm initialement à $8,9 \pm 5,1$ <b>TIME</b> : 112 ms (46+32+34 ms)	Fiduciels (12 patients) GT(DDF) calculée par le logiciel ITK  2D US (2000 coupes) enregistrées (vidéo, 10 Hz) et localisées
Bhardwaj 2021 [55]	2D US / 3D IRM	Prostate	Non-rigide	(i) 3D-Unet $\Rightarrow$ Seg (ii) init. localisation EM (iii) 2D correction $\Rightarrow$ transl. (iv) 2x 2D-Unet $\Rightarrow$ DDF SSL : MSE(STN(Seg))	II.4.1.1-Positionnelle II.4.1.4-Segmentation II.4.3.1-Hiérarchique II.4.4.1-Pénalisation II.4.5.2-Génération GT II.4.5.4-Déroulement	<b>TRE</b> : $2,51 \pm 1,2$ mm <b>TIME</b> : 70 ms	Fiduciels (34 patients) GT(DDF) calculée par le logiciel ITK  2D US (6000 coupes) enregistrées (vidéo, 10 Hz) et localisées
Guo 2021 [56]	2D US / 3D US	Prostate	Rigide	3D-CNN // 2D+3D-CNN 3D ResNeXt $\Rightarrow$ doF SSL :MSE(doF)+SSIM	II.4.1.1-Spatiale II.4.4.1-Pénalisation II.4.5.2-Génération II.4.5.4-Déroulement	<b>Distance</b> : 2,73 mm initialement à 7.68 mm <b>NCC</b> : 0.92	2D extraites du volume reconstruit et localisées par EM sur 600 patients
Guo 2022 [57]	2D US / 3D US	Prostate	Rigide	i) US-MR Manuelle ii) init Tradi. iconique GD( MSE) = 5 sec iii) F2F $\Rightarrow \delta(\text{doF})$ iv) n*F2S $\Rightarrow \delta(\text{doF})$ Stop si SSIM Is2 < Is1 SSL :MSE+PCC+SSIM	II.4.3.1-Hiérarchique II.4.4.1-Pénalisation II.4.5.4-Déroulement	<b>Distance</b> : $6,95 \pm 8$ mm <b>NCC</b> : $0,80 \pm 0,18$  Validation clinique <b>TRE</b> : 1,93 mm <b>TIME</b> : 150 ms	2D US enregistrées (vidéo) localisées par EM sur 600 patients  Validation clinique 2 Fiduciels + points de segmentation

## II.3 Travaux de recalage « 2D/3D » par des approches IA

Nous avons choisi de répertorier les études portant sur du recalage 2D/3D en trois catégories différentes : travaux ne donnant que la coupe à recaler en entrée ; travaux donnant la coupe et le volume en entrée ; travaux donnant une séquence de coupes en entrée.

### II.3.1 Une seule coupe en entrée

Contrairement aux méthodes de recalage classique, les méthodes de recalage 2D/3D par apprentissage profond ne considèrent le plus souvent que la coupe à recaler, et ne prennent en compte aucune information concernant le volume de destination, même lorsque celui-ci est disponible. La plupart des travaux [58–65] se concentrent sur des tâches supervisées de régression, prédisant directement les paramètres de transformation (souvent rigides), en se basant sur des réseaux de type CNN. Cependant, la tâche de localisation 2D/3D étant complexe, la plupart de ces travaux limitent la difficulté à travers de nombreuses étapes de prétraitement et d’initialisation [61, 63, 65], à travers de simulation des données limitant les transformations à une seule direction [58, 60, 62, 63] ou à travers plusieurs niveaux de recalages itératifs et hiérarchiques [60, 62, 65]. Un descriptif plus détaillé de certaines de ces méthodes est également fourni en Table II.4.

TABLE II.4 – Comparaison de certaines études de recalage « 2D/3D », ne donnant qu’une seule coupe en entrée du réseau. Se référer à la table des acronymes. \*Ici, le symbole  $\delta$  correspond à l’écart relatif.

Auteur Année	Recalage	Organe	Transfo	Architecture & Apprentissage	Renvois aux classes d’analyse	Résultats	Évaluation Données & GT
Miao 2016 [65]	2.5 Xray / 3D CT		Rigide	(i)Prétraitement $\Rightarrow$ ROI (ii) CNN $\Rightarrow$ doF  SL : MSE(doF)	II.4.1.4-Segmentation II.4.3.1-Hiérarchique II.4.5.1-Génération	<b>RMSE</b> : 0.04/0.04/0.32 mm 0.06/0.18/0.18 ° <b>TIME</b> : 100 ms	2D extraites de simulation sur volume 1/1/10 mm ; 2/10/10°  940 images pour test
Li 2018 [62]	2D US / 3D US	Foetus cerveau	Rigide	CNN $\Rightarrow$ T (euler/points /quaternions/ matrice) + Incertitude (I)  MTL : L2(T)+ CE(I)	II.4.2.1-Paramétrage II.4.2.4-Incertitude II.4.3.1-Hiérarchique II.4.4.2-Multi-tâches	<b><math>\delta t</math></b> : 3,83 $\pm$ 1,49 mm <b><math>\delta \theta</math></b> : 12,7° $\pm$ 7 <b>SSIM</b> : 0,95 <b>TIME</b> : 40 ms	2D extraites de volumes (72 patients) autour de plan standard TC/ TV
Chen 2019 [63]	2D CT / 3D IRM	Cerveau	Rigide	AlexNet $\Rightarrow$ doF hors-plan (échantillonnage)  SL : CE(softmax(doF)))	II.4.2.1-Paramétrage II.4.2.2-Classification II.4.5.1-Génération	<b>TRE</b> : 4,5 $\pm$ 8 mm <b>MSE</b> : 121,7m <b>SSIM</b> : 0,95 <b>TIME</b> : 40 ms	2D extraites de simulation sur volume $\pm$ 60°, $\pm$ 15mm
Di Vece 2022 [64]	2D US / 3D US	Foetus cerveau	Rigide	3D ResNet $\Rightarrow$ doF  SL : MSE(doF)	II.4.2.1-Paramétrage II.4.4.2-Multi-tâches	<b>L2</b> : 11,84 mm <b>GE</b> : 5,17° <b>TIME</b> : 3 ms	Entraînement : 2D extraites de volume (fantômes), plan standard TV Validation sur 22000 coupes (7 patients)

Miao et al. [65] fût l’une des études pionnières de l’estimation des paramètres de transformation de manière directe. Ils proposent de réaliser un recalage 2D(projection)/3D(CT) rigide de manière très structurée, avec à la fois une hiérarchisation de la difficulté, des sous-études locales, et un prétraitement d’images afin de fusionner les types d’informations dans une seule image d’entrée.

L'image d'entrée prétraitée est partitionnée en plusieurs zones, sur lesquelles un réseau CNN est appliqué (avec des poids partagés). La prédiction est en réalité organisée de manière hiérarchique et subdivisée en trois étapes : estimations des paramètres dans le plan, rotation hors plan, et translation hors plan. À la fin de chaque étape, une nouvelle image d'entrée est rééchantillonnée à partir des estimations précédentes. Leur méthode est illustrée sur la Figure II.12. Leur étude d'ablation a notamment révélé qu'une telle hiérarchisation de la difficulté, sur des zones locales et prétraitées, permet de réduire la complexité du problème et d'améliorer la qualité de la prédiction.

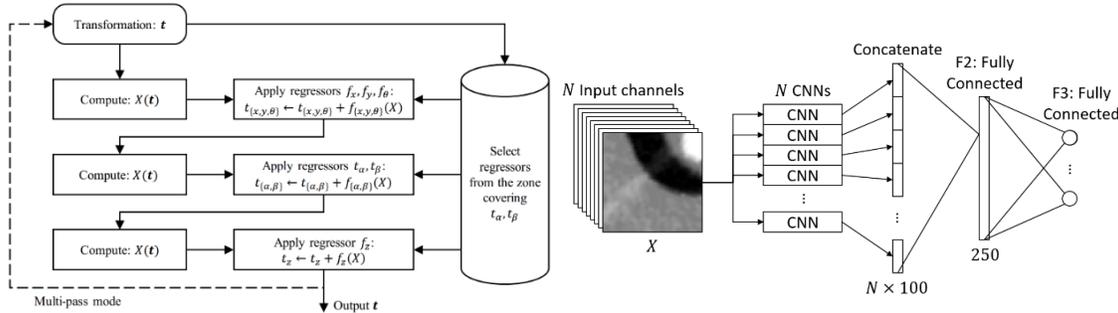


FIGURE II.12 – Illustration de la méthodologie de Miao et al. [65]

Li et al. [62] utilisent une méthode itérative basée sur un réseau CNN supervisé permettant de réaliser un recalage 2D(US)/3D(US) rigide. À chaque itération pendant l'inférence, le réseau prédit les paramètres de la transformation rigide qui vont permettre d'extraire du volume une nouvelle coupe, plus proche de la position recherchée. Le réseau estime également, dans une seconde tâche, la qualité des paramètres prédits à chaque itération, et utilise cette estimation de qualité pour pondérer les paramètres pour l'étape suivante, selon un protocole spécifique développé. Ils démontrent notamment que cette probabilité permet d'améliorer significativement les résultats car elle permet à la fois de mieux généraliser la prochaine initialisation et de partager une optimisation conjointe plus pertinente. Ils comparent différents types de représentation pour les angles (quaternions, angles d'Euler, matrices de rotation, points fiduciels) et finissent par privilégier les quaternions. Leurs données d'évaluation sont déjà toutes orientées autour d'un plan standard, facilitant alors la tâche du recalage et rendant la comparaison des résultats difficiles.

Chen et al. [63] est l'une des seules études à traiter l'estimation de paramètres de transformation sous la forme d'un problème de classification. Ils utilisent un réseau CNN supervisé pour réaliser un recalage 2D/3D rigide et multimodal entre les imageries IRM/CT du cerveau. Les auteurs décrivent chaque coupe par son centre et son vecteur normal permettant de représenter les transformations sous forme de classes (répartition polaire uniforme échantillonnée). Le réseau est entraîné sur des coupes fortement prétraitées, avec notamment segmentation et alignement, permettant notamment de réduire la dimension de recherche aux paramètres hors plan et faciliter le recalage. Les résultats obtenus sont comparables à ceux des approches par régression.

Di Vece et al. [64] étudient la localisation d'un plan US en temps-réel dans le cadre de guidage lors d'examen obstétricaux. Pour cela, ils utilisent un réseau convolutif pour estimer les paramètres de localisation du plan par rapport à une origine d'un modèle générique, sans ajout ni du volume ni d'autre information de localisation. A la place des codages de rotation traditionnels (quaternions,

ou angles Euler, etc.), ils privilégient un codage continu de la rotation à partir de deux vecteurs (base de la matrice de rotation), qui seront par la suite projetés, normalisés et complétés grâce à la méthode de Gram-Schmidt. Leur méthode est illustrée sur la Figure II.13. Leurs données d'évaluation sont simulées de manière simplifiée autour d'un même plan (Trans ventriculaire) sur 7 volumes de patients. Ils obtiennent des erreurs moyennes d'environ 11.84 mm et 25.17° en moins de 3 ms par coupe.

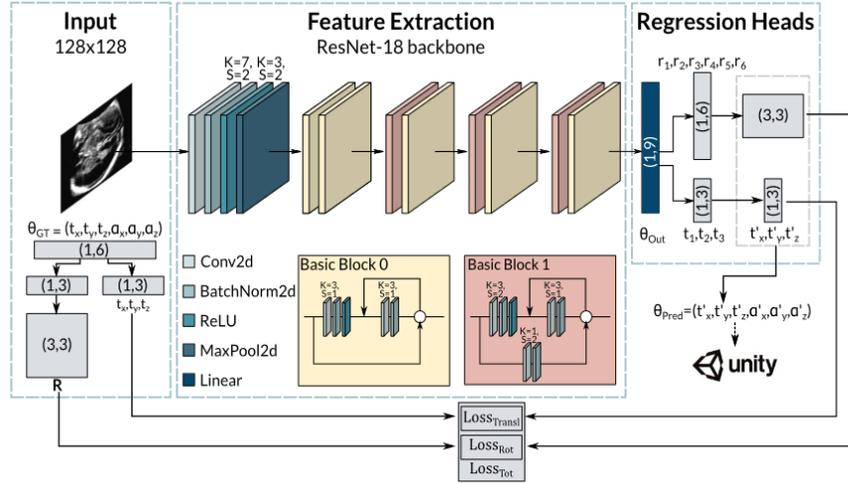


FIGURE II.13 – Illustration de la méthodologie de Di Vece et al. [64]

### II.3.2 Coupe et volume en entrée

Contrairement à la catégorie précédente, ces études [37,48,54,55,66–69] portent sur des recalages 2D/3D qui utilisent les deux types d’images (coupe et volume) en entrée de leur réseau. À cause de la difficulté de combiner ces deux images de représentation et dimension différentes, très peu d’études font une correspondance dans les espaces de dimensions de départ 2D/3D [69]. La plupart de ces études se ramènent à un problème à une seule dimension : (i) en projetant les caractéristiques 2D dans l’espace 3D grâce à la localisation de la sonde [48], ou (ii) en extrayant une coupe d’un volume [54,55], ou (iii) en manipulant les dimensions d’une couche du réseau pour étendre/dupliquer des cartes caractéristiques [37,66,67]. Un descriptif plus détaillé de certaines de ces méthodes est également fourni en Table II.5.

TABLE II.5 – Comparaison de certaines études de recalage « 2D/3D », donnant coupe et volume en entrée du réseau. Se référer à la table des acronymes. \*Ici, le symbole  $\delta$  correspond à l’écart relatif.

Auteur Année	Recalage	Organe	Transfo	Architecture & Apprentissage	Renvois aux classes d’analyse	Résultats	Évaluation Données & GT
Wei 2020 [66]	2D US / 3D IRM	Foie	Rigide	(i) Resnet $\Rightarrow$ classif (ii) 2D-Unet $\Rightarrow$ 2D-Seg + Replicateur (iii) 3D-Unet $\Rightarrow$ US-Seg (iv) Regression $\Rightarrow$ p, n  SSL : CE (classif) / DSC(US-Seg)	II.4.1.2-Spatiale II.4.1.4-Segmentation II.4.2.1-Paramétrage II.4.2.2-Classification	$\delta\theta$ : $12,7 \pm 6^\circ$ <b>Distance</b> : $4,9 \pm 3,1$ mm	2D extraites de simulation sur volume $\pm 20^\circ$ et 20mm 1035 images (52 patients) GT définie manuellement
Markova 2022 [69]	2D US / 3D IRM	Abdos	Rigide	(i) 2 Unet $\Rightarrow$ f2d, f3d (ii) Similarity $\Rightarrow$ Sij (iii) Ransac  SSL : loss (softmax(Sij))	II.4.1.2-Spatiale II.4.3.2-Dynamique II.4.4.1-Pénalisation	$\delta\theta$ : $21,3 \pm 16,3^\circ$ initialement à 44 <b>Distance</b> : $21,5 \pm 16,9$ mm initialement à 40	2D extraites de 30 balayages : 3957 images (16 patients)  GT définie manuellement
Fu 2022 [67]	2D IRM / 3D IRM	Abdos	Rigide	3D-CNN // 2D+3D-CNN $\Rightarrow$ doF  SSL : MSE(doF)+SSIM	II.4.1.2-Spatiale II.4.3.1-Hiérarchique II.4.4.1-Pénalisation	<b>MSE</b> : 27 initialement à 1927	2D extraites de simulation sur volume $\pm 4^\circ$ , $\pm 10$ mm sur 10 patients

Wei et al. [66] proposent un réseau supervisé permettant une initialisation rigide pour un recalage 2D(US)/3D(IRM) durant un examen échographique du foie. Le modèle repose sur un processus hiérarchique avec une première étape de classification déterminant le type de coupes standards (médiales ou trans-costales). Après une étape de segmentation automatique (2D-Unet) ne gardant que les structures de vaisseaux anatomiques importantes, le volume initial est transformé avec les paramètres moyens de la classe prédite, alors que la coupe initiale est, elle, dupliquée dans les mêmes dimensions que ce volume. Les deux “volumes” ainsi obtenus sont fournis à un réseau (3D-Unet) qui prédira la position du plan US à travers la segmentation de ce plan dans le volume. Une régression estime ensuite le centre et le vecteur normal à ce plan segmenté. Cette méthode, illustrée en Figure II.14 est notamment comparée à des méthodes précédemment citées [58,60] et les résultats démontrent l’intérêt de l’ajout du volume en entrée ainsi que d’une étape de classification pour réduire l’espace de recherche.

Markova et al. [69] s’intéressent à l’apprentissage auto-supervisé de caractéristiques multi-dimensionnelles pour la problématique de recalage 2D(US)/3D(IRM) dans le cadre d’échographie

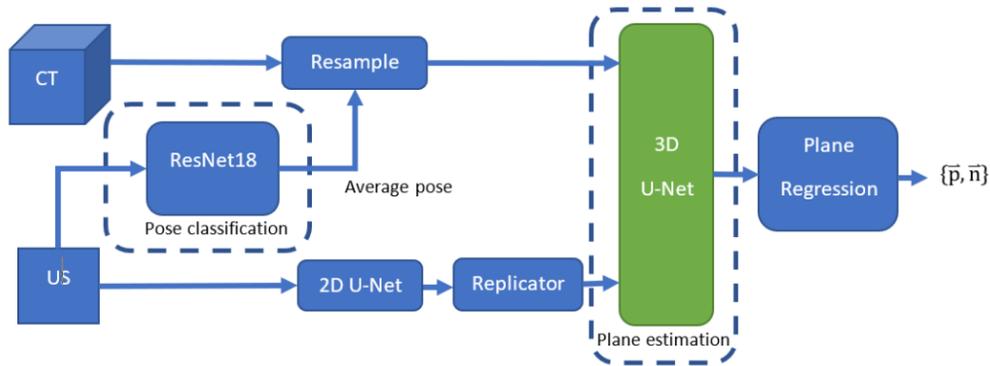


FIGURE II.14 – Illustration de la méthodologie de Wei et al. [66]

abdominale. C'est l'une des seules méthodes à étudier ce problème sans modifier/projeter les dimensionalités. Pour cela, les deux images sont fournies séparément, à deux réseaux d'encodage (2D ou 3D-U-net) illustrés en Figure II.15. Les caractéristiques ainsi obtenues sont combinées via un produit scalaire, dont seules celles dépassant une valeur seuil sont gardées. Par la suite, un algorithme classique RANSAC [70] est appliqué sur ces deux nuages de points pour estimer la transformation finale. Ils démontrent qu'un réseau plus simple, avec moins de paramètres, permet d'obtenir de meilleurs résultats avec notamment de meilleures généralisations.

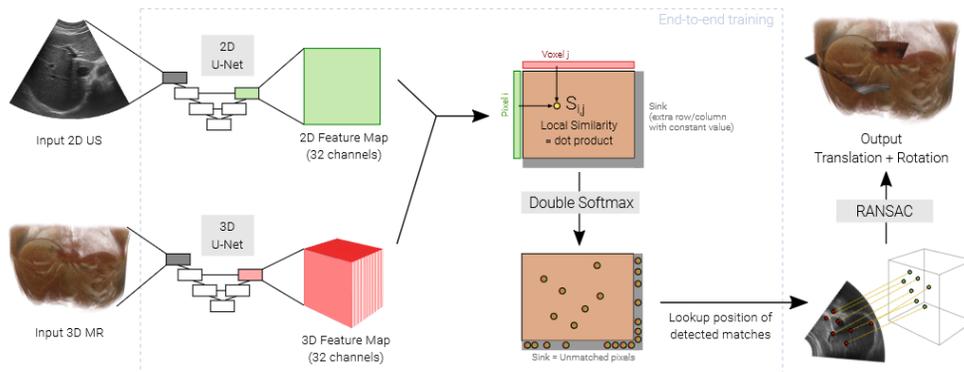


FIGURE II.15 – Illustration de la méthodologie de Markova et al. [69]

Pour rappelsons que deux méthodes précédemment décrites dans la section II.2.2.2, combinent également les entrées 2D et 3D. Guo et al [56] utilisent notamment une architecture équilibrée grâce à une étape de convolution étendue, sur l'entrée 2D, en superposant les cartes caractéristiques 2D jusqu'à obtenir la même dimension que l'entrée 3D. Le problème 2D/3D se retrouve ainsi converti en 3D/3D. Bhardwaj et al. [54] choisissent, quant à eux, de réduire le problème 2D/3D à du recalage 2D/2D en extrayant un coupe à partir du recalage rigide précédemment établi.

### II.3.3 Séquence de coupes en entrée

Les études de cette partie utilisent une séquence de coupes successives en entrées du réseau. Ces études visent principalement les problèmes de reconstruction de volume [71–78], et plus rarement des problèmes de localisation de coupes [79, 80] dans un volume. Plusieurs architectures de réseau sont décrites à travers ces méthodes, dont : des convolutions 3D sur des coupes concaténées [72, 81], des réseaux récurrents [71, 74, 79], des modèles d’attention [80], des réseaux « Transformeurs » [73, 75], ou encore des concaténations d’informations de position relatives entre ces coupes [76–78]. Un descriptif plus détaillé de certaines de ces méthodes est fourni en Table II.6.

TABLE II.6 – Comparaison de certaines études de recalage « 2D/3D », utilisant des séquences d’images en entrée du réseau. Se référer à la table des acronymes. \*Ici, le symbole  $\delta$  correspond à l’écart relatif.

Auteur Année	Recalage	Organe	Transfo	Architecture & Apprentissage	Renvois aux classes d’analyse	Résultats	Évaluation Données & GT
Prevost 2018 [76]	2D US / 2D US	Bras Jambe (Reconstr.)	Rigide	CNN $\Rightarrow \delta(\text{doF})$ SL : MSE(doF)	II.4.1.1-Positionnelle II.4.1.3-Séquentielle	<b>Final drift</b> : 11mm (5% de 200) <b>MAE</b> : 5mm, 1-3° (relatif)	2D extraites de 600 balayages : 300 000 images (15 patients) GT localisateur optique
Guo 2020 [72]	2Ds US / 3D US	Prostate (Reconstr.)	Rigide	3D-ResNet + Attention $\Rightarrow \delta(\text{doF})$ SL : MSE(doF)+PCC	II.4.1.3-Séquentielle II.4.3.2-Dynamique II.4.4.1-Pénalisation	<b>Distance</b> : 10,33 mm <b>Drift</b> : 17,39 mm	2D extraites de trajectoire base-apex (3 s) sur 600 patients GT localisateur EM
Singh 2020 [79]	2Ds IRM / 3D IRM	Foetus	Rigide	Encoder-Decoder LSTM $\Rightarrow \text{doF}$ MTL : L2(doF)	II.4.1.3-Séquentielle II.4.3.2-Dynamique II.4.4.2-Multi-tâches	<b>MSE</b> : 3,55 $\pm$ 0,22 (prédiction) 8,07 $\pm$ 0,72 (estimation)	2D extraites de simulation sur volume ( $\pm 90^\circ$ ) : 1280 séquences d’images (40 patients) GT localisateur optique
Yeung 2021 [80]	2Ds US / 3D atlas	Foetus	Rigide	CNN + pairwise comparison* attention $\Rightarrow \text{coord.}$ MTL : L2(coord.)	II.4.2.1-Paramétrage II.4.1.3-Séquentielle II.4.3.2-Dynamique II.4.5.1-Génération	<b>ED</b> : 6 mm <b><math>\delta\theta</math></b> : 10,3° <b>NCC</b> : 0,84	2D extraites de simulation sur volume plan standard (TT)
Xu 2021 [73]	2Ds IRM / 3D IRM	Foetus (Reconstr.)	Rigide	(i)Transformer $\Rightarrow \text{coord.} + \text{I}$ (ii) Reconstruction V MTL : L1(V)+L2(coord.)	II.4.1.1-Positionnelle II.4.1.3-Séquentielle II.4.2.4-Incertitude II.4.3.1-Hiérarchique II.4.3.2-Dynamique II.4.4.2-Multi-tâches	<b>ED</b> : 4,35 $\pm$ 0,90 mm <b>GD</b> : 0,074 <b>SSIM</b> : 0,916	2D extraites de simulation sur volume (10 patients)

Prevost et al. [76] proposent de reconstruire un volume US à partir de deux coupes US contiguës suivies par capteur inertiel. Ils utilisent un réseau CNN supervisé pour estimer les paramètres de transformation rigide entre deux coupes successives. Ils ajoutent plusieurs types d’informations en entrée de leur réseau (voir Figure II.16) : la paire d’images, le flot optique entre les deux coupes, et le déplacement relatif provenant du suivi inertiel de la sonde. Ils montrent que de tels ajouts sont souvent plus importants que l’architecture du réseau elle-même. Sur des zones balayées et reconstruites par des mouvements relativement simples, ils observent une dérive de reconstruction provoquée par une accumulation d’erreur lors des recalages successifs. Cette dérive illustre cependant deux limites à cette étude : la difficulté de reconstruire de l’information 3D à partir seulement de cohérence locale entre des coupes successives, et les limites d’une analyse statique (réseau non récurrent) pour un problème de suivi de trajectoire.

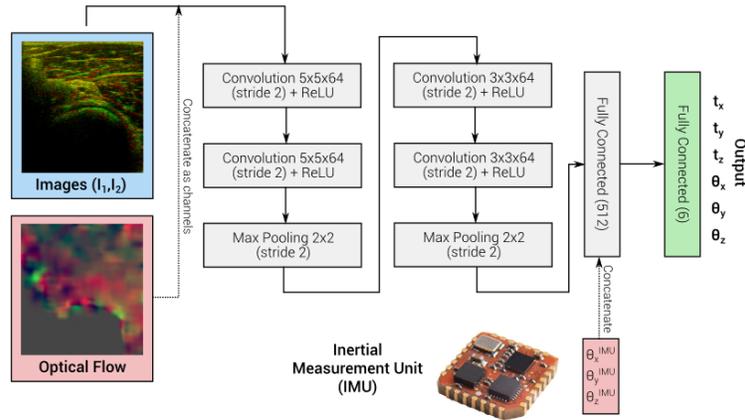


FIGURE II.16 – Illustration de la méthodologie de Prevost et al. [76]

Guo et al. [72] s'intéressent également à la reconstruction de volume US. Ils s'inspirent notamment de l'étude précédente [76] et apportent de l'information contextuelle supplémentaire à travers trois nouvelles contributions : (i) ils utilisent une séquence de 5 coupes US consécutives, concaténées sous forme d'image 3D, comme entrée de leur réseau ; (ii) ils ajoutent un réseau d'attention, pour mieux capter le phénomène de décorrélation de speckle et de mouvement hors-plan entre les coupes ; et (iii) ils rajoutent à la MSE « Mean Square Error », un terme de corrélation par degré de liberté, pour mieux s'adapter aux particularités de la trajectoire générale. Ils utilisent un réseau supervisé (3D-ResNet), illustré en Figure II.17 estimant les paramètres de transformation relatifs moyens sur cette portion de 5 coupes. Sur de nombreux balayages de prostate (base à apex) d'environ 3s, ils obtiennent des erreurs moyennes d'environ 10 mm et de dérive finale d'environ 17mm (moitié du diamètre de la prostate), ce qui est bien plus performant que d'autres études de la littérature réimplémentées sur leur même base de données [76]. Ils mettent cette amélioration sur le compte de l'apport des informations spatio-temporelles à la fois comme données d'entrée, mais également comme terme de pénalisation.

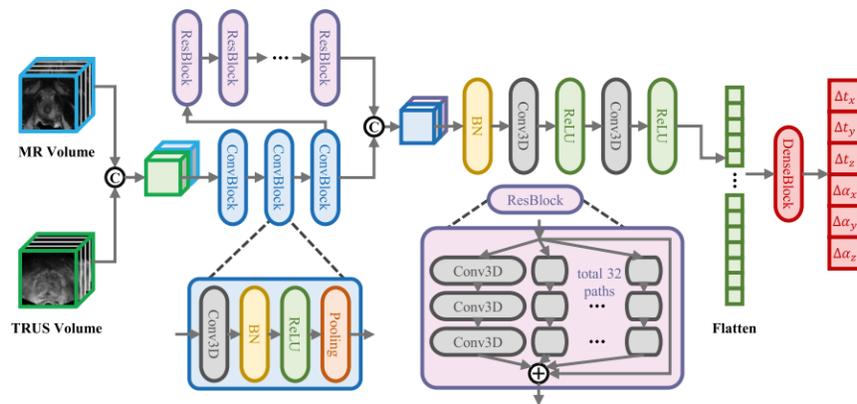


FIGURE II.17 – Illustration de la méthodologie de Guo et al. [72]

Singh et al. [79] proposent un réseau encodeur-décodeur récurrent estimant la localisation de coupes à partir de séquences de coupes IRM. Après un premier encodage des caractéristiques spatiales (via un CNN), l'encodeur temporel détermine la position de chacune des coupes de la séquence et fournit également un vecteur contextuel au décodeur qui estimera la trajectoire future. Le réseau estime pour chaque coupe, 3 types de paramètres de transformation (les rotations hors plan, la rotation du plan, et la translation hors plan), avec des fonctions d'activation adaptées pour chacun avant d'être tous ajoutés dans la fonction de coût totale. L'architecture est illustrée sur la Figure II.18. Comparativement à des études précédemment décrites [58,60] (3D-CNN sur une unique coupe d'entrée), ils démontrent que l'apprentissage à travers une analyse récurrente et partagée sur une séquence de coupes permet de mieux anticiper la dynamique de mouvement. Bien qu'évalué sur des mouvements complexes à partir de trajectoires réelles, leur méthode comporte cependant un pré-traitement important (détection de zone d'intérêt, redimensionnement, segmentation, et alignement au centre).

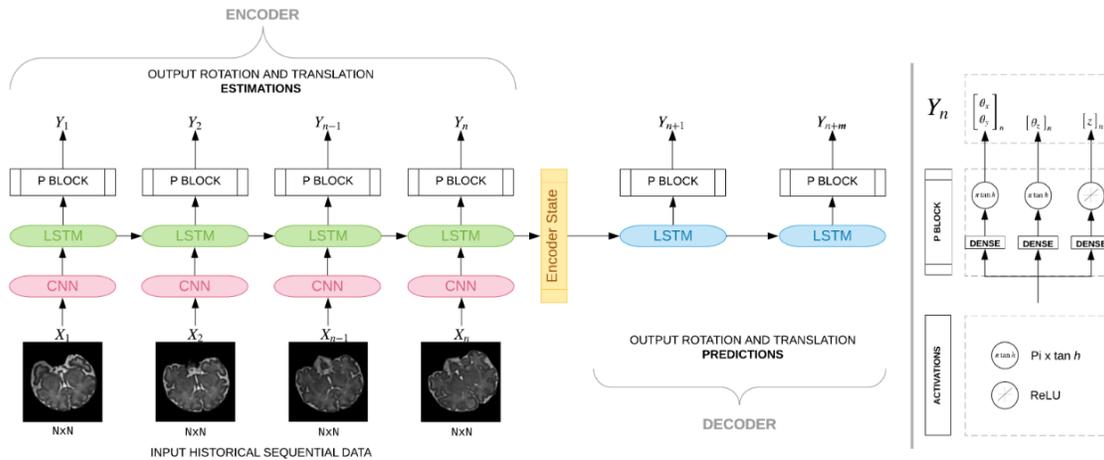


FIGURE II.18 – Illustration de la méthodologie de Singh et al. [79]

Yeung et al. [80] proposent, quant à eux, de combiner les informations contextuelles et spatio-temporelles, contenues dans une séquence de coupes d'entrée, au travers d'un réseau d'attention (voir Figure II.19). Pour cela, une comparaison par paire est calculée et pondérée par un module d'attention, avant d'estimer les coordonnées cartésiennes de chaque coupe en simultanément. Cette méthode multi-tâches permet notamment d'associer et de combiner au mieux les coupes entre elles, pour que la prédiction de position de l'une soit déduite de celles de toutes les autres. Cette utilisation des modules d'attention est justifiée par leur supériorité face aux réseaux récurrents plus coûteux et plus difficiles à faire converger. Ils évaluent leur réseau sur une grande base de coupes US extraites artificiellement de volumes, et fortement prétraitées. Leurs erreurs de localisation dans un atlas 3D, sont inférieures à d'autres études de la littérature [58], et leurs résultats se généralisent très bien sur des données cliniques (coupes non standard) annotées par des experts.

Xu et al. [73] proposent de reconstruire un volume à partir d'une séquence de coupes IRM en utilisant des réseaux « Transformeurs ». À partir de : (i) la séquence d'images (multi-vues); (ii) leur localisation respectivement précédemment estimée; et (iii) leur encodage positionnel au sein de la séquence; le réseau estime à la fois la séquence de localisation mise à jour ainsi que l'incertitude

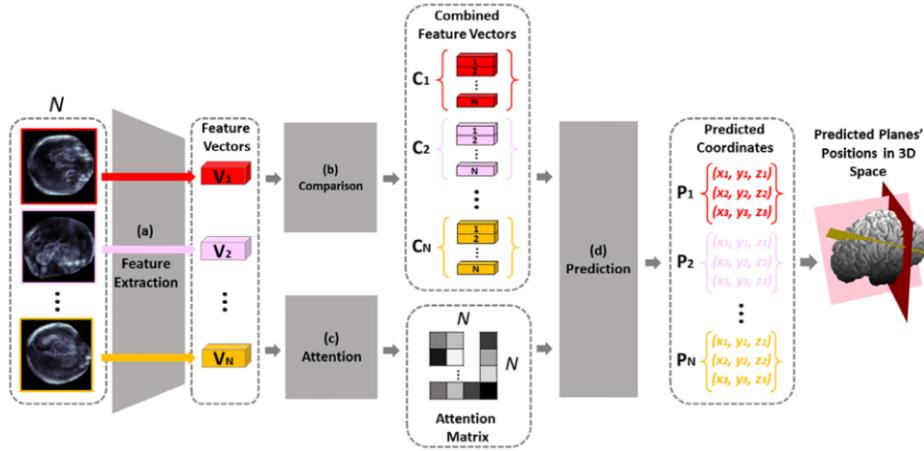


FIGURE II.19 – Illustration de la méthodologie de Yeung et al. [80]

pour chacune, afin d'estimer le volume reconstruit à partir des positions de ces coupes. Une nouvelle séquence de coupes, extraites du volume ainsi reconstruit à partir des localisations précédemment prédites, sera par la suite comparée à la séquence d'entrée. Ce principe rappelle les termes d'auto-pénalisation couramment utilisés comme régularisation et cohérence spatiale durant l'optimisation, mais utilisé ici directement comme entrée du réseau. Ce processus est répété itérativement, et permet d'améliorer alternativement la transformation et la reconstruction du volume. Avec cette méthode, illustrée en Figure II.20, ils démontrent une meilleure précision de recalage par rapport à des méthodes utilisant des modules de convolution [58] ou d'attention [80], précédemment citées dans ce manuscrit. Ils démontrent également la contribution du Transformeur et d'un apprentissage multi-tâches (reconstruction et localisation) bénéficiant d'une optimisation conjointe et hiérarchique.

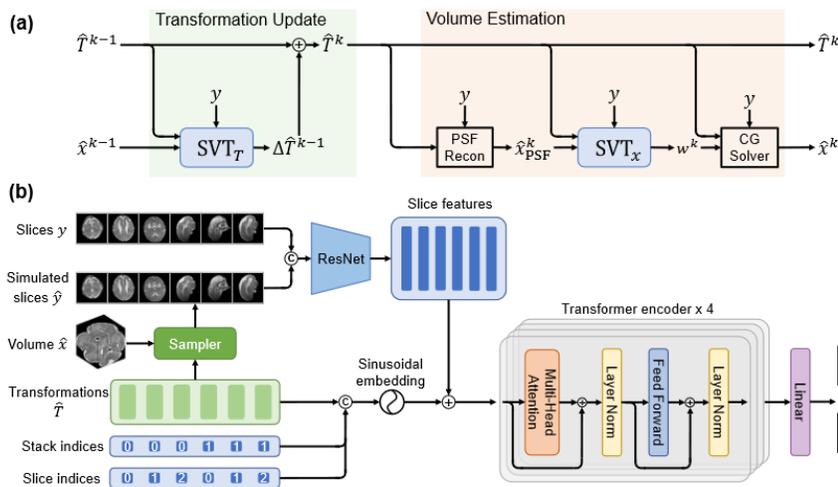


FIGURE II.20 – Illustration de la méthodologie de Xu et al. [73]

## II.4 Analyse et synthèse de l'état-de-l'art

Nous résumons ci-dessous les différentes idées de l'état de l'art qui nous ont inspirées pour la suite de nos travaux. Nous les avons répertoriées en 5 catégories selon qu'elles concernent le type : d'entrée, de sortie, de structure, d'optimisation ou encore d'évaluation.

### II.4.1 Types d'entrée

Dans l'état-de-l'art, plusieurs types d'entrée sont fournies au réseau afin de faciliter la tâche de recalage. Nous retenons les informations positionnelles, spatiales, séquentielles, et structurelle (voir Figure II.21).

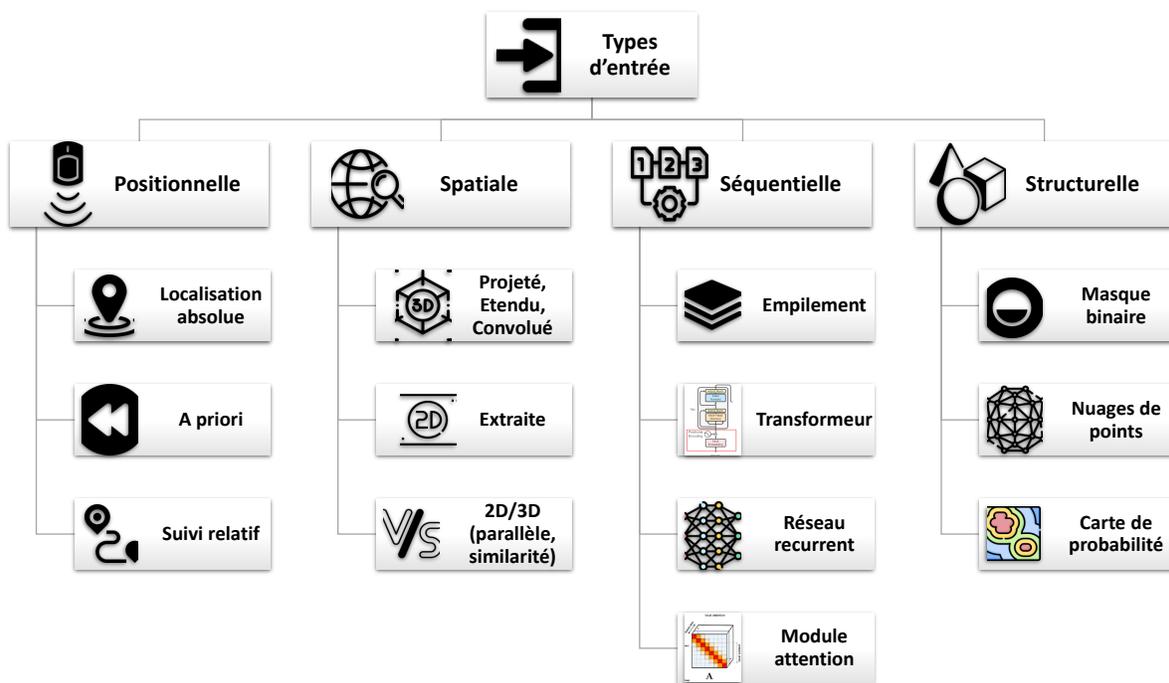


FIGURE II.21 – Taxonomie du type d'information d'entrée : positionnelle, spatiale, séquentielle et structurelle. L'analyse et la description de celles-ci diffèrent selon leurs différentes natures, intégrations ou représentations possibles.

#### II.4.1.1 Entrées positionnelles

Que ce soit pour des méthodes classiques ou par apprentissage profond, des informations supplémentaires aux images à recalcer sont également fournies en entrée de la méthode. Ces informations supplémentaires sont généralement issues de capteurs (électromagnétique, optique, inertiel IMU, etc.), localisant la sonde soit de manière absolue (vers un référentiel physique) ou de manière relative (suivi de mouvement). D'autres types d'informations peuvent concerner directement la localisation de l'image 2D (et de la zone d'intérêt), notamment à partir d'estimation de positionnement

antérieur que l'on nommera « a priori ». Nous définissons l'ensemble de ces informations comme « positionnelles ».

La plupart des études utilisent ces informations positionnelles pour initialiser leur recalage. Elles utilisent pour cela principalement : la localisation de la sonde [20,34,45,49,50] ; ou le positionnement précédent [17,51] parfois combiné à un suivi de sonde [51]. Pour les méthodes classiques [34,45,50–52], cette initialisation est une étape préliminaire intégrée au processus d'optimisation itératif, souvent indispensable pour éviter les problèmes de divergence ou de stagnation. Pour les méthodes par apprentissage profond [54,55], ces initialisations permettent de réduire l'espace de recherche et parfois même la dimensionnalité du problème (2D/2D), en transformant le volume pour en extraire une coupe au plus proche de la bonne localisation.

Sans être des initialisations directes, ces informations peuvent également être ajoutées au réseau comme entrées additionnelles. Certaines approches les intègrent à partir d'entrées vectorielles concaténées sur une couche du réseau [74,76,81], tandis que d'autres les considèrent comme des encodages positionnels [73,75] indispensables lors d'une architecture avec Transformeurs. Contrairement aux initialisations où ces informations « positionnelles » sont complètement intégrées et indissociables de l'image ainsi obtenue, une entrée additionnelle laisse la possibilité de la combiner et pondérer au mieux aux autres types d'entrée du réseau : par exemple, un suivi IMU peut se combiner avec le déplacement dans le plan observé entre plusieurs images (ou flot optique) [76,78] et alors induire une estimation du contexte hors-plan. On retrouve notamment ces informations positionnelles dans des travaux concernant des problématiques de reconstructions de volumes à partir de plusieurs coupes [73–76,78].

### II.4.1.2 Entrées spatiales

Les informations « spatiales » peuvent faciliter l'apprentissage du réseau car elles permettent de mieux considérer l'espace de destination (image de référence) pour les problématiques de recalage 2D/3D. Elles s'illustrent dans la littérature sous plusieurs formes : le volume entier [54,66,67], une sous partie du volume [56], ou plusieurs coupes concaténées [72,81].

Fournir les deux types d'entrée (espace de départ 2D, espace de destination 3D) est systématique lors d'approches classiques de recalage classique. En revanche, ce n'est utilisé que dans peu d'études de recalage par apprentissage profond [54–57,66,67,69]. De telles données d'entrée pourraient apporter un a priori sur l'espace de recherche, mais combiner ces deux types d'informations (de dimensions différentes) est complexe de par la nature différente des caractéristiques obtenues et leur dimensions déséquilibrée. La plupart de ces études se ramènent à la même dimension soit par extension [66] (3D/3D), par extraction [54,55,57] (2D/2D), en manipulant les dimensions des convolutions [56,67] (2D suivies de 3D). Très peu d'études confrontent les caractéristiques 2D et 3D directement [69].

Les informations spatiales sont également très utilisées avec le terme de pénalité, permettant de préserver une cohérence spatiale sur les formes et les structures anatomiques. Cette contribution sera détaillée dans la catégorie « Types d'optimisation » (voir section II.4.4).

### II.4.1.3 Entrées séquentielles

Des séquences de coupes successives peuvent également être fournies en entrée du réseau. On parle d'information « séquentielle ». Elles présentent l'avantage d'accumuler de l'information spatiale, à partir de plusieurs plans au cours du temps. On parle de contexte spatial et temporel combiné, soit spatio-temporel.

Les études utilisant des informations séquentielles concernent notamment des problématiques de reconstruction de volume [71–76] ou plus rarement des problèmes de localisation de coupes [79, 80] dans un volume, où la prédiction de localisation de chacune des coupes est alors pondérée et conditionnée par le contexte plus général de toutes les autres coupes. Pour cela, la bonne intégration du contexte spatio-temporel est traitée de différentes manières : à travers des convolutions 3D sur des coupes concaténées [72, 81], de réseaux récurrents [71, 74, 79], de modèles d'attention [80], ou encore de réseaux « Transformeurs » [73, 75]. Ces différents types d'architecture seront décrites plus en détails dans la catégorie « Types de structure » (voir section II.4.3).

### II.4.1.4 Entrées structurelles

On considère comme informations « structurelles », les informations de contours d'un objet issues notamment de sa segmentation dans une image. Plus précises et spécifiques que des informations spatiales, nous décidons alors de les traiter de manière supplémentaire à la catégorie décrite précédemment. Plusieurs études de recalage d'images médicales se basent sur des informations de segmentation et ont montré qu'elles pouvaient faciliter le processus de recalage [38–41, 43, 47, 82–84]. Nous avons identifié trois façons principales d'utiliser la segmentation pour améliorer le recalage : (i) soit comme « caractéristique » en donnée d'entrée supplémentaire au réseau, (ii) soit comme « label » lors du processus d'optimisation dans un cadre auto-supervisé, (iii) soit comme « tâche conjointe » dans un cadre multi-tâches. Les deux dernières contributions seront détaillées dans la catégorie « Types d'optimisation » (voir sections II.4.4.1 et II.4.4.2). Cette segmentation est le plus souvent définie au préalable, par un réseau supplémentaire ou manuellement. Dans tous les cas, la bonne interprétation de cette information est dépendante de la qualité de la segmentation obtenue.

Un a priori de la forme segmentée est souvent utilisé en entrée du réseau [38, 40, 43, 83]. Certaines de ces études utilisent les segmentations comme seuls données d'entrée [38, 40, 83], tandis que d'autres les combinent avec les images initiales non-segmentées [43] pour ne pas perdre toute l'information contenue dans l'image. Ces études démontrent généralement que l'étape de segmentation permet d'affiner le recalage par rapport à un recalage n'utilisant que des images brutes. La plupart de ces études étudient une tâche de recalage multimodal, où l'utilisation de caractéristiques anatomiques (de forme) est plus pertinente que l'utilisation de caractéristiques d'intensité pouvant être altérées lorsque les modalités d'imagerie sont différentes. Cependant peu ont étudié le cas monomodal [83] ou multi-dimensionnel, correspondant à notre problématique.

Ces segmentations se présentent le plus souvent sous la forme de masques binaires [40, 43, 83], de cartes de probabilité [54], mais des études récentes ont montré que la représentation sous forme de nuages de points (en coordonnées 3D) est plus adaptée [38, 48]. L'analyse à travers un réseau spécifique aux nuages de points (PointNet) permettrait de considérer davantage la rotation interne d'un objet (grâce à des connexions de maillages intra-volumique ordonnées), et de s'affranchir des différences d'isotropie et de vues des plans (souvent perdues dans des couches de convolution).

## II.4.2 Types de sortie

Il existe de nombreuses représentations du résultat du recalage. Parmi les plus fréquentes dans la littérature on retrouve : le paramétrage de transformation (rotation, translation, etc.) [56, 67, 74, 76, 79] ; des champs de déplacements (recalage déformable) [38, 40, 43, 54] ; des coordonnées (points caractéristiques [48, 68, 69, 73, 80] ou plan [63, 66, 68]) ; ou l'utilisation de classes [63, 66, 85] avec parfois des notions d'incertitudes [62, 68, 73, 86]. Seulement certaines d'entre elles seront décrites plus en détails ci-dessous.

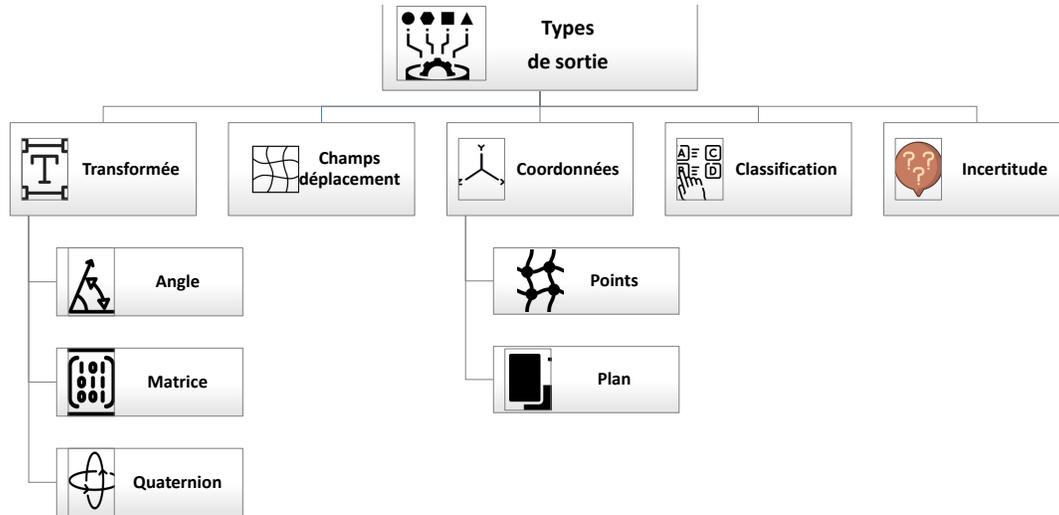


FIGURE II.22 – Taxonomie du type d'information de sortie.

### II.4.2.1 Paramétrage des transformations

La majorité des études traitent ce problème sous la forme d'un réseau « image-vers-paramètres ». Cela revient à résoudre un problème non-déterministe où plusieurs solutions peuvent être valables pour la même entrée, complexifiant d'autant plus le problème. Certaines études s'intéressent particulièrement aux différentes formes de représentation des paramètres de transformation et notamment les rotations [58, 60, 62, 85, 87–89]. Les formes les plus répandues sont : les quaternions [86], les angles d'Euler [74], les matrices de rotation [64] ou le couple sin/cos. Les quaternions sont retenus pour des estimations de grands angles et la correction de l'effet de « Gimbal lock » mais pas nécessairement pour de petits angles.

Cependant, la non-continuité et la non-périodicité de certaines de ces représentations restent des problèmes à considérer. Pour les résoudre, certaines approches proposent de contraindre l'intervalle de solutions en utilisant des fonctions d'activation multipliant  $\pi$  avec la fonction de tangente hyperbolique [60, 79], d'autres proposent d'assurer l'ortho-normalisation des matrices de rotation via l'algorithme de Gram-Schmidt [64].

### II.4.2.2 Coordonnées

La représentation des résultats de recalage sous forme de coordonnées (de points [48, 68, 69, 73, 80] ou plan [63, 66, 68]) présente différents avantages : elle permet de faire une optimisation plus robuste en moyennant les erreurs sur plus de points que seulement les paramètres de transformation, elle permet d'associer des éléments de même nature [69] (points physiques) contrairement à une confrontation avec une information de l'image (pixel ou voxel), et elle permet d'envisager un recalage non-rigide plus facilement. La prédiction et la détection de coordonnées de points se rapprochent de l'idée du recalage classique géométrique. La combinaison d'images d'entrée et de nuages de points serait d'autant plus pertinente car elle permettrait de tirer profit des avantages des deux représentations, tout comme le fait un recalage hybride dans les méthodes classiques. De plus, les prédictions sur les points permettent de considérer des termes de pénalité structurelle [38, 48] (surfactive ou volumique), en plus des mesures de similarités iconiques. Cependant, cette prédiction nécessite des « réseaux de points » spécifiques (Point-Net), plus adaptés pour garder une cohérence « points-vers-points ».

La prédiction sous forme de plans permet de traiter le problème sous la forme « image-vers-image ». Ces plans peuvent se représenter sous la forme de coordonnées cartésiennes des points du plan (image à 3 canaux) ou sous la forme d'une image 3D binaire segmentant ce plan [68]. Cet encodage permet de garder une cohérence spatiale entre les coordonnées des points de sortie et des images d'entrée au moment des convolutions. Cependant, pour garantir la coplanarité de points plans, des travaux proposent d'utiliser une pondération des points par PCA [68].

### II.4.2.3 Classification

Du fait de la continuité de la plupart des représentations des résultats de recalage, mentionnées précédemment, une approche de régression est souvent privilégiée. Cependant, certaines études implémentent soit une partie [66] soit la totalité [63, 85] de leur modèle de recalage en se basant sur le principe de classification.

L'échantillonnage et la discrétisation de ces valeurs continues permettent de représenter la localisation sous formes de classes. Certaines choisissent une répartition polaire uniforme où les coupes sont représentées par leurs coordonnées du centre du plan et de son vecteur normal [63]. Cependant la qualité du recalage prédit dépendra du nombre de classes définies (échantillonnage).

La classification peut également être envisagée comme une première étape, d'initialisation au recalage. Certains déterminent des coupes standards comme classes, et initialisent le recalage à partir des paramètres de localisation moyens de la classe prédite [66]. Ce concept serait également réalisable dans notre cas, où biopsies (notamment aléatoires) peuvent s'avérer très stéréotypées créant alors une cartographie de prélèvement de 12 zones (base, milieu, apex, à droite et à gauche).

### II.4.2.4 Incertitude

Des méthodes existent (Monte Carlo Dropout, Network ensembliste, test time augmentation) pour estimer le niveau de confiance à chaque prédiction. Cependant, elles nécessitent pour cela plusieurs prédictions (différents réseaux, avec différentes initialisations ou dropout), à moyennner, et sont donc coûteuses en temps. De plus en plus de méthodes visent à prédire en temps-réel et en complément de leur prédiction, un indicateur de la confiance de cette prédiction. Cette notion

de probabilité définit la confiance accordée à la prédiction et est estimée soit par le réseau lui-même [62, 73, 86] soit par un autre réseau [68]. L'apprentissage de ces estimations d'incertitudes ne repose pas sur des vérités terrain, mais sur l'optimisation de la fonction de coût de la prédiction principale.

Dans un processus itératif ou hiérarchique [62, 73] ces mesures servent à quantifier la qualité des résultats obtenus et à les pondérer soit pour l'étape suivante [62], soit lors du calcul d'optimisation [73]. Ce concept peut également s'avérer intéressant sur des données séquentielles, concernant un suivi de trajectoire ou de reconstruction de volume, où les erreurs d'accumulation peuvent très vite fausser les résultats.

Certaines méthodes utilisent directement un réseau de type « Hypothèse-multiple » [86] pour choisir la prédiction qui aura la plus petite incertitude parmi toutes les hypothèses. D'autres méthodes [68] réutilisent un deuxième réseau, de même architecture mais préalablement entraîné seulement sur des données d'entrée basées sur les vérités terrain bruitées, permettant d'affiner la prédiction du premier réseau.

### II.4.3 Types de structure

L'architecture d'un réseau est choisie pour être la plus adaptée aux données d'entrée, de sortie, et à la problématique. Certaines architectures sont très répandues dans la théorie du deep learning et ont déjà fait leurs preuves dans certaines situations. Les études de notre état-de-l'art exploitent des architectures déjà connues, et les adaptent à de nouvelles applications. Nous décrivons ci-dessous des concepts applicatifs généraux (voir Figure II.23).

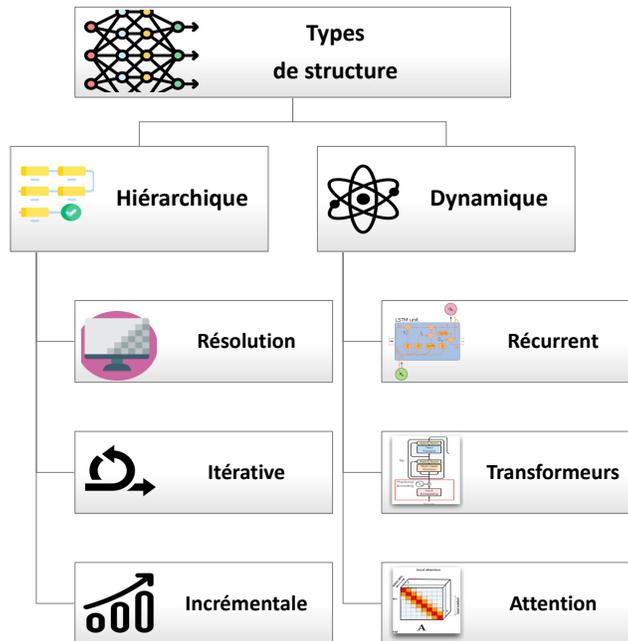


FIGURE II.23 – Taxonomie du type de structure.

### II.4.3.1 Structures hiérarchiques

On considère comme processus « hiérarchique », un réseau structuré en plusieurs étapes distinctes permettant d'améliorer la précision du recalage prédit. Les successions d'étapes les plus utilisées dans la littérature concernent notamment : des résolutions différentes [65, 68], une incrémentation de la difficulté [40, 41, 54, 57, 65, 66], ou parfois plusieurs itérations d'une même étape [57, 62, 67, 73].

Parfois une étude du problème sur la totalité de l'espace et avec une résolution importante peut s'avérer plus complexe, plus coûteuse en temps de calcul, et plus difficile à faire converger. Certains [65] entreprennent alors de décomposer leurs problèmes en sous-problèmes, plus simples et à portée réduite, avant de les rassembler sur une même couche du réseau. D'autres [68] dégradent la résolution, dans un premier temps, pour y détecter une région d'intérêt, sur laquelle la résolution sera rétablie par la suite. De plus, dans la plupart des méthodes recensées dans la littérature, des étapes de prétraitements (manuels ou automatiques) sont souvent nécessaires pour extraire des régions d'intérêt permettant une étude plus localisée. Ces prétraitements seront décrits plus en détails dans la catégorie « Types d'évaluation » (voir section II.4.4).

Incrémenter la difficulté étape par étape se rapproche de l'idée d'initialisation et permet de réduire l'espace de recherche initial. On trouve des passages de paramètres plan vers des paramètres hors-plan [65], des passages de localisations discrètes vers des localisations continues [66], ou encore des passages de recalage affine vers du recalage déformable [40, 41, 54].

Enfin, réitérer le même réseau plusieurs fois durant l'entraînement, est également envisagé en réutilisant les prédictions précédentes comme nouvelle entrée. Des études [57, 62, 73] démontrent que le nombre d'itérations utilisées a un impact significatif sur la qualité du recalage final. Cependant, ces étapes nécessitent une transformation des entrées (étapes de rééchantillonnage et d'interpolation d'un volume), pouvant prendre un certain temps de calcul [57, 73]. D'autres études [67, 74] réalisent également plusieurs itérations et optimisations, cette fois-ci au moment de l'inférence, pour volontairement contraindre le réseau à s'adapter à un nouveau type de données. On parle d'optimisation adaptative temps-réelle et auto-supervisée. Bien que ces études montrent une amélioration de la généralisation des résultats, ces étapes supplémentaires nécessitent de nouvelles alternatives aux vérités terrain (pas toujours disponibles ou adaptées) et sont parfois trop coûteuses pour de la prédiction en temps-réel.

### II.4.3.2 Structures dynamiques

On considère comme structure « dynamique », tout réseau appliqué sur une séquence logique d'éléments d'entrée, dont l'architecture permet l'interaction entre chacun d'entre eux. Ce type de structure est notamment utilisé sur des séquences d'entrées temporelles concernant des problématiques de langage, de reconstruction d'espace, ou d'analyse de trajectoires.

Plusieurs études utilisant des réseaux récurrents (RNN) ont été recensées dans cet état de l'art [71, 74, 79]. Les modèles les plus utilisés sont notamment les LSTM (« Long Short Term Memory »), avec un facteur d'oubli, qui leur permettent de retenir plus ou moins l'information sur une fenêtre de temps évolutive. Cependant, beaucoup font remonter la difficulté de les entraîner et de les faire converger sans sur-apprentissage [90]. Les modèles d'attention font de plus en plus leur apparition dans des études de recalage [69, 72, 80], en réponse aux limites des réseaux récurrent. Ils permettent

notamment de combiner et de pondérer chacun des éléments de la séquence, en fonction de leur interaction et similarité entre eux, quantifiées grâce à des mesures de produits scalaires. Certains l'interprètent et le comparent aux mécanismes d'attention de décorrélation de speckle [72].

Les réseaux de types « Transformeurs » [91] sont de plus en plus présents depuis quelques années [73, 75]. Nous les dissocions du paragraphe précédent car ils combinent à la fois des modules d'attention et un encodage positionnel de la séquence. Cet encodage traduit initialement la position de l'élément au sein de la séquence, mais on trouve de plus en plus de représentations diverses : indication d'une orientation standard [73], suivi du mouvement relatif [75], ou même des a priori de résultats précédents [73, 75]. Ces informations correspondent notamment aux informations que l'on a définies précédemment par le type « positionnelle » (voir section II.4.1). À travers ces encodages (adaptés à chaque problématique), les informations positionnelles peuvent alors contribuer au réseau différemment d'une simple concaténation d'entrée additionnelle.

### II.4.4 Types d'optimisation

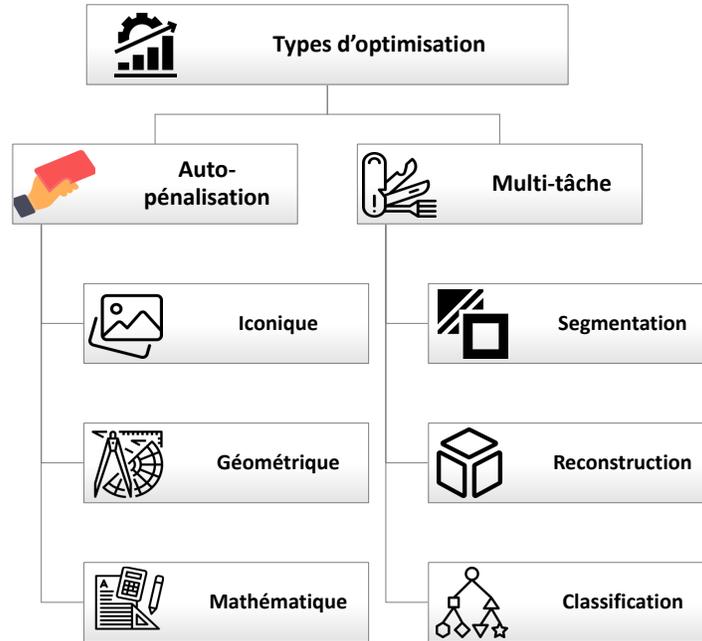


FIGURE II.24 – Taxonomie du type d'optimisation.

#### II.4.4.1 Auto-pénalisation

Des termes de pénalités sont souvent utilisés, dans les fonctions de coût, pour les rendre plus optimales et plus adaptées face aux problèmes à résoudre. Les termes d'auto-pénalisation se présentent comme un bon compromis entre apprentissage supervisé et non-supervisé, permettant d'extraire des labels internes (provenant des données d'entrée) qui seront utilisés pour le calcul d'optimisation de l'entraînement. On évalue ainsi, d'une manière détournée, des éléments que l'on ne pouvait utiliser au départ en s'affranchissant d'une quantification sous forme de vérité terrain, parfois difficile à obtenir. Ces termes de pénalité permettent parfois d'ajouter une mesure plus pertinente et plus révélatrice de la qualité du recalage, en respectant des cohérences structurelles par des mesures iconiques, géométriques, ou mathématiques. La bonne pondération de tous ces termes dans la même fonction de coût reste cependant à déterminer.

L'auto-pénalisation la plus fréquente pour des problématiques de recalage d'images reste la mesure de similarité iconique (SSIM ou NCC) [40, 41, 54, 56, 57, 67, 73]. À partir de modules de rééchantillonnage différentiable (STN [92]), une coupe extraite de l'espace de destination à la localisation prédite sera comparée à la coupe initiale. Cette pénalité nécessite d'avoir ce contexte global de destination à disposition, et est systématiquement utilisée lorsque les deux espaces (départ et destination) sont fournis en entrée [40, 41, 54, 56] ou reconstruits [71, 73]. Cette même méthode est également utilisée sur les données segmentées plutôt que sur les données d'entrée, notamment dans

un cadre multimodal [40, 41, 43, 47], pour mesurer le recouvrement (DSC, NCC) des segmentations transformées.

D'autres pénalisations permettent de renforcer l'aspect structurel en préservant une cohérence sur la réalité géométrique et anatomique de l'espace de destination. Certains [68] se conforment à la planéité géométrique en minimisant la 3ème (moindre) composante principale d'un plan prédit ; d'autres contraignent la distance de points surfaciques et volumiques via des contraintes biomécaniques [38] ; d'autres encore utilisent des classificateurs quantifiant la probabilité de vraisemblance d'un volume reconstruit [71].

Inspirées par les méthodes classiques, des études par apprentissage profond rajoutent des termes de pénalisation mathématique de type cohérence inverse [78]. Cela permet de vérifier la robustesse de la prédiction du recalage en sens inverse. D'autres études [71, 72] privilégient des mesures par degré de liberté (avec PCC) plutôt que des moyennes qui font perdre la spécificité de certaines variations de mouvements.

#### II.4.4.2 Multi-tâches

Certaines études utilisent plusieurs chemins de sortie pour établir plusieurs types de prédiction : d'où la désignation « multi-tâches ». Pour chaque tâche, une fonction de coût, adaptée au type de données, est déterminée avant d'être pondérée et ajoutée dans la fonction de coût totale. Ce partage, au niveau de l'architecture et au niveau de l'optimisation, s'illustre le plus souvent avec des tâches de segmentation, de reconstruction de volume, ou de dissociation de classes.

Les tâches de segmentation et de recalage étant étroitement liées et souvent complémentaires, c'est l'association multi-tâches la plus fréquente de notre état de l'art [39, 82–84]. En effet, l'a priori de déformations du recalage peut aider à la segmentation, ou inversement, l'a priori de segmentation peut initialiser un recalage lors d'une transformation importante. Réaliser ces tâches, de manière simultanée et conjointe, permet alors de combiner les avantages de chacune. Ces méthodes reposent sur des architectures avec des encodages partagés où les poids sont identiques et optimisés en même temps, et des fonctions de coût combinant ces deux termes. Ces études ont montré que la stratégie multi-tâches surpasse celle de la tâche unique [39, 82, 84].

Une estimation de reconstruction d'un volume à partir des résultats de localisation de plusieurs coupes, conjointement prédites durant la tâche de recalage, est également fréquente [71, 73]. Cette reconstruction du volume peut permettre d'extraire de nouvelles séquences de coupes prédites qui pourront être comparées aux images d'entrée [73], ou de vérifier la cohérence sur les formes anatomiques reconstruites [71].

Enfin, les prédictions multiples peuvent être considérées comme des classes de nature différentes selon si elles traitent des paramètres différents (translation ou rotation) [60, 62–64, 79] ou des mesures d'incertitude additionnelles [62, 73, 86]. Certaines études dissocient les paramètres de transformation à prédire, ne partageant pas les mêmes ordres de grandeur, en plusieurs sorties pour normaliser leur impact durant l'optimisation. À travers différentes fonctions de coût spécifiquement adaptées aux types de mesures.

### II.4.5 Qualité des données et d'évaluation

La qualité des données et des procédés de validation est cruciale pour la bonne interprétation des résultats. Dans cette partie, nous prenons soin d'analyser plus en détail les conditions expérimentales des travaux présentés en section II.2.2, qui portent sur des problématiques de recalage 2D(US)/3D(US) par apprentissage profond pour la navigation de biopsie prostatique, donc similaires aux nôtres. Selon le type d'évaluation choisi, beaucoup de résultats sont finalement à interpréter avec précaution. Nous mettons en évidence certains critères d'importance ci-dessous (voir Figure II.25).



FIGURE II.25 – Critères d'importance durant l'évaluation du recalage.

#### II.4.5.1 Génération des données d'entrée

La plupart des méthodes n'ont pas accès aux images (2D US) directement issues des procédures cliniques ni à leur localisation précise dans l'espace (3D). Elles doivent donc établir un processus de simulation pour définir leurs données d'entrée et de sortie. Des étapes de prétraitements sont également souvent envisagés avant d'utiliser les données d'entrée telles quelles.

Dans les problématiques de recalage 2D/3D, beaucoup d'études simulent artificiellement des coupes en les rééchantillonnant à partir de volumes cliniques fixes [51, 52, 56, 67, 73]. Ces simulations proviennent de trajectoires : le plus souvent simplifiées [52, 56, 72, 76, 78, 81], ou avec des intervalles de variations limités [51, 52, 56], avec parfois une seule orientation de coupe (canonique/standard) [58, 62, 66, 67, 73, 80], ou plus rarement de trajectoires complexes et réalistes [55]. D'autres méthodes encore [20, 45, 52] n'ont pas accès à une acquisition 3D, et reconstruisent un volume à partir d'un balayage d'acquisitions 2D combinées à la localisation de la sonde. Les coupes rééchantillonnées présentent les mêmes informations anatomiques (sans déformation) et les mêmes bruits (sans cohérence de speckle) que le volume dont elles sont extraites. Ce même volume repré-

sente également l'espace de destination dans lequel on cherche à les localiser. Cette simulation des données par rééchantillonnage ne reflète donc pas toute la difficulté et le réalisme rencontrés dans des conditions cliniques réelles, mais sert comme un premier niveau de difficulté d'évaluation de la tâche de recalage.

Les prétraitements de données permettent d'améliorer significativement la qualité des images d'entrée et de réduire l'espace de recherche. De nombreuses études de l'état de l'art réalisent pour cela une étape de définition (automatique ou manuelle) de zone d'intérêt [54, 55, 63, 65], parfois à partir de segmentation [54]. D'autres utilisent un alignement préalable, permettant de s'affranchir de certaines dimensions de recherche, notamment celle du plan pour ne se concentrer que sur la dimension hors-plan. Les initialisations [34, 45, 50, 54–56] de recalage à partir d'informations positionnelles, précédemment citées en section II.4.1.1, sont également très fréquemment incluses dans les étapes de prétraitements. À notre connaissance, ces études n'ont cependant pas évalué la généralisation de leur réseau sur des données non prétraitées.

### II.4.5.2 Génération des vérités terrain

La confiance accordée aux vérités terrain (données de sortie) est aussi un élément très important lors de l'évaluation. Ces vérités terrain (segmentations, paramètres de transformation, etc.), sont le plus souvent construites avec les procédés de génération de données décrits dans section II.4.5.1, ou obtenues grâce à la localisation des équipements. Parfois, les travaux se comparent simplement aux résultats obtenus par des méthodes comparables de la littérature.

Bien que des données simulées et extraites d'un même volume soient très simplifiées, leur localisation est parfaitement contrôlée et vérifiée comme vérité terrain.

Des localisateurs (magnétique, optique, etc.) sont utilisés dans certaines études [54, 55], pour localiser la position de la sonde échographique en parallèle de l'enregistrement du flux échographique durant des procédures cliniques. En revanche, cette localisation ne donne que la position de la sonde qui a acquis l'image, dans le repère « physique ». La véritable position de l'organe par rapport à la sonde reste inconnue à cause des déformations et bougés anatomiques. Une utilisation de données enregistrées en clinique, avec des déformations réalistes, mais sans vérité terrain rigoureuse peut s'avérer tout aussi biaisée qu'une utilisation de données simulées et simplifiées.

Pour finir, certains travaux utilisent des résultats provenant de méthodes comparables et reconstruites (FEM [38], recalage ITK [54]) pour définir des vérités terrain qui devront servir de supervision aux méthodes à établir. Cela ne garantit pas la qualité de la vérité terrain estimée par cette autre méthode.

### II.4.5.3 Validation

La validation de la qualité du recalage doit être évaluée avec une mesure pertinente, évaluée sur un grand nombre d'objets caractéristiques indépendants du processus d'entraînement, avec des annotations précises. Une validation pertinente dépend de la cohérence de la métrique choisie par rapport aux données, et des objets caractéristiques établis.

Parmi les métriques les plus utilisées dans notre état-de-l'art, on retrouve le plus souvent une analyse géométrique entre plusieurs points (TRE), une analyse iconique sur la similarité des images (SSIM, NCC), et une analyse statistique sur la régression obtenus (MAE, MSE, R2, PCC). Des

études [58,60,89] ont notamment comparé les métriques statistiques sur les angles et recommandent l'utilisation de mesures géodésiques plus spécifiques aux quaternions ou aux matrices de rotation.

Les objets caractéristiques de validation avec leurs annotations sont le plus souvent définis à partir de mesures indépendantes de l'apprentissage. Les plus utilisés en recalage sont les fiduciels anatomiques, définis par des experts. Ces fiduciels sont souvent présents en faible nombre (entre 3-5 par image), et sont parfois tous situés sur un même plan ou surface, ne reflétant pas parfaitement la qualité en 3D. Pour finir, les annotations manuelles nécessitent cependant un certain contrôle de la répétabilité inter et intra-expert qui n'est pas systématiquement vérifiée ou publiée.

#### II.4.5.4 Déroulement

L'idée de réaliser un recalage continu, automatique et sans intervention externe, est assez peu présente dans ces travaux. Cela permet cependant de s'approcher davantage de la complexité d'une situation clinique réaliste.

De nombreuses études de l'état-de-l'art utilisent des étapes de prétraitement ou d'initialisation de leur recalage [34, 37, 45, 50, 54, 55]. Cependant, très peu de méthodes incluent ces traitements de manière automatique dans le processus. La plupart se basent sur une intervention externe et manuelle [34, 37, 45, 54, 55] ne permettant pas une validation réaliste parfaitement indépendante. D'autres encore nécessitent la localisation de la sonde (dans le référentiel "physique") [20, 34, 45, 49, 50], présentant la contrainte d'avoir recours à un équipement avec capteur (et préalablement calibré). À notre connaissance, ces études n'ont cependant pas évalué la généralisation de leur réseau sans ce type de données, qui n'est pas accessible à tous en routine clinique.

L'idée de réaliser un recalage continu est de plus en plus envisagée en utilisant les résultats du recalage précédent comme initialisation [51, 52, 57]. Certaines méthodes évaluent même les performances sur des simulations séquentielles de procédures complètes pour mieux évaluer un impact clinique réaliste [52, 57]. Cependant, certaines méthodes [56] proposent une évaluation qui repose sur une initialisation des résultats précédents à partir de vérités terrain bruitées, très avantageuses. Excepté en phase d'apprentissage, cela ne serait pas faisable lors d'un déroulement clinique réaliste.

Finalement, l'évaluation dépend de plusieurs éléments : du processus de génération des données d'entrée (simulation plus/moins simplifiées) ; de la nature des données (fantômes/cliniques, avec/sans prétraitement) ; de son déroulement entièrement automatique (intervenant externe, validation séquentielle) ; de la confiance accordée aux vérités terrain d'apprentissage (localisation de la sonde seule ou avec localisation de l'organe) ; de la métrique d'évaluation choisie ; et des fiduciels de validation (nombre d'exemples, nature 2D, manuels/automatiques). Tout obstacle au bon réalisme ou à la faisabilité du déroulement clinique entraîne donc des biais dans l'évaluation. Beaucoup de résultats sont alors à interpréter avec précaution.

## II.5 Notre positionnement

Afin de faciliter la réalisation des biopsies, le principal objectif de cette thèse est d'établir un dispositif d'assistance à la navigation, permettant à l'urologue de repérer continuellement les zones de biopsies qu'il cible. Pour cela, nous nous basons sur des méthodes de recalage rigide 2D(US)/3D(US) en temps-réel, permettant de d'aligner au mieux la coupe échographique courante (de l'imagerie peropératoire) par rapport à un volume de référence (pris en début d'examen). En s'inspirant des études prometteuses de l'état de l'art concernant cette problématique, nous nous sommes appuyés sur des réseaux d'apprentissage profond fortement adaptés au besoin du temps-réel et à la quantité d'informations présentes lors d'un traitement d'images.

La principale nouveauté de notre étude est de considérer des informations liées à la trajectoire réalisée afin de faciliter la tâche de recalage durant la navigation. Nous avons vu dans notre état-de-l'art différents types d'informations d'entrée possibles représentant le concept de trajectoire :

- Des entrées positionnelles (résultats du recalage précédents, suivi de sonde, flot optique)
- Des entrées spatiales (coupe voisine, volume ou sous-volume de référence)
- Des entrées séquentielles (séquences de coupes)
- Des entrées structurelles (masque binaire de segmentation).

Elles présentent toutes l'avantage de contenir de l'information contextuelle spatiale et temporelle combinée, permettant de mieux considérer la trajectoire réalisée. Nous faisons le choix de tester la combinaison de toutes ces entrées, en les intégrant et les évaluant progressivement, jusqu'à obtenir la meilleure configuration possible. La bonne intégration de telles entrées à une architecture profonde nécessitera de porter une attention particulière aux structures utilisées (convolution, dynamique, attention, etc.) ainsi qu'aux types d'optimisation associées (simple, auto-pénalisation, multi-tâches).

Plusieurs évolutions de notre étude ont donc été conduites selon la configuration et la combinaison des types d'entrée précédemment citées. La description de chacune sera allouée à un chapitre dédié dans ce manuscrit (voir Chapitre IV-VI).

Établir des validations adaptées reste un enjeu de taille, notamment dans des études de traitement d'images par apprentissage profond qui nécessitent une très grande quantité de données. Beaucoup d'études présentent des biais au bon réalisme ou à la faisabilité du déroulement clinique résultant alors en de mauvaise évaluation. Afin de répondre aux limites des méthodes de validation existantes, nous proposons, comme seconde contribution de cette thèse, de nous concentrer sur le respect des conditions cliniques de réalisme et de faisabilité de notre application. Cela s'illustrera notamment au travers des données utilisées (simulation d'entrée et de vérités terrain) et des méthodes de validation choisies (évaluation sans intervention extérieure, déroulement réaliste, métriques pertinentes). Cette problématique est justement la thématique du chapitre suivant (voir Chapitre III), qui explicitera la mise en place de notre protocole de validation expérimentale.

**Points clé**

- ❖ Le guidage d'intervention sous imagerie médicale se base essentiellement sur des méthodes de recalage d'images, où les approches par apprentissage profond apparaissent comme une nouvelle opportunité. Cependant, très peu d'études se concentrent sur des problématiques de recalage 2D/3D, pour de l'échographie, et de la prostate.
- ❖ Des méthodes de recalage 3D/3D avec fusion IRM existent pour le repérage des biopsies de prostate. En revanche, seule l'utilisation des images peropératoires (2D US) permet de considérer les déformations anatomiques en temps-réel.
- ❖ Des méthodes de recalage d'images 2D/3D existent pour l'assistance dynamique à la navigation des biopsies prostatiques. Les méthodes les plus prometteuses présentent des fréquences et des précisions de recalage satisfaisantes mais dont les conditions expérimentales (génération de données et évaluation) sont encore limitées et ne représentent pas assez un réalisme clinique.
- ❖ Grâce à l'état-de-l'art des méthodes « 2D/3D » générales, nous avons proposé une classification des différentes approches méthodologiques. Pour ce qui concerne l'ajout d'informations liées à la trajectoire réalisée, nous retenons particulièrement les différents types d'entrée référencées (positionnelles, spatiales, séquentielles et structurelles) ainsi que les structures et optimisations associées.

---

## Mise en place d'un protocole de validation expérimentale

---

*Avant* de décrire les différents choix méthodologiques envisagés pour réaliser ce recalage 2D/3D, ce chapitre s'intéresse à la mise en place d'un protocole de génération des données nécessaires à notre étude et à l'élaboration d'un protocole de validation expérimentale pertinent.

Ces protocoles ont une influence capitale sur la cohérence de la tâche de recalage simulée avec celle réalisée en clinique. Dans un premier temps, nous définirons l'ensemble des données d'entrée et de sortie nécessaires pour un bon apprentissage de la tâche de recalage. Nous présenterons pour cela les éléments dont nous disposons, à partir de bases de données cliniques, et ce qu'il nous reste à simuler. Trois différents procédés de simulation seront explicités, avec leurs avantages et inconvénients. Enfin, nous détaillerons notre protocole de validation des résultats qui respecte un déroulement clinique réaliste et faisable, avec des métriques d'évaluation adaptées et pertinentes.

---

## Sommaire

---

<b>III.1 Génération de données</b>	<b>57</b>
III.1.1 Définitions et notations des données nécessaires à notre étude	57
III.1.2 La base de données clinique disponible	60
III.1.3 Simulation de trajectoire	62
III.1.3.1 Trajectoire de biopsie avec simulateur	62
III.1.3.2 Trajectoire reconstruite par interpolation	68
III.1.3.3 Trajectoire simplifiée de balayage base-apex	71
III.1.3.4 Conclusion sur les simulations	74
III.1.4 Entrées optionnelles : représentation de la trajectoire réalisée	76
III.1.4.1 Données positionnelles	76
III.1.4.2 Données spatiales	77
III.1.4.3 Données structurelles	78
<b>III.2 Protocole de validation</b>	<b>79</b>
III.2.1 Métriques	79
III.2.1.1 Métrique géométrique (TRE)	79
III.2.1.2 Métrique iconique (NCC)	80
III.2.1.3 Métrique statistique ( $R^2$ )	81
III.2.2 Évaluation séquentielle sur une trajectoire complète	82
III.2.2.1 Analyse séquentielle	82
III.2.2.2 Navigation par portion	83

---

## III.1 Génération de données

### III.1.1 Définitions et notations des données nécessaires à notre étude

Notre tâche principale concerne la localisation optimale des images échographiques (collectées durant le guidage d'un examen de biopsie prostatique) par rapport à un volume échographique (acquis en début de navigation). Ce volume est défini comme « volume de référence » ( $V_{ref}$ ), et représente le concept d'espace de destination par rapport auquel on cherche à définir l'alignement des images le plus juste possible. Le choix de ce volume est variable et peut être représenté par différentes acquisitions selon les étapes de navigation. L'objectif est de recalculer à chaque instant  $t$ , l'image 2D US courante  $I_t$  avec  $V_{ref}$ , en estimant la transformation rigide 2D/3D  $T_t$ . Cette problématique est illustrée en Figure III.1. Nous définissons alors :

- Les données d'entrée de notre réseau par la succession d'images échographiques ( $I_1, \dots, I_t$ ), avec l'image courante dénommée  $I_t$ .
- Les données de sortie du réseau (« vérités terrain ») par la localisation optimale de chaque image par rapport au volume de référence, définie à partir de la transformation rigide  $T_t$ .

Ces informations sont les données d'entrée et de sortie principales de notre méthode et seront communes à chacune des études présentées par la suite (Chapitre IV-VI). Les définitions des entrées optionnelles et spécifiques à chaque méthode (positionnelles, spatiales, séquentielles ou encore structurales) seront détaillées plus particulièrement dans la section III.1.4.

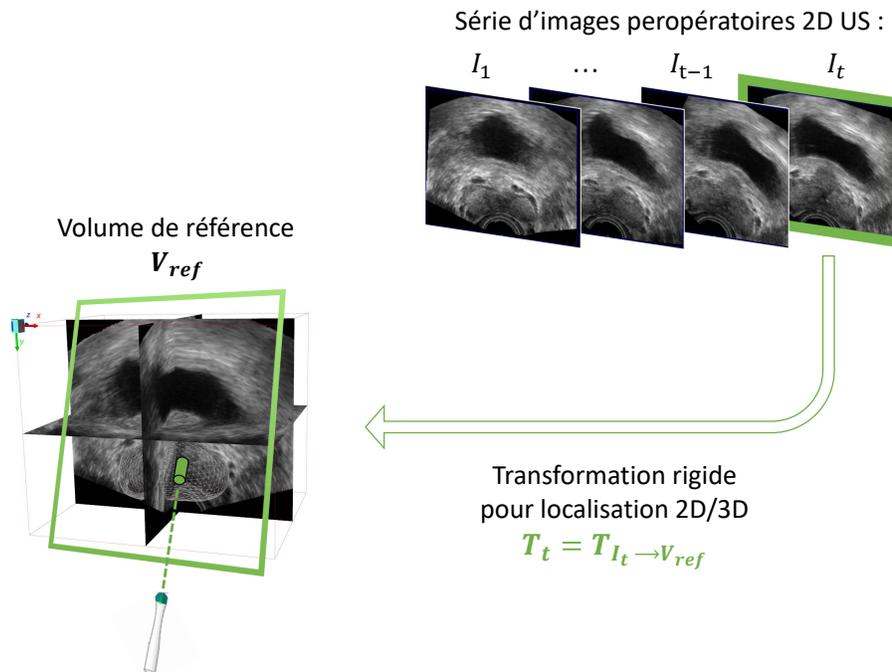


FIGURE III.1 – Définitions et notations des données d'entrée et de sortie.

En complément des dénominations définies en Figure I.7, nous précisons les différents repères utilisés pour décrire nos données dans la suite de ce manuscrit (voir Figure III.2) :

- Le repère échographique  $R_{US}$  (déjà introduit précédemment), prend son origine au bout de la sonde échographique. Dans ce référentiel, l'axe  $y$  correspond à l'axe de rotation de la sonde sur elle-même, le plan  $(x,y)$  représente le plan de la coupe tandis que  $z$  est l'axe d'élévation orthogonale hors-plan. Ce repère est associé à l'image 2D US acquise à chaque instant  $t$ , durant le guidage échographique.
- Le repère image  $R_{I_t}$  caractérise l'image  $I_t$  sous la forme d'une matrice de pixels, où chaque pixel possède une valeur entière d'intensité (entre 0 et 255) représentant le contenu de l'image. La correspondance entre le repère image et le repère de la sonde peut être connue par calibration : à partir des coordonnées physiques de l'origine de l'image, ainsi que la valeur de mise à l'échelle (espacement de pixels en coordonnées physiques). En revanche, la position physique du contenu de l'image n'est pas connue, et est justement celle à déterminer en fonction des bougés et déformations anatomiques.
- Le repère volumique  $R_{V_i}$  caractérise le volume 3D  $V_i$ , obtenu par reconstruction suite à un balayage d'acquisitions 2D de la sonde échographique. Chaque nouvelle acquisition de volume possède une origine de repère placée à différente position dans l'espace physique selon l'orientation de la sonde au moment de l'acquisition.
- On définit en particulier le repère associé au volume de référence  $R_{V_{ref}}$ , correspondant à une acquisition de volume réalisée en début de navigation. C'est notamment par rapport à ce référentiel que l'on cherche à définir la localisation optimale de l'image  $I_t$  associée au repère  $R_{I_t}$ . La transformation vérité terrain  $T_t$  est ainsi définie entre les repères  $R_{I_t}$  et  $R_{V_{ref}}$ .

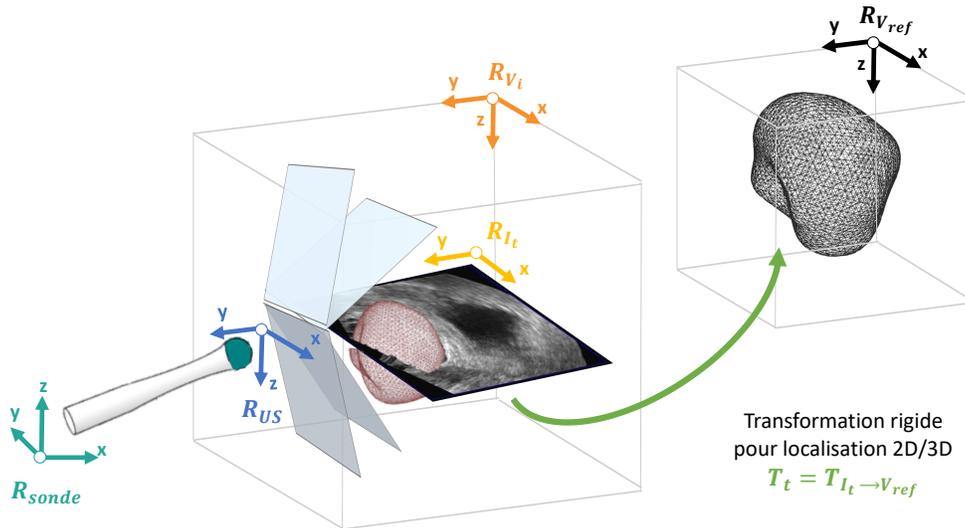


FIGURE III.2 – Visualisation et dénomination des différents repère en jeu lors de la localisation d'une coupe  $I_t$  dans un volume de référence  $V_{ref}$ .

Réussir à avoir accès à une série d'images échographiques 2D durant une procédure clinique tout en connaissant l'alignement optimal par rapport au volume de référence, compte tenu des bougés et/ou déformations, est un processus impossible à réaliser rigoureusement à l'heure actuelle.

Une localisation des images à partir du suivi par capteur ou par localisateur externe (optique, électromagnétique), ne permet de considérer que le bougé de la sonde seule. En revanche, cela ne permet pas de prendre en compte d'éventuels bougés et déformations de la prostate (induites par les bougés de l'organe/patient/praticien), pourtant cruciaux pour localiser correctement des biopsies. De plus, la localisation par un système de suivi optique présente certaines problématiques liées à son installation et son champ de vision dans le bloc opératoire, à la visibilité continue des marqueurs selon la position de l'image acquise, à la mise en place d'une étape de calibration de la sonde avec le localisateur, et aux autorisations réglementaires avant de pouvoir acquérir de nouvelles données.

Une estimation de la localisation la plus probable de chaque image, même par plusieurs praticiens experts, ne semble pas être une solution plus précise ou plus facilement envisageable. La représentation mentale de la position de ces images, pouvant être arbitrairement orientée dans des orientations non-standard et dans un volume présentant une anatomie différente, complexifie d'autant plus cette tâche. De plus, une telle procédure serait bien trop coûteuse pour obtenir une base de données suffisante pour être exploitable en apprentissage supervisé. Il faudrait également évaluer la qualité et la répétabilité de la labellisation, qui peut varier selon l'expertise entre praticiens (inter-variabilité) ou encore entre plusieurs réalisations d'un même expert (intra-variabilité).

Pour s'affranchir de ces problématiques, nous avons opté pour la simulation de nos données d'entrée et de sortie à partir de données cliniques acquises lors d'examens cliniques. Différentes simulations ont été envisagées, chacune présentant leurs avantages et limites. Dans la suite de cette partie, nous présenterons la base de données clinique dont nous disposons et qui a été notre support de travail principal, avant de présenter les différents procédés de simulation avec leur réalisme et complexité associés.

### III.1.2 La base de données clinique disponible

La plateforme d'assistance Trinity (Koelis<sup>®</sup>), commercialisée à partir de 2016, permet la visualisation de la cartographie des carottes dans un volume de référence durant le guidage échographique de biopsies prostatiques (voir Figure III.3). Grâce à son utilisation courante durant de réelles procédures de routine de biopsies, notamment au CHU de Grenoble, nous avons pu collecter un grand nombre de données cliniques (déclaration MR2711140520 CNIL).



FIGURE III.3 – Dispositif d'assistance à la navigation avec fusion 3D US-IRM : Trinity<sup>®</sup> (Koelis, France). La plateforme intègre un échographe associé à une interface de visualisation de cartographie des carottes. La plateforme a été commercialisée et approuvée FDA en 2016 et est depuis couramment utilisée par les urologues pour des procédures de biopsie. Toutes les images sont traitées en B-mode.

Les données associées à chaque procédure regroupent les éléments décrits ci-dessous et illustrés en Figure III.4 :

- Le volume IRM ( $V_{IRM}$ ), lorsque celui-ci est réalisé en préopératoire. Des zones suspectes peuvent potentiellement être repérées sur cette modalité.
- Le volume échographique « panorama » ( $V_{pano}$ ), pris en début de procédure. Il provient en réalité de la fusion de trois acquisitions, réalisées à partir de trois positions de sonde orientées différemment. Avec une meilleure qualité d'image et un plus grand champ de vision, ce volume sert de référence anatomique pour les cartographies de biopsies. En effet, tous les autres volumes seront recalés sur celui-ci, permettant ainsi de tous les décrire dans le même espace anatomique. Les zones suspectes de la modalité IRM peuvent également être transposées sur cette modalité grâce à un algorithme de fusion élastique 3D(US)/3D(IRM) intégré [17].
- Le maillage de la prostate, estimé dans le panorama, par l'urologue expert à l'aide d'un algorithme de segmentation semi-supervisé. On le note par la suite  $M_{pano}$ .
- Les volumes échographiques, dits de « biopsies » ( $V_i$ ), acquis à chaque nouvelle biopsie  $i = (1, j, \dots, N)$  au cours de l'examen. Chaque acquisition dure environ 3 secondes durant lesquelles l'opérateur et le patient doivent essayer de rester immobiles.

- Le recalage  $T_{V_i \rightarrow V_{pano}}$ , entre chaque volume de biopsie et le panorama. Chaque transformation est obtenue à partir d'un algorithme de recalage iconique 3D/3D rigide intégré dans la plateforme [17]. Ces transformations sont toutes décrites par rapport au repère  $R_{V_{pano}}$  et illustrées sur la Figure III.4. Ce recalage, combiné à la fusion multimodale préopératoire, permet de contrôler la localisation des biopsies réalisées ou en cours de réalisation (positionnement de l'aiguille « virtuelle »). Le temps de recalage a pu être réduit à 1 seconde, grâce notamment à l'optimisation sur GPU.

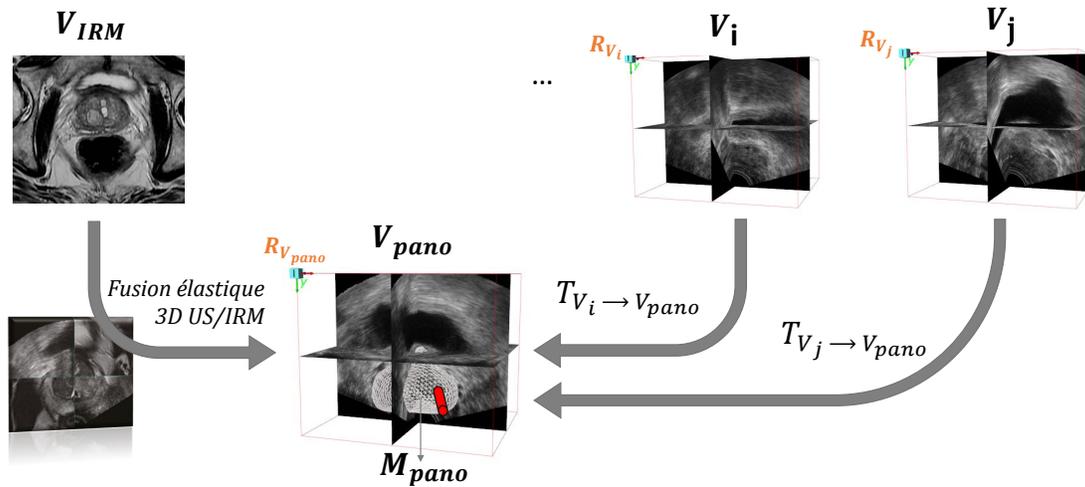


FIGURE III.4 – Localisation des volumes à partir des informations collectées durant des procédures cliniques de biopsies.

Nous recensons, à ce jour, des données de plus de 300 procédures cliniques, collectées depuis 2017, sur 277 patients différents. Chaque patient dispose des éléments décrits précédemment, avec entre 5-15 acquisitions de volumes de biopsies et leur transformation correspondante vers le panorama. Ces volumes sont généralement d'une dimension de  $304 \times 256 \times 256$  voxels avec une résolution isotropique de 0.306 mm.

Cette base de données cliniques permet de collecter des données réelles, obtenues durant de vraies procédures cliniques et sur une grande diversité de patients. Ceci nous offre un support de travail conséquent et pertinent par rapport aux enjeux de la génération de données pour un apprentissage profond supervisé. En revanche, le flux des images peropératoires n'est pas accessible et leurs localisations optimales par rapport au volume de référence restent inconnues. La partie suivante décrit les processus de simulation que nous avons élaboré, à partir des données dont nous disposons, afin de répondre à ce manque.

### III.1.3 Simulation de trajectoire

À défaut d’avoir des images (2D US) directement issues des procédures cliniques, nous avons opté pour un processus de simulation pour extraire artificiellement des coupes en les rééchantillonnant à partir d’informations cliniques disponibles. Différentes simulations sont présentées dans cette partie : trajectoire de biopsie via un simulateur, reconstruction de trajectoire par interpolation, ou trajectoire par balayage base-apex. Différents degrés de réalisme et complexité y sont intégrés. Le choix de la simulation de la trajectoire la plus adaptée est d’autant plus important dans le cadre de notre étude, où l’on cherche à exploiter les informations liées à la trajectoire réalisée pour faciliter la tâche de recalage durant une procédure. Nous présenterons chaque méthode de génération, avec leurs avantages et limites, pour mieux comprendre l’évolution de notre simulation de données.

#### III.1.3.1 Trajectoire de biopsie avec simulateur

Dans un premier temps, nous avons utilisé un simulateur de biopsie (« Biopsym »), précédemment développé par notre équipe [19, 93]. Celui-ci permet de reproduire des procédures de biopsies avec un retour visuel sous la forme d’image 2D informant sur la position de la biopsie virtuellement réalisée au sein de la prostate. Ce simulateur se compose d’une sonde échographique factice, réalisée en impression 3D sur la base d’une sonde échographique de la plateforme Trinity, et rattachée à un dispositif haptique à retour de force (Phantom Omni, Sensable Devices Inc.). Cette sonde est placée au centre d’un anneau rigide dans lequel un insert en silicone a été ajouté pour mimer l’anus du patient. Cette dernière version du simulateur, illustrée sur la Figure III.5-A, a été pensée pour représenter au mieux les conditions cliniques, avec une sonde ultrason factice, des déplacements contraints et des frottements réalistes. Tous les angles sont mesurés dans le référentiel du simulateur, représenté par le référentiel de la sonde ( $R_{US}$ ).

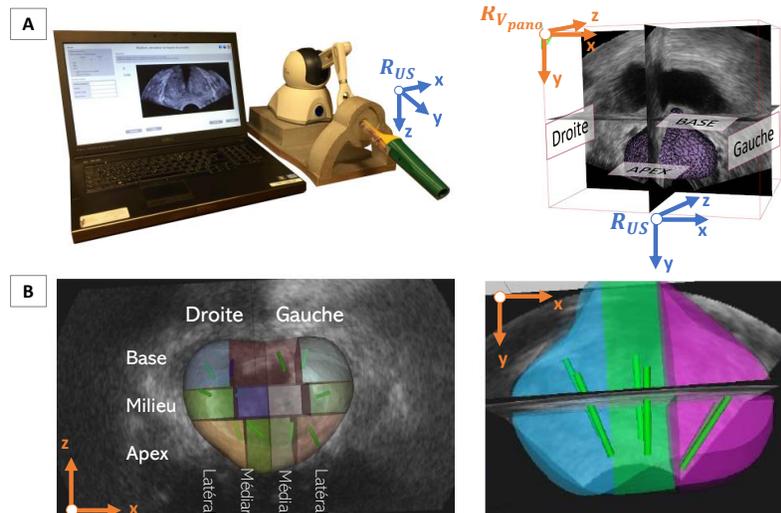


FIGURE III.5 – A) Biopsym, le simulateur de biopsie. B) Interface de visualisation des biopsies virtuelles réalisées (vue sagittale à gauche, vue coronale à droite). Évaluation des performances obtenues grâce au score d’intersection des biopsies avec les 12 secteurs de prostate respectant le schéma de biopsie standard.

Ce simulateur est utilisé essentiellement dans un but pédagogique pour favoriser l'apprentissage du geste aux étudiants praticiens [94, 95]. En effet, celui-ci enregistre les positions et orientations à l'extrémité de la sonde, à une fréquence d'environ 100 Hz, lorsque celle-ci est déplacée par l'utilisateur. À partir de certaines données cliniques de la base de données présentée précédemment ( $V_{pano}$  et  $M_{pano}$  uniquement), il affiche l'image 2D extraite du volume clinique, correspondant au plan de coupe échographique obtenue à la position de la sonde. Le calcul et l'affichage sont réalisés en temps-réel et permettent de simuler un guidage échographique continu en fonction des mouvements enregistrés par le dispositif haptique. Par la suite, une carotte virtuelle peut également être réalisée, à partir d'une touche clavier, dont le déclenchement du tir de biopsie est simulé par un mouvement de recul du bras. Une visualisation de la position de la biopsie virtuelle est fournie à l'utilisateur (Figure III.5-B), lui permettant ainsi d'évaluer sa performance compte tenu des 12 secteurs à viser s'il s'agit d'une biopsie standard ou de la position d'une cible IRM s'il s'agit d'une biopsie ciblée.

Grâce au principe de l'algorithme du simulateur, décrit dans le paragraphe précédent et illustré sur la Figure III.6, il est alors possible d'obtenir une suite d'images 2D ( $I_1, \dots, I_t$ ), avec une connaissance précise de leur localisation au sein d'un volume de référence, selon la trajectoire réalisée. Les vérités terrain  $T_t$  sont définies par la mesure de l'orientation de la sonde du simulateur et le volume de référence est défini par le panorama ( $V_{ref} = V_{pano}$ ). Au cours de ce processus, le maillage de la prostate ( $M_{pano}$ ), associé au volume panorama, est également transformé selon les transformations rigides successives associées à la trajectoire de la sonde. Les maillages sont sauvegardés à chaque instant et notés  $M_t$ . Ils seront notamment utilisés au cours du processus d'évaluation ultérieur.

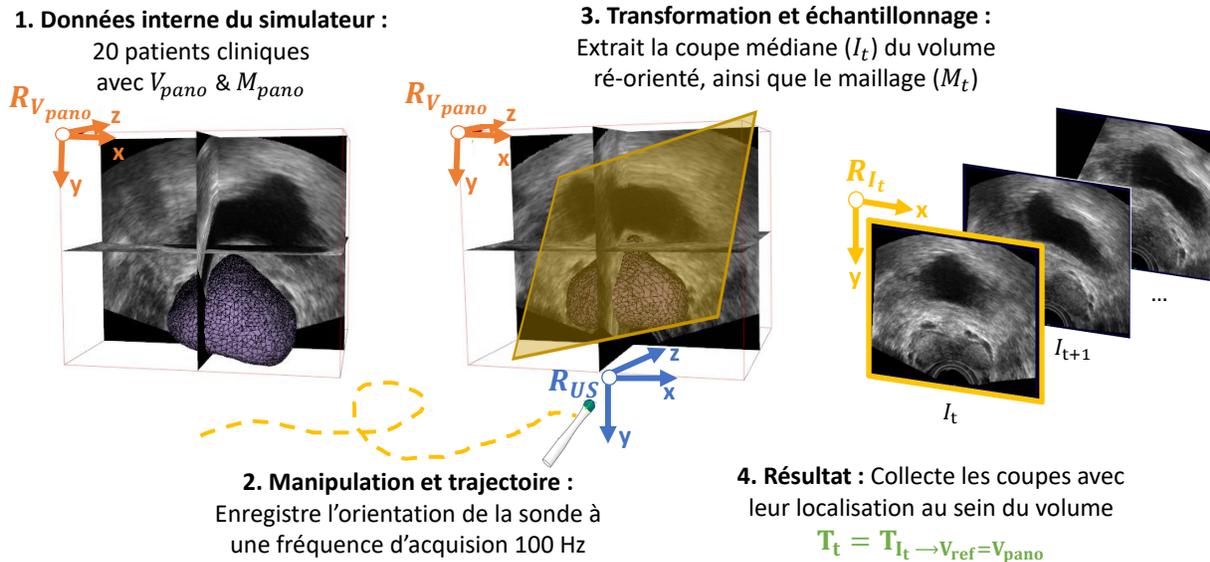


FIGURE III.6 – Principe de reconstruction d'une image 2D, correspondant à la position de la sonde factice, par le simulateur.

Nous avons collecté plusieurs trajectoires de biopsies standardisées, réalisées par plusieurs pra-

ticiens sur le simulateur. La procédure de biopsies standardisées, est supposée respecter un certain ordre de passage entre les différentes zones de la glande à balayer (notamment au CHU de Grenoble) : successivement la base (médian puis latéral), milieu (médian puis latéral) et apex (médian puis latéral). Ce schéma, très stéréotypé et régulier, est répété sur les deux moitiés de la glande (droite et gauche, ou inversement) après une rotation à  $180^\circ$  de la sonde sur elle-même pour atteindre l'autre côté avec l'aiguille. Ces 12 secteurs de biopsies standardisées sont rappelés dans la Figure III.5.

Un exemple de deux trajectoires, réalisées par deux praticiens de différents niveaux d'expertise, est illustré sur la Figure III.7. Sur ces courbes, on reconnaît notamment les motifs consécutifs et répétitifs décrits précédemment, avec la présence plus ou moins importante de bruit supplémentaire selon la dextérité et l'efficacité de l'opérateur. Les trajectoires obtenues sont très stéréotypées avec des amplitudes de mouvements cohérentes entre tous les utilisateurs :  $\pm 20^\circ$  autour de l'axe z pour le mouvement médian vers latéral, une rotation  $180^\circ$  autour de l'axe y pour la rotation de la sonde d'un côté à l'autre de la glande, et jusqu'à  $30^\circ$  autour de x pour le balayage de la base à l'apex. Une description de la distribution moyenne des trois rotations, selon l'expertise des praticiens, est également fournie sur l'histogramme de la Figure III.8, où les classes de mouvements sont plus ou moins distinctes selon la précision du praticien (voir Figure I.4).

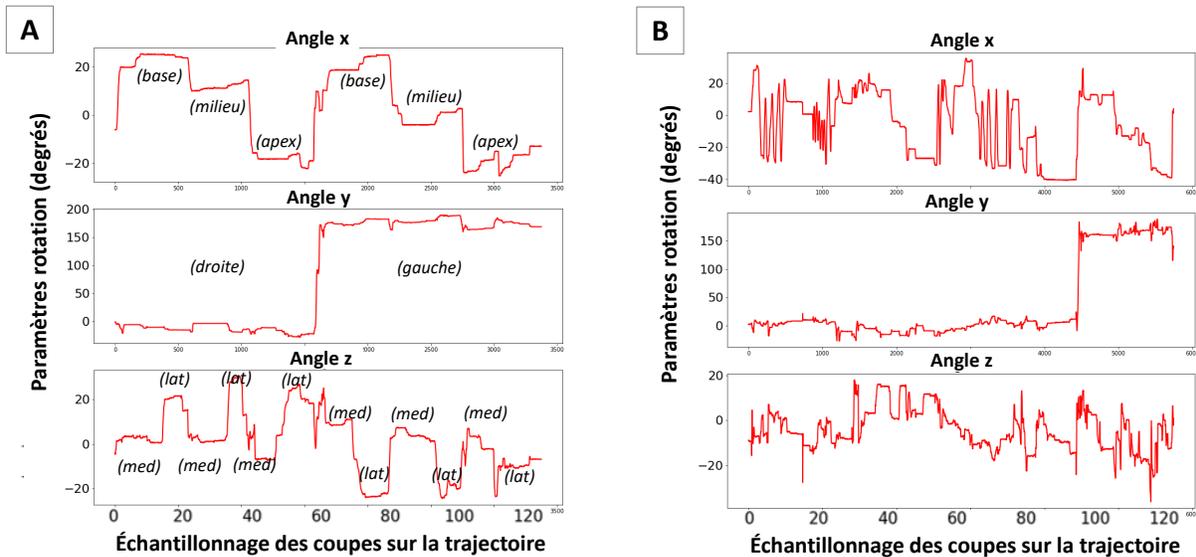


FIGURE III.7 – Visualisation des trajectoires de type biopsie standardisée pour chaque paramètre de rotation, et selon différent niveau d'expertise : expert à gauche (A), débutant à droite (B).

Les coupes, dont on cherche à définir l'alignement optimal du contenu iconique par rapport au volume de référence, sont issues de ces trajectoires. Il est important de savoir quantifier l'erreur d'alignement initial (avant recalage) entre ces deux images. Cela permet à la fois de vérifier l'amélioration de l'alignement estimé par l'algorithme de recalage et de quantifier la difficulté des tâches en jeu (étroitement liée à l'initialisation du recalage). Nous allons utiliser la métrique de recalage TRE (décrite en II.1.1.1), pour mesurer l'erreur avant recalage tout au long des trajectoires simulées. Pour cela, la TRE sera calculée entre les maillages  $M_t$  et  $M_{pano}$  pour mesurer la correspondance

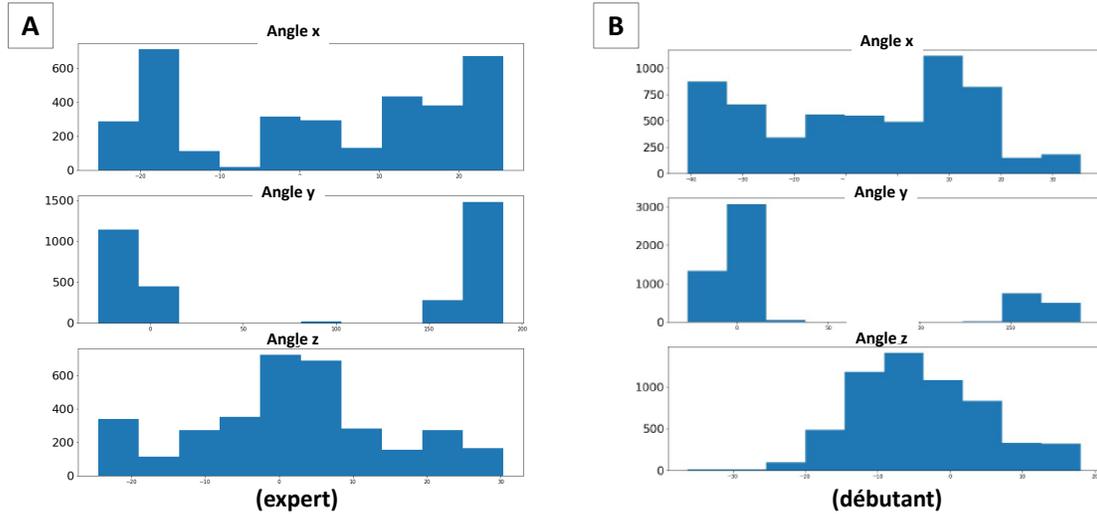


FIGURE III.8 – Histogramme de la distribution des paramètres de rotation pour une trajectoire de type biopsie, réalisée selon différents niveaux d'expertise : expert à gauche (A), débutant à droite (B).

géométrique entre les points surfaciques du maillage de la prostate. La description de ce calcul, également utilisé pour la validation de nos résultats, sera plus détaillée en section III.2.2.

L'évolution du profil de l'erreur de recalage en fonction d'une trajectoire appliquée est illustrée sur la Figure III.9. Certains motifs, consécutifs et répétitifs (de type créneaux), sont cohérents avec le profil de la trajectoire stéréotypée appliquée. On remarque que l'erreur varie significativement à partir de la moitié de la trajectoire (de 10 mm à 40 mm), correspondant à la rotation de la sonde sur elle-même pour atteindre l'autre hémiprostate avec l'aiguille. L'erreur moyenne avant recalage, mesurée sur la totalité des procédures simulées, est d'environ  $30.41 \pm 19.05$  mm. Cette erreur, largement impactée par la deuxième moitié de l'examen, est conséquente par rapport à la taille totale d'une glande prostatique d'environ 40 mm de diamètre. Cela reflète une tâche de recalage complexe et difficile, où l'anatomie peut parfois être complètement inversée par rapport à l'anatomie du volume de référence.

Cette utilisation du simulateur présente de nombreux avantages et permet de simuler une trajectoire au plus proche d'une utilisation en clinique :

- Tout d'abord, les praticiens peuvent réaliser des mouvements de trajectoire de biopsies réalistes où les caractéristiques cliniques sont conservées : mêmes amplitudes de mouvements, vitesse du mouvement adaptée, fréquence d'acquisition identique. Les trajectoires obtenues sont complexes et représentatives d'une procédure de biopsie standardisée.
- Les simulations sont patient-spécifiques, i.e. les praticiens adaptent leurs mouvements en fonction du patient et notamment de la forme et du volume de sa prostate.
- Il est possible de comparer des trajectoires issues de plusieurs praticiens, avec notamment des niveaux d'expertises différentes (trajectoire plus longue, mouvement hésitant, etc.). Ceci permettrait par la suite d'évaluer la bonne généralisation de notre modèle à tous les niveaux,

ou au contraire d'envisager des modèles spécifiques pour chaque niveau d'expertise.



FIGURE III.9 – Évolution du profil de la TRE au cours d'une trajectoire de type biopsie standardisée.

Nous retenons cependant quatre limites majeures à cette simulation de trajectoires de biopsies :

- Les coupes échographiques de suivi (qui correspondent au guidage) sont extraites directement d'un seul et même volume. Elles présentent donc une absence de déformation anatomique ainsi qu'une incohérence de speckle. Ce volume sert également d'espace de référence dans lequel on cherche la localisation optimale des coupes. Les images présentent donc la même anatomie, rendant la tâche de localisation plus simple que la réalité.
- Les techniques de reconstruction et d'interpolation des coupes ainsi extraites ajoutent également du biais dans la qualité des images. Exceptée l'image issue d'un plan identique au plan qui a servi à réaliser l'acquisition du volume, l'image 2D reconstruite à une orientation quelconque, présentera une qualité inférieure de celle obtenue durant l'échographie.
- Contrairement à des images échographiques réelles, présentant un « cône » échographique constant et fixe sur l'image, nos images simulées présentent des cônes échographiques mobiles et variables (voir Figure III.9). Durant la trajectoire simulée, le plan de la coupe à rééchantillonner peut intersecter le volume de référence hors de son contenu iconique prédéfini. L'image ainsi obtenue est alors incomplète et contient des « zones noires » (sans information iconique), que l'on peut qualifier de cône échographique variable. La Figure III.9 illustre plusieurs exemples de coupes issues de l'algorithme de rééchantillonnage du simulateur, avec la présence plus ou moins importante de zones noires selon les orientations des coupes extraites du volume de référence. Ce problème, lié aux variations de zones noires lors du rééchantillonnage, occasionne un biais de réalisme des images échographiques dont il faudra étudier l'impact lors de l'apprentissage (voir Chapitre IV).
- Seules les rotations de la sonde sont enregistrées par le simulateur. Sans translation, nous ne pouvons considérer que des transformations rigides incomplètes, se limitant aux trois paramètres de rotation. D'autres versions du simulateur ont cherché à simuler des déformations afin de considérer les translations [19,93], en revanche elles résultent en des zones noires encore plus importantes.

En conclusion, le simulateur nous permet de reproduire la génération de données la plus utilisée de notre état-de-l'art : des coupes rééchantillonnées de volumes cliniques fixes à partir de trajectoires

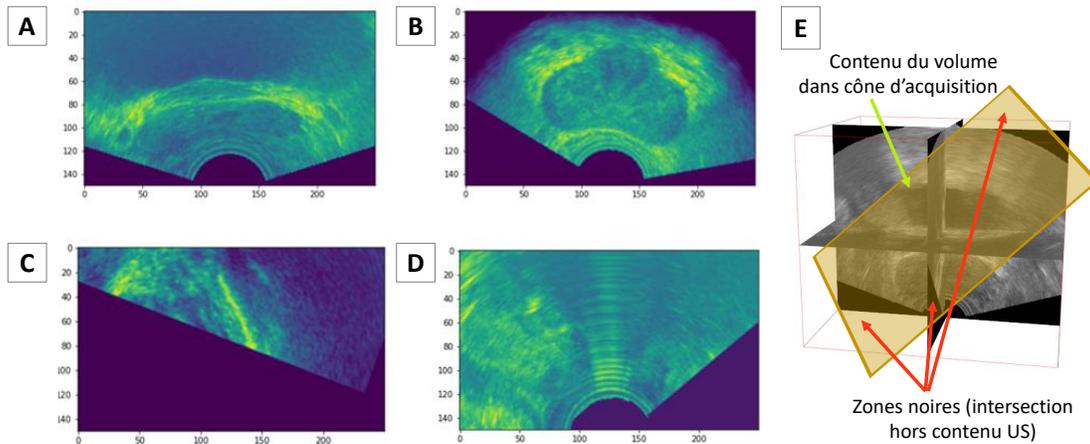


FIGURE III.10 – Exemples d’orientation de cône échographique variable avec « zones noires » selon la coupe extraite.

(voir section II.4.5.1). La localisation de chaque image (vérité terrain  $T_t$ ) est définie par l’orientation de la sonde factice du simulateur, qui correspond à la transformation appliquée sur le volume de référence. Le concept d’espace de référence est défini, dans ce cas-ci, par le volume panorama ( $V_{pano}$ ) qui est le seul volume disponible dans les données du simulateur.

L’intérêt de l’utilisation de ce simulateur, contrairement à un algorithme de rééchantillonnage avec des trajectoires simulées aléatoirement, est la conservation du réalisme et de la complexité du geste de biopsie. En effet, la manipulation est réalisée par des experts, elle est adaptée à une anatomie spécifique du patient, tout en conservant des caractéristiques de mouvements (vitesse, amplitude, etc.) représentatives de biopsies cliniques standardisées. L’application d’amplitude de rotations jusqu’à  $180^\circ$ , où l’anatomie se retrouve complètement inversée, reflète notamment une tâche complexe de localisation. Cette simulation, utilisée à la fois pour la génération des données et pour la procédure de validation, permet une évaluation plus réaliste et plus complexe que la plupart des études référencées dans la littérature.

L’avantage lié à la complexité des trajectoires simulées est également la limite majeure de cette simulation : des zones noires diffèrent selon l’orientation de la coupe rééchantillonnée et ne sont pas représentatives de la réalité. Ceci peut alors biaiser l’apprentissage de notre modèle. Cette problématique est commune à toutes les études de notre état-de-l’art présentant des coupes arbitrairement orientées dans un volume. Cependant, très peu d’études mentionnent ce biais ou évaluent son impact. Pour finir, les erreurs initiales (avant recalage) simulées sont conséquentes (jusqu’à 30 mm), et bien supérieures à celles retrouvées dans la littérature (autour de 9 mm, voir II.3).

La prochaine section détaille notre nouvelle proposition de simulation visant à améliorer le réalisme de la tâche de recalage et s’affranchir des limites de cette première approche.

### III.1.3.2 Trajectoire reconstruite par interpolation

Notre deuxième processus de simulation a cherché à répondre aux limites majeures relevées dans la simulation de biopsies standardisées précédente. L'objectif principal est de réussir à générer une suite d'images 2D, localisées par rapport au volume de référence, tout en s'affranchissant des problématiques (i) d'absence de déformation et (ii) d'absence des paramètres de translation. Pour cela, nous ne cherchons plus à imiter de nouvelles procédures en simulant de nouvelles trajectoires sur de nouvelles données, mais au contraire, nous cherchons à reproduire la procédure clinique réalisée à partir des données enregistrées et disponibles.

Les données collectées pour chaque procédure clinique sont enregistrées ponctuellement durant l'examen : notamment au début de la procédure (acquisition de  $V_{pano}$ ,  $M_{pano}$ ) et à chaque nouveau site de biopsie  $i = (1, j, \dots, N)$  avec l'acquisition de  $V_i$  et le calcul du recalage  $T_{V_i \rightarrow V_{pano}}$ , intégré dans la plateforme [17]. Les 6 paramètres du recalage obtenus sont calculés par rapport au repère volumique  $R_{V_{pano}}$ .

Il est alors possible d'interpoler ces 6 paramètres de transformation enregistrés entre deux estimations de recalage successives et de reconstruire l'évolution de ceux-ci au cours de la procédure complète. L'interpolation de chacun des paramètres, permet d'approximer la trajectoire continue ayant été réalisée durant cette procédure clinique.

Des exemples d'interpolation quadratique de paramètres sont illustrés sur la Figure III.11. Par rapport aux trajectoires précédentes, le schéma stéréotypé se distingue toujours mais est visuellement moins marqué. Cette simulation permet d'obtenir tous les paramètres de transformation rigide (rotation et translation), en s'assurant d'un certain réalisme de la trajectoire ainsi reconstituée. Ceux-ci définiront alors nos valeurs de localisation de vérités terrain  $T_t$ .

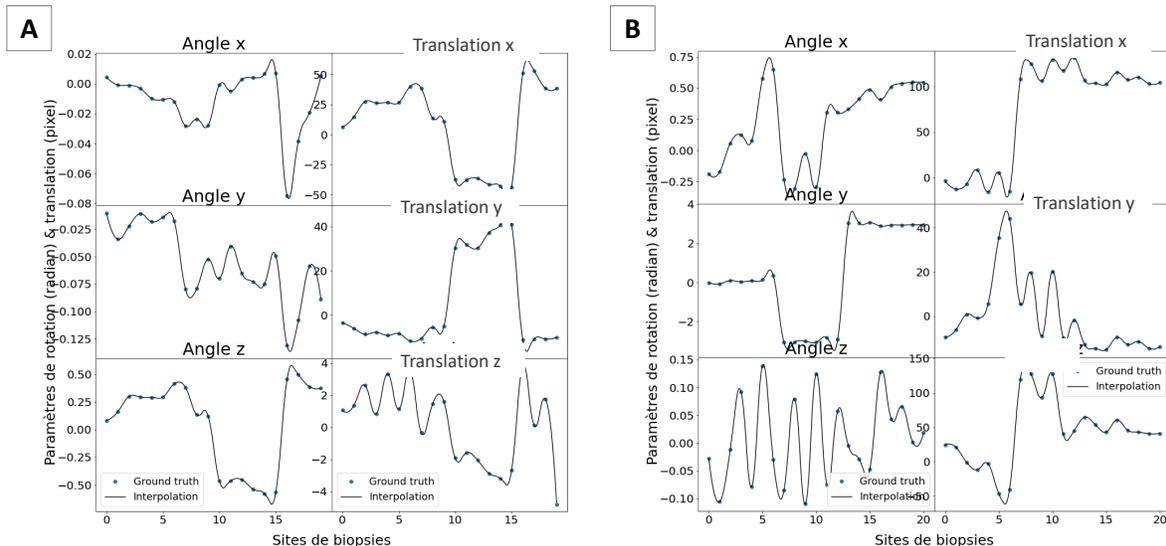


FIGURE III.11 – Visualisation des trajectoires reconstruites par interpolation quadratique des paramètres de transformation enregistrés entre les sites de biopsies successifs.

L'idée est d'extraire des coupes issues des différents volumes échographiques  $V_{i=(1,\dots,N)}$ , selon chaque portion de trajectoire reconstruite entre deux sites de biopsies successifs. En effet, à partir du recalage mesuré pour un site de biopsie  $i$  ( $T_{V_i \rightarrow V_{pano}}$ ) ainsi que le recalage imposé par l'interpolation calculée pour un instant  $t$  ( $T_t$ ), on peut alors déterminer le déplacement relatif ( $T_{V_i \rightarrow V_{pano}}^{-1} \cdot T_t$ ) nécessaire pour rééchantillonner  $V_i$ . L'image ainsi extraite représente  $I_t$  dont la localisation par rapport à  $V_{pano}$  correspond bien aux paramètres de transformation  $T_t$  (voir Figure III.12). Cette génération permet d'extraire des coupes issues de plusieurs volumes échographiques, présentant tous des déformations anatomiques différentes et indépendantes du volume de référence par rapport auquel on cherche à les localiser ( $V_{ref} = V_{pano}$ ).

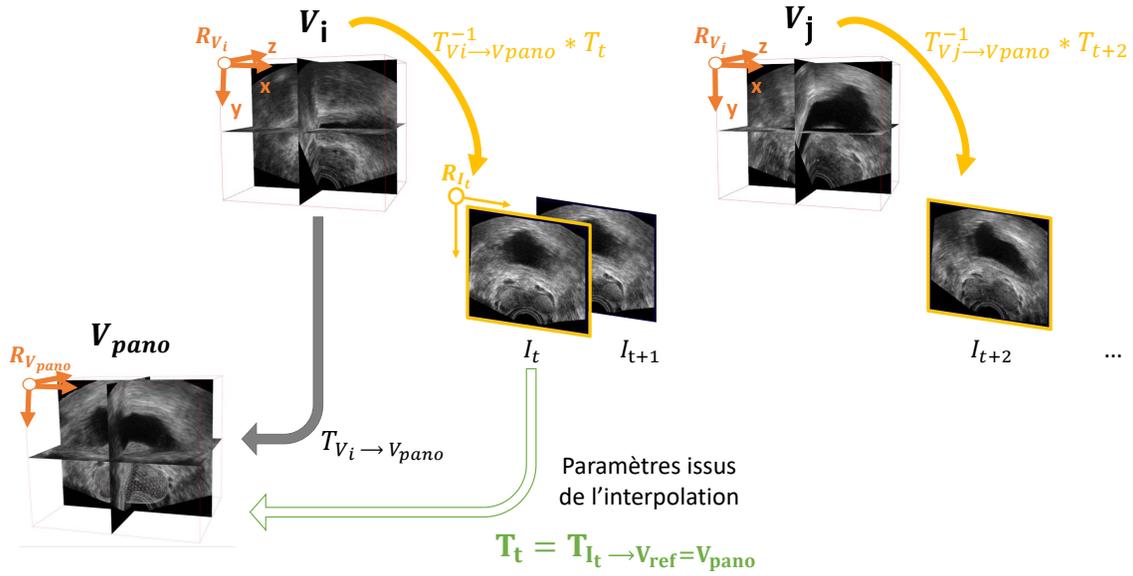


FIGURE III.12 – Second processus de génération de données : extraction d'une série d'images 2D, issues de plusieurs volumes successifs, échantillonnées à partir d'une trajectoire imposée (étape d'interpolation préalable).

De la même façon que pour la simulation précédente, nous avons appliqué ces mêmes trajectoires pour transformer le maillage de la prostate ( $M_{pano}$ ) et les enregistrer à chaque instant (notés  $M_t$ ). La Figure III.13 illustre l'évolution de l'erreur initiale (avant recalage) au cours de différents exemples de trajectoires reconstruites par interpolation. Cette erreur de recalage est toujours très importante, illustrant de façon réaliste l'ampleur des mouvements d'une sonde échographique d'une biopsie à la suivante. De plus, le volume panorama (comme volume de référence) présente des déformations de plus en plus différentes des images rééchantillonnées (d'autres volumes) au cours de la trajectoire. L'association de trajectoires complexes et réalistes avec différentes déformations anatomiques rend la tâche de recalage simulée encore plus complexe.

A première vue, ce nouveau processus de simulation permet de résoudre certaines limites de la simulation précédente. Cependant, des problèmes demeurent :

- Sans compter les imprécisions liées à l'interpolation de la trajectoire, celle-ci ne correspond en

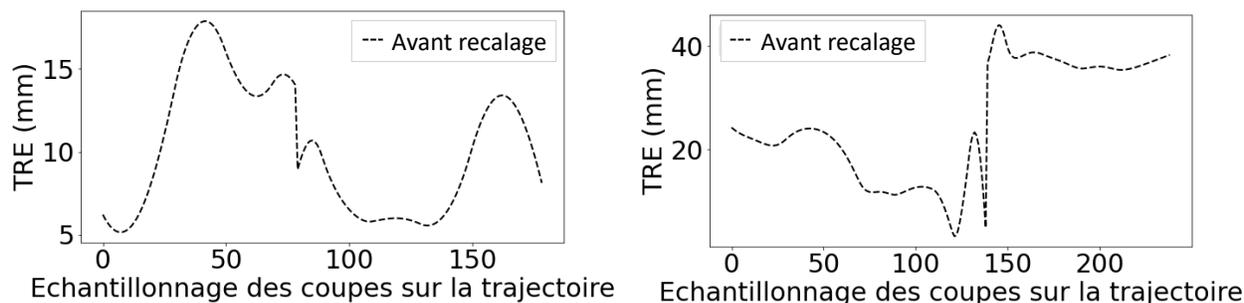


FIGURE III.13 – Évolution du profil de la TRE au cours d'une trajectoire reconstruite par interpolation.

réalité pas aux déplacements réalisés par le praticien durant l'examen. En effet, les paramètres issus du recalage rigide traduisent les déplacements de la sonde mais également ceux de la prostate. La trajectoire, ainsi « reconstruite » à partir de l'interpolation des paramètres, correspond en réalité aux bougés de la prostate, combinés d'une manière indissociable à ceux de la sonde. Il n'est donc pas possible de reconstituer une trajectoire réaliste de la sonde seule, comme réalisée durant la procédure. De plus, l'interpolation indépendante, sur chacun des paramètres, ne respecte pas leurs réelles corrélations. Enfin, selon la fréquence d'interpolation choisie, la vitesse des mouvements ne respecte plus parfaitement celle des gestes initialement produits.

- Le rééchantillonnage au sein de plusieurs volumes, à partir d'une portion de trajectoire limitée, ne permet pas de réduire l'apparition des zones noires. Avec des paramètres de translation supplémentaires, celles-ci sont au contraire encore plus importantes.

En reconstruisant une trajectoire proche de celle réalisée durant la procédure, en disposant des données enregistrées ponctuellement, l'objectif est de régénérer les images s'approchant du guidage échographique tout en maîtrisant leur localisation. L'utilisation de plusieurs volumes successifs, pour générer des images d'anatomies différentes du volume de référence, est une idée prometteuse qui permet de reproduire la notion de déformations anatomiques en temps-réel. Ceci permet de se rapprocher du réalisme et de la complexité de la tâche de recalage à effectuer. Cependant, utiliser le panorama comme volume de référence, alors que celui-ci présente une anatomie de plus en plus différente au cours de la trajectoire, complexifie d'autant plus la tâche de localisation. Il faut ainsi songer à utiliser un volume de référence bien différent du volume utilisé lors du rééchantillonnage, mais dont l'anatomie reste suffisamment proche pour faciliter la tâche de localisation. Enfin, l'utilisation d'une trajectoire complexe et continue, bien que représentative de celle effectuée en clinique, ne permet pas le rééchantillonnage d'images sans présence de zones noires. C'est pourquoi dans notre troisième approche de génération de trajectoires, nous avons choisi de simplifier le mouvement, le plus possible, pour s'affranchir de ce biais potentiel toujours présent dans nos données.

### III.1.3.3 Trajectoire simplifiée de balayage base-apex

Ce dernier processus de génération vise à la fois à supprimer la présence de zones noires (selon l'orientation du rééchantillonnage) et à diminuer la complexité de la tâche simulée, pour être comparable aux autres études de l'état de l'art. Pour cela, les trajectoires mises en jeu dans cette partie, devront respecter ces conditions aux dépens du réalisme et de la complexité de trajectoires d'examen.

Nous proposons de balayer un volume selon l'axe de motorisation de la barette échographique pour garantir un même cône échographique pour toutes les coupes et pour garantir des erreurs initiales moins grandes. Nous avons généré des trajectoires de balayage ( $T_{Sweep,t}$ ), correspondant à plusieurs aller-retours entre les extrémités supérieures (base) à inférieures (apex) de la prostate. Ces trajectoires sont obtenues via une rotation autour de la tête de la sonde ( $\pm 40^\circ$  autour de l'axe des x) et permettent d'extraire des coupes successives ( $I_1, \dots, I_t$ ) issues de ce balayage dans un volume échographique. Ces coupes sont alors contraintes au même type de balayage que celui qui a participé à la reconstruction de ce volume. Les zones noires, associées au cône échographique, sont fixes et identiques sur chacune des images ainsi rééchantillonnées.

Afin de considérer des déformations anatomiques différentes, mais suffisamment proches, nous avons envisagé plusieurs niveaux de complexité de notre tâche de recalage. Alors que les coupes ( $I_t$ ) sont toujours mentionnées comme étant extraites de  $V_i$ , le choix de l'espace de référence dépendra du niveau de complexité que l'on veut simuler : soit celui-ci correspond au même espace ayant servi au rééchantillonnage des coupes ( $V_{ref} = V_i$ ), soit il s'agira de localiser ces coupes dans un espace différent ( $V_{ref} = V_{j \neq i}$ ). Ces deux processus de génération de données sont désignés respectivement par «  $I_{i,t}$  à  $V_i$  » ou «  $I_{i,t}$  à  $V_j$  » et sont illustrés dans la Figure III.14.

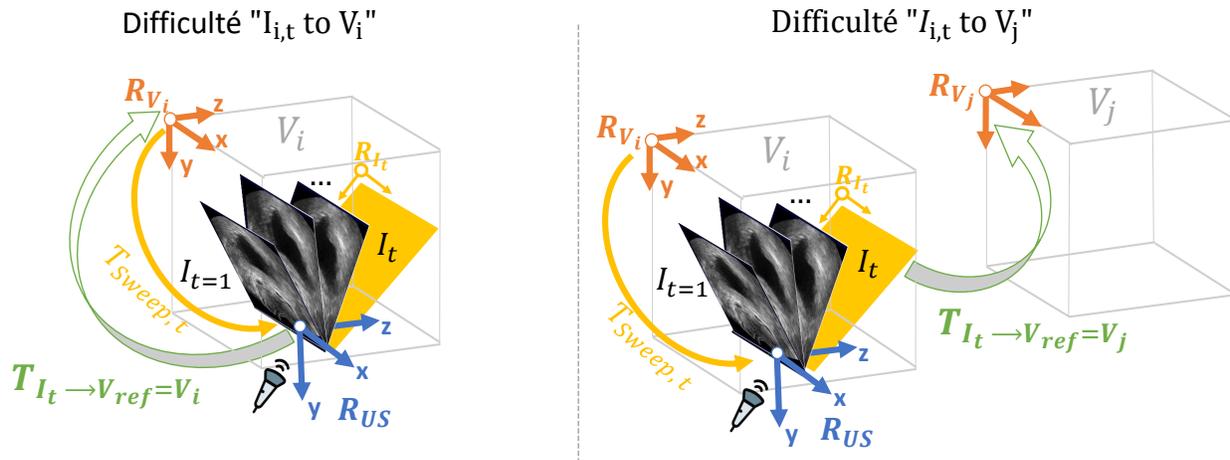


FIGURE III.14 – Troisième processus de génération de données. Le mouvement de balayage correspond à plusieurs allers-retours entre les extrémités supérieures (base) à inférieures (apex) de la prostate. Les coupes  $I_t$  sont rééchantillonnées d'un volume  $V_i$ . Deux niveaux de difficultés sont simulés en fonction de l'espace de localisation choisi : soit identique ( $V_i$ ) soit différent ( $V_j$ ).

Alors que le premier jeu de données présente des dépendances et des biais d'image qui ne représentent pas toute la complexité des cas réels (contenu iconique parfaitement identique, incohérence de speckle), le second comprend des coupes d'entrée qui sont maintenant indépendantes et anatomiquement différentes du volume de référence par rapport auquel nous voulons les localiser. Contrairement au second processus de simulation (III.1.3.2) où le panorama est utilisé tout au long de la procédure comme volume de référence, ici, le choix de  $V_j$  est évolutif et dépend des volumes voisins. En effet, utiliser le volume acquis au site de biopsie précédent  $V_{i-1}$ , s'illustre comme une complexité intermédiaire où la tâche de localisation se limite à une anatomie similaire. Les vérités terrain de localisation peuvent alors être calculées via la composition des différentes transformations à disposition. Le calcul est décrit dans l'équation III.1 selon le niveau de complexité simulé.

Les Figures III.15-A et C illustrent toutes deux l'allure de trajectoires ainsi obtenues. Les mouvements du balayage base-apex (rotations successives autour de l'axe x) sont bien illustrés dans les graphes, tandis que les autres rotations sont soit nulles (pour «  $I_{i,t}$  à  $V_i$  ») soit non nulles du fait de la composition avec la transformation  $T_{V_i \rightarrow V_{ref}}$  (pour «  $I_{i,t}$  à  $V_j$  » selon l'équation III.1.3.3). Les paramètres de translation sont considérés mais ne représentent en réalité que le changement de repère entre les mouvements réalisés (exprimés depuis le centre de sonde  $R_{US}$  associé au volume rééchantillonné  $V_i$ ) et le repère du volume de référence ( $R_{V_i}$  ou  $R_{V_j}$ ). Ces paramètres ne traduisent donc pas explicitement les translations de la prostate par rapport à l'anatomie de référence.

$$T_t = T_{I_t \rightarrow V_{ref}} = T_{V_i \rightarrow V_{ref}} * T_{sweep,t}^{-1}, \text{ avec :} \quad (\text{III.1})$$

$$T_{V_i \rightarrow V_{ref}} = \begin{cases} T_{identite}, & \text{si } V_{ref} = V_i \\ T_{V_j \rightarrow V_{pano}}^{-1} * T_{V_i \rightarrow V_{pano}}, & \text{si } V_{ref} = V_j \end{cases}$$

Pour mesurer les erreurs obtenues avec ces trajectoires avant recalage, nous avons appliqué ce même processus de simulation aux maillages de prostate, pour les utiliser comme points géométriques lors du calcul de TRE. Pour cela, le maillage  $M_i$ , associé au volume initial  $V_i$ , est transformé en fonction de la trajectoire appliquée et noté  $M_t$ . Selon le niveau utilisé, il s'agit de comparer  $M_t$  et  $M_i$  (pour le niveau «  $I_{i,t}$  à  $V_i$  ») ou  $M_t$  et  $M_j$  (pour le niveau «  $I_{i,t}$  à  $V_j$  »). Dans le dernier cas, les maillages  $i$  et  $j$  n'intègrent pas les déformations anatomiques observées au sein des différents volumes, mais seulement les différences rigides. En effet, nous rappelons que seul le maillage du panorama  $M_{pano}$  est calculé par un algorithme de segmentation semi-automatique d'après le contenu du volume  $V_{pano}$ . Tous les autres maillages ( $M_i$ ,  $M_j$ , etc.) sont obtenus par la transformation rigide de ce même maillage.

Les Figures III.15-B et D illustrent l'erreur de TRE au cours d'un balayage base-apex selon les deux niveaux de simulation possible. L'erreur moyenne mesurée est d'environ 12 mm et retranscrit bien une diminution de la difficulté par rapport aux deux simulations précédentes, justifiant que la localisation par rapport aux volumes voisins (et anatomiquement plus similaires) est une tâche plus facilement réalisable que celle vers le panorama (pouvant être beaucoup plus différent).

Cette simulation combine certains avantages des simulations précédentes, tout en s'affranchissant des derniers biais que l'on voulait corriger, avec notamment :

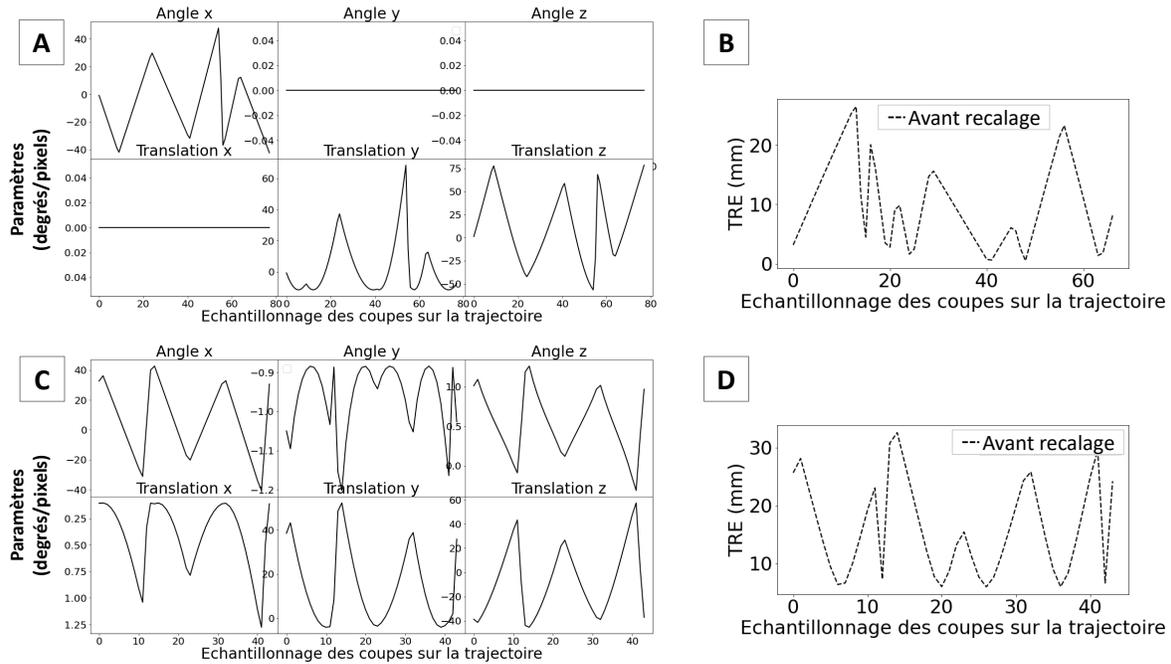


FIGURE III.15 – Visualisation des trajectoires obtenues par balayage base-apex et évolution du profil de la TRE : pour «  $I_{i,t}$  à  $V_i$  » en A et B respectivement, et pour «  $I_{i,t}$  à  $V_j$  » en C et D respectivement.

- L'intégration de déformations anatomiques entre les images d'entrée ( $I_1, \dots, I_t$ ) et l'espace de référence dans lequel on cherche à les localiser (pour le niveau «  $I_{i,t}$  à  $V_j$  »).
- Une suppression des zones noires (cône échographique d'orientation variable) qui pouvait altérer le bon apprentissage de notre modèle.
- Une réduction de la difficulté de la tâche initiale, grâce notamment à l'utilisation des volumes « voisins », i.e. acquis à un instant précédent et dont l'anatomie est assez similaire aux images du guidage en temps-réel que l'on cherche à localiser.
- La possibilité de nous comparer aux autres études de notre état-de-l'art, qui présentent le plus souvent des générations de données équivalentes à notre premier niveau de difficulté (pour le niveau «  $I_{i,t}$  à  $V_i$  »).

Cependant, la simplification des trajectoires simulées nous contraint alors à ne travailler qu'avec des mouvements unidirectionnels, comme un balayage autour d'un seul axe. Bien que ce mouvement, présentant une amplitude sur la totalité de la glande prostatique, soit réaliste, il ne représente pas toutes les possibilités d'orientations arbitraires que l'on peut avoir au cours d'un examen clinique. De plus, cette simulation de trajectoire ne présente aucune contrainte ou correspondance temporelle réaliste, il faudra alors adapter notre fréquence de rééchantillonnage pour reproduire au mieux une temporalité des mouvements de balayage base-apex réalistes.

La partie suivante résume les différents avantages et limites des trois procédés de simulation, décrits précédemment.

### III.1.3.4 Conclusion sur les simulations

La Table III.1 résume et compare les trois procédés de simulation proposés. Le simulateur présente l'avantage de générer des données à partir des mouvements très réalistes (en amplitude, vitesse et complexité), réalisés par plusieurs experts. Parmi toutes les données collectées, il ne dispose que du volume panorama pour chaque patient, ce qui contraint les rééchantillonnages et les localisations au sein de ce même volume. Bien que ce biais de génération (contenu iconique identique) simplifie la problématique de recalage, les mouvements complexes des trajectoires simulées génèrent des erreurs initiales (avant recalage) conséquentes et bien supérieures à celles trouvées dans la littérature. De plus, ils occasionnent des zones noires perturbant le réalisme de nos données.

Notre deuxième processus de simulation se concentre sur une complexification de la tâche avec un recalage vers un autre volume. L'utilisation des différents volumes successifs, enregistrés au cours d'une procédure, ainsi que la reconstruction de la trajectoire réalisée par interpolation, nous ont permis à la fois de générer des images présentant des déformations différentes et de simuler les translations de la sonde. Les trajectoires, encore trop complexes, mènent aux mêmes conséquences de zones noires et d'erreurs initiales trop importantes.

Enfin, notre dernier processus de simulation découle de la combinaison de deux premiers. Il permet de générer deux niveaux de difficultés, où les images rééchantillonnées à partir de balayage base-apex peuvent être localisées soit dans ce même volume, soit dans un volume voisin. Le premier niveau de difficulté est comparable à la plupart des processus de génération de la littérature, tandis que le niveau de difficulté supérieur permet d'évaluer la bonne généralisation de notre modèle sur des tâches plus réalistes. Cela permet également de générer des images sans zones noires arbitraires (même cône échographique), et des erreurs avant recalage moins grandes. Même si ce processus n'intègre pas toutes la complexité de trajectoires réelles (trajectoire incomplète, vitesse du geste), cette utilisation reste le meilleur compromis entre nos différentes simulations et celles de la littérature.

TABLE III.1 – Comparaison des avantages et limites des trois procédés de simulation envisagés selon plusieurs critères de réalisme des conditions cliniques et de biais inclus dans la génération.

Type de simulation	Complexité et réalisme	Multi experts	Amplitude du geste	Vitesse du geste	Patient spécifique	Paramètres rigides complets	Cohérence et déformation	Présence de zones noires	Erreur initiale élevée
Simulateur de biopsie	✓	✓	✓	✓	✓	✗	✗	✓	✓
Interpolation de recalage	✓	✗	✓	✗	✓	±	±	✓	✓
Balayage simplifié	±	✗	✓	✗	✓	✗	±	✗	✗

L'évolution de nos différents procédés de simulation vient donc de notre volonté de conserver un certain réalisme correspondant à la tâche clinique espérée, tout en nous affranchissant des biais pouvant être engendrés. Cette évolution, et les trois simulations qui en résultent, se sont construites tout au long de ce travail de thèse. Les résultats présentés dans les chapitres suivants seront donc obtenus à partir de ces différents procédés de génération de données.

Pour finir, toutes les données images présentées jusqu'à présent peuvent être recadrées en une

dimension particulière selon le jeu de données utilisé. On note la dimension ( $H \times W$ ) pour les images 2D et ( $H \times W \times D$ ) pour les images 3D. Lors de ce recadrage, on choisit de rogner les bords latéraux (selon  $x$  ou  $z$ ), et sur la profondeur du plan où la vessie apparaît (en  $y=0$ ). Le but est de garder un maximum d'information contextuelle échographique, de meilleure qualité au plus proche de l'insertion de la sonde ( $y=H$ ), là où les déformations de prostate dues à la pression de la sonde sont le plus visible. La Figure III.16 illustre les exemples de redimensionnement en 2D et 3D.

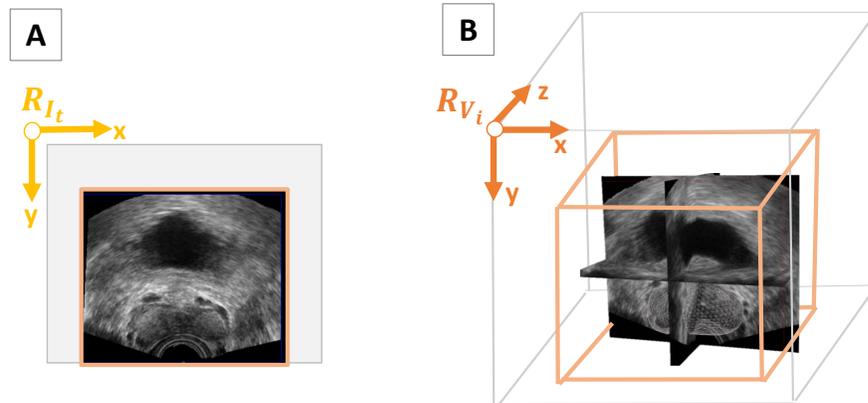


FIGURE III.16 – Exemple du processus de recadrage selon des axes préférentiels, pour des images 2D en A et des images 3D en B.

### III.1.4 Entrées optionnelles : représentation de la trajectoire réalisée

Parmi les différents types de représentation possibles liées à la trajectoire déjà réalisée, nous avons retenu les types de données positionnelles (estimation du recalage précédent, suivi sonde, flot optique), spatiales (coupe voisine, volume ou sous-volume de destination), ou encore structurales (masque binaire de segmentation). Elles découlent principalement des informations d'entrée et/ou de sortie issues des processus de génération décrits précédemment.

#### III.1.4.1 Données positionnelles

Nous définissons comme information positionnelle, toute information issue de capteurs ou d'a priori permettant de connaître des positions antérieures (relatives ou absolues) durant une trajectoire. Les définitions et notations des différentes données positionnelles envisagées sont décrites ci-dessous et illustrées dans la Figure III.17.

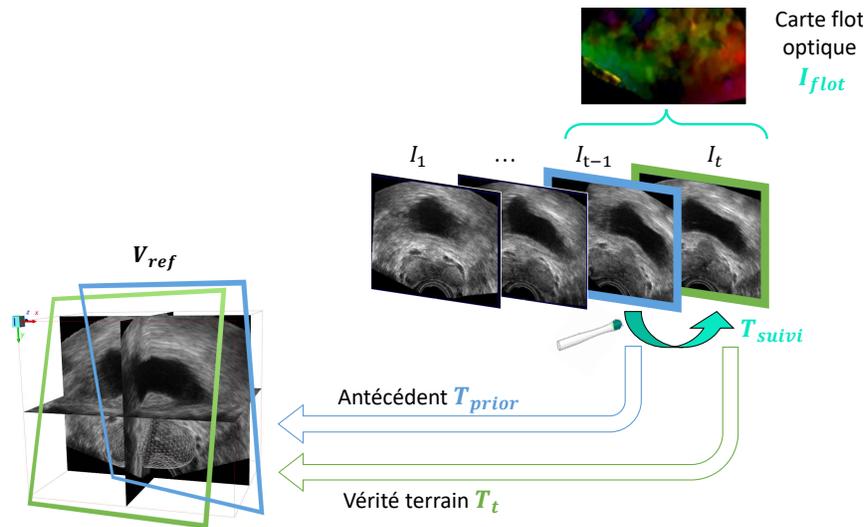


FIGURE III.17 – Définitions et notations des données optionnelles d'entrée positionnelles.

- Pour représenter un a priori de trajectoire, nous définissons comme antécédent du recalage courant, la variable  $T_{prior}$  correspondant à la localisation absolue de la coupe précédente, autrement dit, les résultats du recalage précédent. Ces antécédents peuvent être fournis de manière séquentielle théorique grâce aux labels de la coupe précédente ( $T_{prior} = T_{t-1}$ ) ou de manière séquentielle expérimentale avec les résultats des prédictions successives ( $T_{prior} = \widehat{T}_{t-1}$ ). Le premier antécédent est, quant à lui, toujours initialisé à partir de la valeur du premier label bruité ( $\widehat{T}_0 = T_0 + 1^\circ$ ). Ce dernier cas permet de simuler un cas réel d'utilisation clinique, où les prédictions seraient réalisées une à une de manière itérative. Dans les expériences, le même encodage sera utilisé pour représenter les entrées  $T_{prior}$  et les transformations à estimer  $\widehat{T}_t$  afin de conserver une cohérence dans les représentations de localisation.
- Nous définissons la variable  $T_{suivi}$ , s'apparentant à l'information issue d'un capteur inertiel IMU fixé à la sonde. Cette information représente alors un suivi de la sonde, correspondant

au déplacement relatif de la sonde entre deux temps successifs d'acquisition d'images. Dans notre cas, cette variable peut être calculée à partir des déplacements absolus successifs (voir équation III.2). Pour plus de réalisme, ce calcul peut être bruité avec un bruit additif aléatoire maximum de  $1^\circ$ , correspondant à l'erreur relative moyenne d'un capteur inertiel [96]. Avec ce calcul,  $T_{suivi}$  ne correspond pas uniquement aux mouvements de la sonde, mais peut combiner également les bougés de la prostate (notamment pour les simulations de déformations). L'utilisation de transformations absolues successives ne reflètent pas non plus de réels déplacements relatifs d'IMU, non calibrés par rapport aux autres référentiels. Là encore, cette entrée est fournie avec le même type d'encodage que celui utilisé pour toutes les localisations.

$$T_{suivi} = T_t^{-1} * T_{t-1} \pm 1^\circ \quad (\text{III.2})$$

- À partir de l'algorithme de Farnebäck [97], nous définissons la carte de flot optique  $I_{flot}$  entre deux images successives. Ce flot se représente par une image à deux canaux, une pour chaque composante du champ vectoriel bidimensionnel, décrivant le mouvement relatif dans le plan entre deux images.

#### III.1.4.2 Données spatiales

Nous définissons comme information spatiale, toute information concernant l'espace de destination par rapport auquel on cherche à définir la localisation optimale. Dans notre cas, l'espace de destination est  $V_{ref}$ , et peut représenter différentes acquisitions de volume selon le concept du volume de référence choisi. Cet espace peut être fourni de manière locale, globale ou partielle d'après les définitions suivantes :

- On définit comme information spatiale locale, la coupe voisine  $I_{t-1}$ . Celle-ci correspond à l'image 2D du guidage à l'instant précédent et communique des informations sur l'anatomie antérieure, selon la trajectoire réalisée.
- L'information spatiale globale correspond directement à l'ajout du volume de référence  $V_{ref}$ .
- Pour finir, l'information spatiale partielle correspond à une sous-partie du volume de référence. Ce « sous-volume » sera judicieusement déterminé en fonction d'un espace restreint en fonction de la trajectoire courante d'un instant  $t$ . On le note alors  $V_{sub,t}$ . Le processus de génération de ce sous-volume est illustré dans la Figure III.18. Celui-ci est obtenu après avoir transformé et orienté le volume de référence  $V_{ref}$  selon une transformation  $T_{sub,t}$ , qui repose sur l'estimation de la localisation de la coupe précédente  $I_{t-1}$ . Cette estimation du temps précédent reprend le concept d'antécédent ( $T_{prior}$ ), défini plus haut. L'espace de recherche est ainsi restreint, centré et recadré autour d'une position proche de celle recherchée. La localisation de la coupe courante est calculée par rapport au sous-volume intermédiaire ( $T_t = T_{I_t \rightarrow V_{sub,t}}$ ) comme décrit dans l'équation III.3a. Ensuite, des transformations relatives vers chaque sous-volume sont successivement composées (eq. III.3b) pour obtenir une transformation finale vers le volume de référence initial (eq. III.3c). Ceci est pratique pour toujours rééchantillonner à partir du même volume de référence complet et s'affranchir de potentielles zones noires.

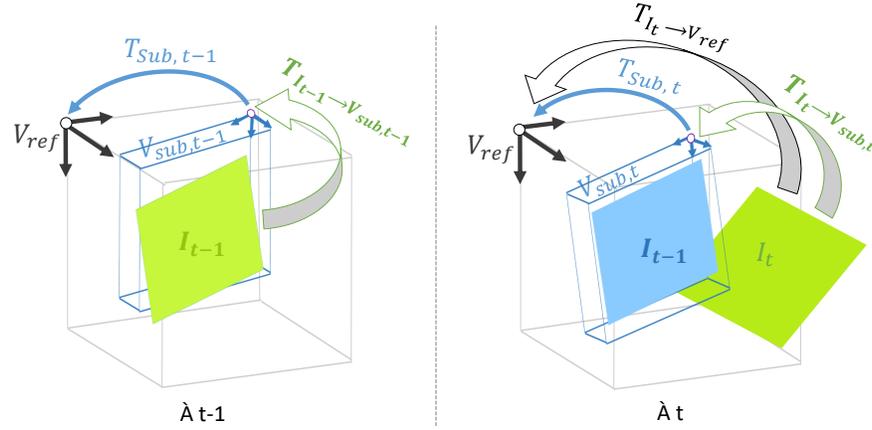


FIGURE III.18 – Génération des sous-volumes et calcul de la localisation de la coupe vers celui-ci.

$$T_t = T_{I_t \rightarrow V_{sub,t}} = T_{sub,t}^{-1} * T_{I_t \rightarrow V_{ref}} \quad (\text{III.3a})$$

$$\text{avec } T_{sub,t}^{-1} = T_{I_{t-1} \rightarrow V_{sub,(t-1)}}^{-1} * T_{sub,t-1}^{-1} \quad (\text{III.3b})$$

$$\text{et } T_{sub,0} = T_0 = T_{I_0 \rightarrow V_{ref}} \quad (\text{III.3c})$$

### III.1.4.3 Données structurelles

Nous définissons les informations structurelles comme des informations spatiales présentant les contours de de l'anatomie. Nous utilisons notamment les informations de segmentations de la prostate au sein des images disponibles grâce au maillage de référence ( $M_{ref}$ ). Les maillages  $M_t$  associés à chaque acquisition 3D sont obtenus à partir de la transformation rigide du maillage de référence. Une image binaire (3D) peut être calculée à partir du maillage de référence ( $M_{ref}$ ) à partir de l'application d'un filtre et d'un masque. Tout comme notre processus de simulation des données, cette image peut ensuite être transformée et rééchantillonnée, afin d'obtenir les segmentation successives ( $S_1, \dots, S_t$ ) associées aux images échographiques ( $I_1, \dots, I_t$ ). Le processus est illustré sur la Figure III.19.

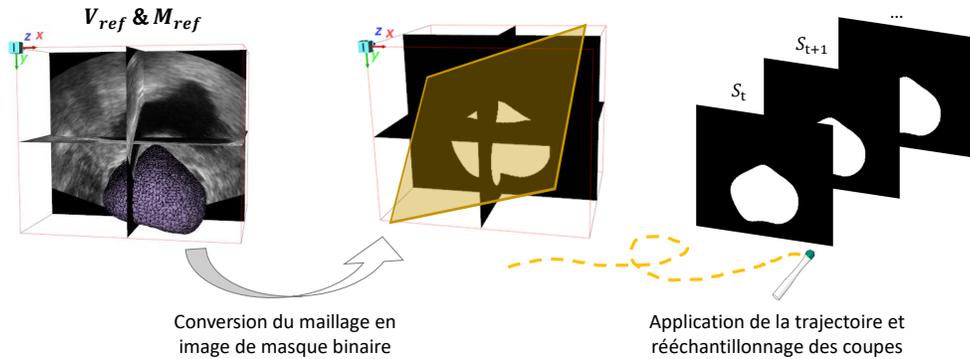


FIGURE III.19 – Définitions et notations des données optionnelles d'entrée structurelles.

## III.2 Protocole de validation

Il est également nécessaire d'évaluer la tâche de localisation à partir d'un protocole de validation des résultats. Celui-ci nécessite d'utiliser des métriques pertinentes et adaptées à la finalité clinique, qui sont calculées en respectant le bon réalisme et la faisabilité du déroulement clinique.

### III.2.1 Métriques

La validation de la qualité du recalage doit être une mesure pertinente évaluée sur un grand nombre d'éléments avec des annotations précises. Nous avons choisi d'évaluer la qualité du recalage à l'aide de métriques proposant des analyses complémentaires : (i) une analyse géométrique en utilisant la correspondance anatomique entre plusieurs points (TRE), (ii) une analyse iconique basée sur la similarité du contenu des images (NCC), et (iii) une analyse statistique comparant les résultats de régression avec la vérité du terrain (R2, MAE).

#### III.2.1.1 Métrique géométrique (TRE)

Une des métriques les plus répandues dans la littérature du recalage est la TRE. Celle-ci permet de mesurer la correspondance géométrique entre plusieurs points repères anatomiques, définis et localisés manuellement par des experts. Son calcul, simple et rapide, permet d'avoir des résultats comparables avec la plupart des études trouvées dans la littérature.

Dans notre cas, nous utilisons des points surfaciques du maillage de prostate  $M_{ref}$ , fournis par notre base de données et associés à l'espace de référence (soit panorama, soit volume de biopsie, selon le procédé de simulation). En notant les points de ce maillage  $p_k$  avec  $k = (1, \dots, N)$ , nous calculons la TRE entre l'ensemble des points appariés et transformés par  $T_t$  (transformation de vérité terrain utilisée durant notre simulation) et par  $\hat{T}_t$  (transformation prédite par notre modèle). Ce calcul permet de mesurer l'alignement et la superposition entre les deux maillages  $M_t$  et  $\hat{M}_t$  (voir Figure III.20). Le calcul de TRE est décrit dans l'équation III.4.

Il est important de rappeler que la métrique TRE, de par sa formule, donne une estimation de l'écart-type de la distribution normale de l'erreur. Dans notre cas, elle reflète l'écart entre la position prédite de la biopsie et sa position réelle. Selon les exigences cliniques, rappelons que cette erreur ne doit pas dépasser 5 mm (plus petite tumeur traitable). Pour répondre à ces exigences de précision clinique et pour obtenir environ 95% (correspondant à 2 écarts-types) des biopsies localisées dans la limite du rayon autorisé, la valeur de TRE doit être inférieure à 2.5 mm.

$$TRE(mm) = \sqrt{\frac{\sum_{k=1}^N \left\| (T_t(p_k) - \hat{T}_t(p_k)) \right\|^2}{N}} \quad (III.4)$$

Une telle évaluation basée sur le maillage présente différents avantages : (i) elle permet de calculer une TRE robuste, sur un grand nombre de points disponibles, souvent bien plus nombreux que les repères internes (fiduciels) qui nécessitent d'être annotés manuellement ; (ii) la nature 3D du maillage permet de considérer l'impact anatomique total et directement sur l'organe d'intérêt dans lequel les cibles cliniques sont situées. Cependant, la comparaison des points des maillages, tous obtenus par la transformation rigide du maillage de référence, se limite à l'évaluation du recalage

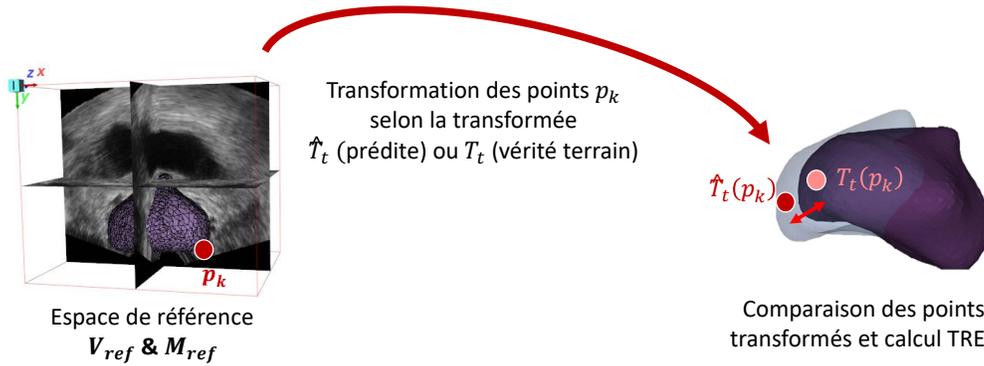


FIGURE III.20 – Calcul de TRE basé sur des points du maillage, transformés selon la prédiction de recalage ou la vérité terrain. Si la superposition n'est pas parfaite, on mesure la distance euclidienne entre les deux positions obtenues pour un même point  $p_k$ .

rigide. Ceci n'est pas un problème dans la plus grande majorité de nos processus de génération de données, où nous simulons des transformations rigides. En revanche, concernant le niveau de difficulté supérieure («  $I_{i,t}$  à  $V_j$  »), les images (issues de  $V_i$ ) sont simulées dans un espace présentant des déformations anatomiques où une évaluation rigide n'est pas suffisante pour évaluer la qualité du recalage. Pour ces cas-ci, il faut établir d'autres métriques complémentaires, permettant notamment de mesurer la qualité du recalage en fonction des déformations entre ces volumes. Les métriques iconiques, décrites par la suite, en sont un bon moyen.

### III.2.1.2 Métrique iconique (NCC)

La métrique NCC est une métrique iconique couramment utilisée en recalage permettant de mesurer la similarité entre deux images via un calcul de différence de pixels. Cette métrique permet donc de refléter la qualité du recalage directement à partir des images échographiques en jeu en considérant notamment les déformations anatomiques, visibles sur les images.

Dans notre cas, nous allons comparer l'image courante  $I_t$ , générée selon notre procédé de simulation de guidage échographique, à l'image obtenue à la position prédite. Cette image est obtenue en rééchantillonnant l'espace de référence ( $V_{ref}$ ), selon la transformation estimée par le modèle ( $\hat{T}_t$ ). Nous la notons  $I(V_{ref}, \hat{T}_t)$ . Pour toutes nos simulations où les deux images à comparer sont issues de la même anatomie (simulateur, interpolation, et «  $I_{i,t}$  à  $V_i$  »), une prédiction parfaite de recalage peut résulter en deux images parfaitement identiques et avec une valeur de similarité maximale. En revanche, pour la simulation «  $I_{i,t}$  à  $V_j$  » où les deux images illustrent différentes anatomies, le recalage optimal (vérité terrain) reflètera d'une valeur de similarité idéale moindre. Le calcul de NCC est décrit dans l'équation suivante (voir eq.III.5), où les pixels  $x_i$  et  $y_i$  correspondent aux pixels de l'image  $I_t$  et  $I(V_{ref}, \hat{T}_t)$ . Les valeurs sont bornées entre 0 et 1, où la valeur maximale correspond à la meilleure similarité entre les images.

$$NCC = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{III.5})$$

### III.2.1.3 Métrique statistique ( $R^2$ )

L'analyse de régression nécessite de comparer quantitativement et statistiquement les prédictions ( $\hat{T}_t$ ) aux vérités terrain ( $T_t$ ). Pour cela, nous avons utilisé le coefficient de détermination  $R^2$  qui mesure la capacité d'un modèle statistique à prédire un résultat.

En quantifiant en pourcentage, dans quelle mesure les prédictions correspondent aux observations (par rapport à l'équation de régression), le coefficient de détermination est utilisé pour identifier la force d'un modèle. Il est particulièrement adapté à l'évaluation de prédictions d'un modèle [98], contrairement aux coefficients de corrélation, qui eux mesurent la dépendance entre deux variables et sont donc sujets à des biais de magnitude. Le calcul du  $R^2$  est détaillé dans l'équation III.6. Ce score est compris entre 0 et 1, où 1 indique une correspondance parfaite tandis qu'un score égal (voire inférieur) à 0 indique des prédictions de l'ordre du hasard.

Cependant, de par la nature et la complexité d'un encodage de transformation géométrique, une analyse statistique sur les paramètres de transformation (rotation, translation, etc.) ne retranscrit pas systématiquement la qualité du recalage. En effet, il n'y pas unicité de la solution, et de faibles variations sur plusieurs paramètres combinés seront peut-être préférables à l'unique variation d'un autre paramètre. C'est pourquoi ce type de métrique statistique ne permet pas de s'affranchir des métriques mesurant l'impact anatomique du recalage (métriques géométriques ou iconiques présentées précédemment).

$$R^2 = 1 - \frac{(T_t - \hat{T}_t)^2}{\sqrt{(T_t - \hat{T}_t)^2}} \quad (\text{III.6})$$

Toutes ces métriques sont évaluées au cours d'une trajectoire, correspondant à une procédure d'examen de biopsie, où les erreurs sont moyennées sur la totalité de la trajectoire. La partie suivante détaille le protocole d'évaluation séquentielle proposé.

### III.2.2 Évaluation séquentielle sur une trajectoire complète

En utilisant différentes entrées liées à la trajectoire réalisée, ou plutôt « supposée » réalisée, il y a un risque d'accumulation d'erreur qui se répercute tout au long d'une trajectoire. Pour établir une évaluation réaliste, nous proposons de mesurer les erreurs à travers une série de recalages successifs, réalisés au cours d'une trajectoire complète. Cette évaluation séquentielle est cruciale pour mesurer les erreurs de déviation au cours d'une trajectoire, reflétant mieux l'impact qu'auraient des erreurs de recalage durant une procédure d'examen de biopsie clinique complète.

#### III.2.2.1 Analyse séquentielle

Afin de fournir une évaluation réaliste de la dérive observée, nous avons évalué une série de recalages, réalisés de manière séquentielle le long d'une trajectoire. Cette simulation séquentielle est réalisée en reproduisant une chronologie d'examen clinique, où les prédictions sont faites les unes à la suite des autres, en utilisant les résultats des prédictions précédentes ( $T_{prior} = \widehat{T}_{t-1}$ ). Nous qualifions cette évaluation d'analyse séquentielle cumulative, et l'illustrons sur la Figure III.21-B. La valeur de départ s'appuie sur la localisation exacte bruitée ( $T_0 \approx T_0 \pm 1^\circ$ ), qui pourra facilement être initialisée en pratique clinique par les résultats du recalage 3D/3D intégré aux plateformes cliniques que nous utilisons. Nous évaluons également l'analyse séquentielle théorique, où les résultats précédents sont fournis à partir des localisations exactes ( $T_{prior} = T_{t-1}$ ), afin de quantifier les performances maximales et idéales que l'on peut espérer obtenir (voir Figure III.21-A).

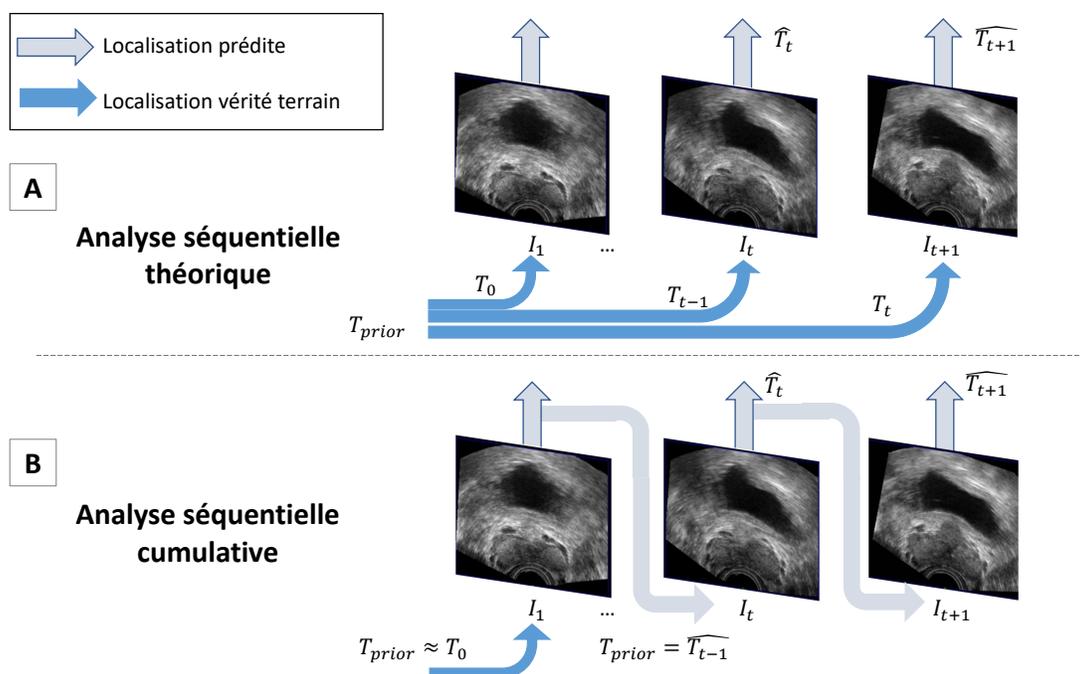


FIGURE III.21 – Évaluation séquentielle avec une série de recalage successifs réalisés le long d'une trajectoire.

### III.2.2.2 Navigation par portion

En pratique clinique, la navigation se réalise par portion, entre le moment du changement de l'aiguille jusqu'à la réalisation de la prochaine biopsie. Avant le prélèvement d'une nouvelle biopsie, le praticien réalise généralement une acquisition 3D lui permettant de vérifier son ciblage grâce au recalage 3D/3D dont il dispose à partir des plateformes cliniques utilisées [17]. L'exploitation de ce recalage à chaque nouvelle biopsie, couramment utilisée dans un cadre clinique réaliste, nous a alors incités à proposer un protocole d'évaluation séquentielle par portion.

Nous envisageons, pour cela, une procédure de biopsie avec une assistance à la navigation entre chaque site de biopsie. Les trajectoires simulées ont été divisé en 12 sous-parties afin de reproduire un « pseudo » schéma de biopsie standard. Ces 12 moments de biopsies, bénéficiant des recalage 3D/3D (intégrés aux plateformes cliniques), peuvent alors être considérés comme 12 corrections clés permettant de réinitialiser la valeur de la prédiction  $T_{prior}$ . Ce procédé est illustré sur la Figure III.22. Cela permet ainsi de réajuster et limiter un potentiel phénomène d'accumulation d'erreur possible au cours de la trajectoire complète.

Ainsi, durant notre protocole d'évaluation séquentielle, l'erreur mesurée est accumulée sur une succession de recalages le long d'une trajectoire complète, jusqu'au prochain réajustement de recalage. Ce protocole respecte parfaitement le réalisme des enjeux cliniques.

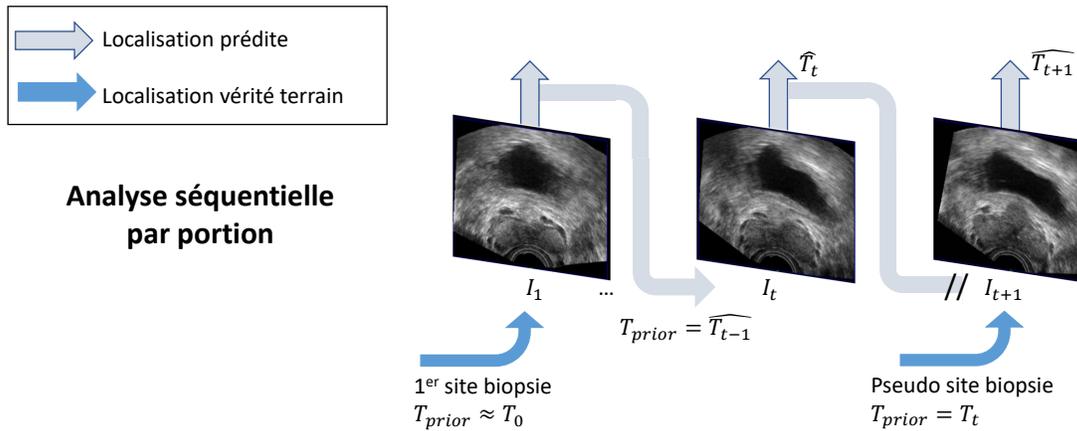


FIGURE III.22 – Évaluation séquentielle par portion, jusqu'au prochain réajustement de recalage au «pseudo» site de biopsie.

**Points clé**

- ❖ L'objectif est de simuler une succession d'images échographiques, où la localisation optimale de chaque image courante  $I_t$  par rapport à un volume de référence  $V_{ref}$ , est connue et est définie par la transformation rigide  $T_t$ .
- ❖ Nous avons proposé plusieurs procédés de simulation, visant à trouver le meilleur compromis entre réalisme et biais occasionné. Les rééchantillonnages, à partir de trajectoires trop complexes, génèrent des « zones noires » perturbant le réalisme de nos données. L'utilisation de différents volumes d'échantillonnage permet de représenter différentes déformations anatomiques, tandis que la localisation au sein de ce même volume ou d'un volume voisin permet de générer plusieurs niveaux de difficultés. On peut alors mieux se comparer à la littérature mais également mieux évaluer la bonne généralisation sur des tâches plus réalistes par rapport aux objectifs cliniques.
- ❖ Nous proposons la combinaison de métriques géométriques, iconiques et statistiques pour la bonne évaluation de notre méthode de recalage. Toutes ces métriques seront utilisées au travers d'une évaluation séquentielle où l'erreur est mesurée et accumulée à travers une série de recalages successifs, réalisés au cours d'une trajectoire complète.
- ❖ Les recalages 3D/3D US estimés à chaque nouveau site de biopsie, par des plateformes d'assistance, sont couramment utilisés dans un cadre clinique réaliste. Leur exploitation est envisageable dans notre protocole d'évaluation séquentielle par portion, comme correction d'un potentiel phénomène de dérive.

---

# Recalage 2D/3D avec des informations positionnelles

---

*Ce* chapitre propose une première méthode, par apprentissage profond, pour la réalisation d'une tâche de recalage 2D/3D durant la navigation de biopsie prostatique. Cette méthode s'intéresse particulièrement à l'apport d'informations « positionnelles », permettant de considérer certains a priori de la trajectoire réalisée et pouvant faciliter la prédiction de recalage durant la navigation. Les objectifs principaux de ce chapitre sont d'évaluer les performances de recalage 2D/3D en temps-réel et de quantifier l'apport des informations liées aux a priori positionnels de trajectoire pour cet apprentissage. Plusieurs expériences ont été réalisées : selon l'ajout et la combinaison des données d'entrée (étude d'ablation) ; selon le type d'encodage des transformations (angle euler, coupe sin/cos, coordonnées) ; et selon différentes caractéristiques des jeux de données testés (fréquence d'échantillonnage, expertise du praticien). Nous présentons les résultats de ces différentes expériences, avant de conclure quant aux contributions méthodologiques et expérimentales à retenir, en vue de possibles améliorations et perspectives.

---

## Sommaire

---

<b>IV.1 Objectifs et contributions</b>	<b>87</b>
<b>IV.2 Matériel &amp; Méthodes</b>	<b>88</b>
IV.2.1 Approche proposée	88
IV.2.1.1 Définitions des données d'entrée	88
IV.2.1.2 Architecture du réseau	89
IV.2.2 Apprentissage	90
IV.2.2.1 Fonction de coût principale	90
IV.2.2.2 Paramètres d'apprentissage	90
<b>IV.3 Expériences et configurations</b>	<b>91</b>
IV.3.1 Étude d'ablation	91
IV.3.2 Étude d'interprétabilité	92
IV.3.3 Comparaison des encodages	92
IV.3.4 Jeux de données	93
IV.3.4.1 Caractéristiques de la base de données principale	93
IV.3.4.2 Autres expériences et caractéristiques	94
<b>IV.4 Résultats</b>	<b>95</b>
IV.4.1 Étude d'ablation	95
IV.4.1.1 Analyse quantitative	95
IV.4.1.2 Analyse temporelle	97
IV.4.1.3 Analyse qualitative	99
IV.4.2 Étude d'interprétabilité : biais des zones noires	100
IV.4.3 Variation des caractéristiques du jeu de données	102
IV.4.3.1 Fréquence d'échantillonnage	102
IV.4.3.2 Niveaux d'expertises	103
<b>IV.5 Discussion</b>	<b>104</b>
IV.5.1 Discussion sur les contributions méthodologiques	104
IV.5.1.1 Différents scénarios testés	104
IV.5.1.2 Ajout de $T_{prior}$	105
IV.5.2 Discussion sur les contributions expérimentales	106
IV.5.2.1 Validité de la génération de données	106
IV.5.2.2 Évaluation séquentielle sur des trajectoires complètes	106
IV.5.2.3 Réalisme et faisabilité clinique	106
IV.5.3 Conclusions sur ce premier modèle	107
IV.5.3.1 Conclusions sur les performances obtenues	107
IV.5.3.2 Comparaison avec la littérature	107
IV.5.3.3 Conclusions sur les objectifs	108

---

## IV.1 Objectifs et contributions

Afin de permettre l’assistance à la navigation pendant les procédures de biopsie, l’objectif de nos travaux est de développer et d’évaluer des méthodes de recalage 2D/3D en temps-réel permettant de localiser à chaque instant  $t$  l’image courante  $I_t$  (issue du guidage 2D US) par rapport au volume défini comme « volume de référence »  $V_{ref}$  (3D US, acquis en début de navigation). La transformation rigide 2D/3D résultante est désignée par  $T_t$ . Rappelons que ce recalage doit satisfaire les exigences cliniques, tant en termes de précision ( $TRE \leq 2.5$  mm) qu’en termes d’efficacité de calcul (prédiction  $\leq 50$  ms par image).

Alors que les méthodes de recalage classiques exploitent des informations sur le volume de destination ou des initialisations de transformation (lorsque celles-ci sont disponibles), les méthodes de recalage 2D/3D par apprentissage profond ne considèrent généralement que la coupe à localiser (voir section II.3.1). Ce manque d’information peut cependant impacter les performances d’une telle tâche. Fortement motivés par l’idée de considérer des informations contextuelles de déplacements absolus et relatifs, nous décidons de nous concentrer sur l’apport d’a priori positionnel pour mieux considérer la trajectoire réalisée au sein de l’espace de référence.

Les trois principaux objectifs de ce premier travail sont :

- De vérifier qu’une telle tâche de recalage 2D/3D est réalisable en utilisant des outils d’apprentissage profond. Pour cela, nous proposons une architecture supervisée de réseau convolutif (CNN), capable de prédire les paramètres de transformation de manière directe. Nous vérifierons qu’un tel cadre de travail permet d’obtenir des résultats cohérents et comparables avec ceux de la littérature, tout en respectant nos attentes cliniques.
- De démontrer qu’en considérant des informations liées à la trajectoire réalisée, via des a priori « positionnels », on peut faciliter la tâche de recalage durant la navigation. Pour cela, nous proposons des configurations flexibles pour mener des études complètes d’ablation où nous comparons plusieurs formes et combinaisons d’entrées : (i) la paire de coupes successives, (ii) le flot optique relatif, (iii) le suivi de sonde relatif, et (iv) les résultats du recalage précédent. Les manières de les intégrer à notre réseau seront également étudiées.
- De valider la robustesse de notre méthode sur nos procédés de génération de données réalistes et sur nos protocoles d’évaluation séquentielle. Pour cela, nous nous concentrons sur la base de données provenant du simulateur de biopsie Biopsym, décrite dans le chapitre précédent (section III.1.3.1). Nous évaluerons notamment plusieurs fréquences d’échantillonnage et plusieurs niveaux d’expertise des praticiens, participant aux caractéristiques de cette base de données. Nous analyserons les résultats des différentes métriques proposées sur des trajectoires complètes, afin d’évaluer le phénomène d’accumulation d’erreur.

La section suivante présente la méthode proposée avec les différentes stratégies d’ajout d’informations positionnelles et les détails de leur intégration au sein de l’architecture du réseau utilisé.

## IV.2 Matériel & Méthodes

### IV.2.1 Approche proposée

Les travaux de cette première méthode ont été inspirés de l'étude de Prevost et. al [76] (voir section II.3.3), notamment en ce qui concerne l'architecture du réseau et l'ajout de données d'entrée. La Figure IV.1 illustre notre approche qui comprend plusieurs éléments : une branche principale de réseau convolutif utilisant (i) une paire d'images successives en données d'entrée principales, des entrées optionnelles de type (ii) flot optique, (iii) suivi de sonde, ou (iv) antécédents. Tous ces éléments sont concaténés à différents niveaux de notre architecture.

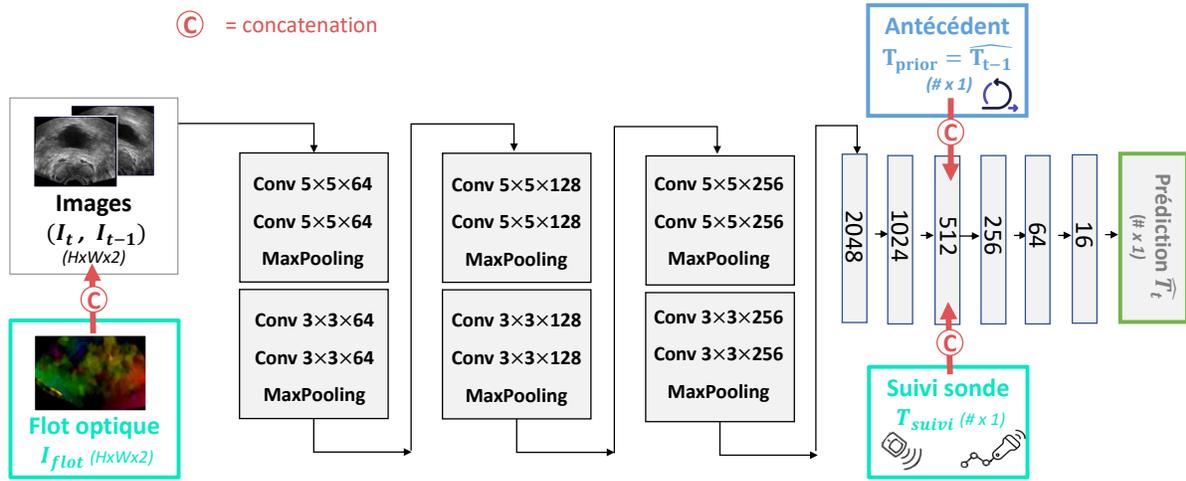


FIGURE IV.1 – Architecture du réseau de convolution (CNN) proposé ainsi que les différentes entrées selon le scénario envisagé.

#### IV.2.1.1 Définitions des données d'entrée

Les différentes entrées possibles sont décrites ci-dessous et illustrées dans la Figure IV.1. Chacune de ces entrées est ajoutée séparément, par le biais de différentes expériences qui seront détaillées dans la section IV.3.1. Pour rappel, le chapitre précédent (voir III.1.4.1) décrit plus en détail la manière dont ces entrées sont calculées et simulées.

- La paire d'images (entrée principale) :  $I_t$  et  $I_{t-1}$ . Chaque image est recadrée au préalable à la taille  $(H \times W)$ , selon les dimensions de hauteur ( $H$ ) et de largeur ( $W$ ) du jeu de données testé (section IV.3.4). La paire est introduite dans le réseau sous la forme d'une image 3D à deux canaux (concaténées sur la profondeur) de sorte que les informations spatiales soient couplées entre elles dès la première couche de convolution du réseau. Cette entrée est l'entrée principale de notre réseau et sera commune à chacune des expériences envisagées.
- Le flot optique (entrée optionnelle) :  $I_{flot}$ , de dimension  $(H \times W \times 2)$ . Cette image, est concaténée à la paire d'images principales, de façon à former une image d'entrée à 4 canaux. Le

but est également de préserver la correspondance spatiale, localement traitée sur chaque zone de noyau de convolution, afin de mieux combiner l'information de déplacement relatif du flot avec les informations initiales contenues dans les images.

- Suivi de la sonde (entrée optionnelle) :  $T_{suivi}$ , correspondant au déplacement relatif de la sonde entre deux temps successifs d'acquisition d'images. Cette entrée se présente sous la forme d'un vecteur de taille  $(\# \times 1)$  où  $\#$  correspond aux nombres de paramètres selon le type d'encodage des transformations envisagé (voir section IV.3.3). Quelle que soit sa taille, cette entrée est concaténée sur la couche de FC-512 du réseau (voir Figure IV.1), similairement à [76].
- Antécédent (entrée optionnelle) :  $T_{prior}$ , correspondant à la localisation de la coupe précédente. Celui-ci est ajouté sur la même couche du réseau (FC-512) que l'entrée optionnelle  $T_{suivi}$ , mais à travers une concaténation de vecteur indépendante. De même que précédemment, ce vecteur se compose des paramètres de transformation, de taille  $(\# \times 1)$ , variable selon l'encodage envisagé (voir section IV.3.3). Nous rappelons que ces antécédents peuvent être fournis grâce aux labels de la coupe précédente ( $T_{prior} = T_{t-1}$ ) dans un cadre séquentiel théorique (comme l'apprentissage) ou dans un cadre séquentiel expérimental avec les résultats des prédictions successives ( $T_{prior} = \widehat{T}_{t-1}$ ).

#### IV.2.1.2 Architecture du réseau

Nous avons utilisé une architecture de réseau convolutif, simple et très répandue dans notre état-de-l'art. Celle-ci est illustrée dans la Figure IV.1 et plus amplement détaillée en Annexe A (voir Table 1).

Le réseau se compose de 12 couches de convolution 2D (noyau= $5 \times 5$  ou  $3 \times 3$ ), adaptées à la nature 2D des images d'entrée. Chaque couche de convolution est suivie d'une fonction d'activation ReLU ("Rectified Linear Unit") et d'une couche de "pooling" (noyau= $2 \times 2$ ). Nous avons déterminé le nombre de couches de convolution et les paramètres d'espacement ("stride") de façon à obtenir une taille d'image en fin d'encodage suffisamment grande et interprétable, tout en présentant un concentré du contenu iconique. Enfin, le réseau se termine par une succession de couches entièrement connectées (de l'anglais "Fully Connected", noté FC), qui sont appliquées sur l'image encodée en vecteur 1D ("flatten"). Le nombre de paramètres à estimer diminue progressivement jusqu'à obtenir la couche finale d'estimation des paramètres de transformation  $\widehat{T}_t$ , de taille variable selon l'encodage choisi ( $\# \times 1$ ).

Comparativement à l'étude dont nous nous sommes inspirés [76], nous avons choisi d'utiliser un encodage de convolution plus profond (12 couches contre 4), avec un nombre de filtres par convolution croissant et plus élevé (jusqu'à 256 filtres). Une architecture plus profonde permet notamment de traiter des problèmes plus complexes grâce à de plus nombreuses analyses non-linéaires. Il convient cependant de trouver un équilibre entre la profondeur traitée et le nombre de paramètres à optimiser. Pour cela, le paramètre d'espacement ("stride") entre les noyaux permet de réduire la taille de l'image obtenue, et donc d'agrandir le champ de vision de la zone traitée.

Le nombre total de paramètres à optimiser est décrit en Annexe A et se calcule en fonction de la taille de l'image d'entrée du réseau ( $H \times W$ ). Dans notre cas, ceux-ci totalisent environ 24 M de paramètres sur l'ensemble des couches du modèle (voir Table 1).

## IV.2.2 Apprentissage

### IV.2.2.1 Fonction de coût principale

Le modèle est entraîné d’une manière supervisée où les paramètres du réseau (poids, filtre, biais, etc.) sont optimisés et mis à jour en minimisant la valeur calculée par la fonction de coût. Choisir une fonction de coût pertinente et adaptée est essentiel pour une optimisation correcte, robuste, et rapide du réseau. Pour cette première approche, nous avons choisi d’utiliser une fonction de coût basée sur le calcul d’erreur quadratique moyenne des paramètres (MSE pour “Mean Square Error”). L’erreur est calculée à partir de la différence entre les valeurs des paramètres de transformation prédites  $\widehat{T}_k$  et les vérités terrains  $T_k$ , pour chaque élément  $k = (1, \dots, K)$  de l’entraînement. Le calcul est défini dans l’équation IV.1a. Cette fonction sera la fonction de coût principale de notre réseau, applicable pour n’importe quel type d’encodage de sortie envisagé.

La fonction MSE présente l’avantage d’être fortement convexe, privilégiant la correction des erreurs les plus aberrantes dès les premières itérations. Cependant, comme toutes les métriques statistiques (précédemment discutées en III.2.1.3), cette mesure ne retranscrit pas systématiquement la qualité du recalage 2D/3D des images en termes de bonne superposition des structures anatomiques. En effet, sans unicité de la solution ni sans même ordre de grandeur entre les paramètres, une telle métrique peut être limitante durant l’optimisation.

$$L_{MSE} = \frac{1}{K} \sum_{k=1}^K MSE_k \quad (\text{IV.1a})$$

$$\text{avec } MSE_k = \left\| T_k - \widehat{T}_k \right\|^2 \quad (\text{IV.1b})$$

### IV.2.2.2 Paramètres d’apprentissage

Toutes les expériences ont été réalisées en entraînant notre réseau à partir de l’optimiseur Nadam [99] et avec un taux d’apprentissage initial de  $1e-4$ . Les poids du réseau ont tous été initialisés avec une distribution Gaussienne (moy=0, std=0.01). Le réseau a été entraîné sur plus de 50 epochs, avec une taille de lot  $K= 500$ , où le terme "epoch" est un anglicisme désignant un cycle complet d’apprentissage sur l’ensemble du jeu de données. C’est avec toutes ces conditions réunies que nous avons obtenu la convergence la plus robuste et rapide. Le modèle a été entraîné sur une GPU Titan RTX en utilisant une implémentation via TensorFlow [100]. L’entraînement dure environ 5 min par epoch, tandis que la durée de prédiction est d’environ  $30 \pm 10$  ms.

## IV.3 Expériences et configurations

### IV.3.1 Étude d’ablation

Afin de déterminer indépendamment quelle entrée positionnelle permet de faciliter la tâche de recalage, nous avons réalisé différentes expériences d’ablation, où les entrées sont ajoutées séparément. L’objectif est de choisir le scénario le plus robuste et le plus précis. Les 4 scénarios envisagés selon les combinaisons possibles de ces données d’entrée sont :

- Scénario 1 : seulement la paire d’images  $(I_t, I_{t-1})$
- Scénario 2 : paire d’images  $(I_t, I_{t-1})$  + le flot optique  $I_{flot}$
- Scénario 3 : paire d’images  $(I_t, I_{t-1})$  + suivi de la sonde  $T_{suivi}$
- Scénario 4 : paire d’images  $(I_t, I_{t-1})$  + flot optique  $I_{flot}$  + suivi de la sonde  $T_{suivi}$ .

Chacun de ces scénarios peut être également testé avec ou sans ajout de l’antécédent  $T_{prior}$  afin d’inclure davantage le concept de trajectoire déjà réalisée grâce aux résultats de localisations précédentes. Ces expériences seront respectivement intitulées « Avec  $T_{prior}$  » et « Sans  $T_{prior}$  ».

Lors de l’ajout de cet antécédent, celui-ci sera intégré de deux manières différentes : à partir des labels théoriques précédents ( $T_{prior} = T_{t-1}$ ) ou à partir des résultats des prédictions précédentes ( $T_{prior} = \widehat{T}_{t-1}$ ). Dans le deuxième cas, l’évaluation séquentielle peut être réalisée sur la trajectoire complète ou par portion, en utilisant les résultats successifs entre deux « pseudo » sites de biopsie (voir section III.2.2.2). Ces expériences seront respectivement intitulées : évaluation séquentielle « théorique », « complète », ou « par portion ». Par défaut, si aucune mention n’est spécifiée, le protocole d’évaluation privilégié pour présenter les résultats est le protocole d’évaluation séquentielle *par portion*, jugé plus adapté pour notre application clinique.

Cependant, les informations de l’entrée  $T_{prior}$  (résultats de la localisation de  $I_{t-1}$  par rapport à  $V_{ref}$ ), notamment lorsque combinées aux informations de  $T_{suivi}$  (déplacement de la sonde entre  $t-1$  et  $t$ ), peuvent être très proches du résultat  $T_t$  que l’on cherche à estimer. Il faut alors surveiller le risque d’une potentielle surpondération de ces entrées par le réseau. Pour confirmer que ces deux informations de localisation (combinées ou non) ne biaisent pas la capacité du réseau à apprendre à partir de la contribution de toutes les données (notamment les images), nous avons considéré une évaluation de contrôle qui calcule la position de l’image en utilisant uniquement les transformations géométriques :  $T_{prior}$  seule (pour les scénarios 1 ou 2) ou la composition  $T_{prior} * T_{suivi}$  (pour les scénarios 3 ou 4). Cette évaluation, sans utilisation ni du réseau ni des images d’entrée, sera identifiée par l’appellation « contrôle ».

Chaque évaluation séquentielle *théorique/complète/par portion*, peut alors être associée à sa propre évaluation de contrôle, en fonction de l’évolution spécifique de  $T_{prior}$  au cours de la trajectoire ainsi évaluée. Dans notre cas, nous calculons le *contrôle* par rapport à l’évaluation séquentielle *par portion*, que nous privilégions. La comparaison entre ces deux expériences permet de vérifier la bonne contribution du réseau et le bon traitement de toutes les données d’entrée.

### IV.3.2 Étude d’interprétabilité

Étudier l’interprétabilité en apprentissage profond, c’est viser à comprendre les raisons qui ont motivé le modèle à prédire tel ou tel résultat. Il existe plusieurs outils de visualisation au service de l’explicabilité des prédictions, notamment dans le domaine du traitement d’images et des réseaux de convolution [101]. Parmi celles-ci :

- La visualisation de « cartes d’activation » (de l’anglais “Activation maps”), obtenues à partir de l’algorithme Grad-CAM (Gradient-weighted Class Activation Mapping) [102], permet de spécifier le rôle et l’importance de chaque élément d’entrée (pixels de l’image) par rapport à une décision de classification. L’évolution du gradient par rapport aux sorties de chaque couche de convolution permet d’obtenir des cartes thermiques, appliquées et moyennées à toutes les couches du réseau, mettant en évidence les régions de l’image ayant le plus d’importance pour la prédiction du résultat. Ce calcul est obtenu au moment de la rétropropagation et est spécifique à une unique classe de sortie. Dans notre cas, chaque paramètre de transformation (#) pourra alors être analysé de manière indépendante pour mieux étudier les attributs de l’image d’entrée ayant participé à la bonne prédiction de ce paramètre.
- La visualisation de « cartes caractéristiques » (de l’anglais “Features maps”), permet d’observer les filtres construits lors de la phase d’entraînement. Les filtres des couches supérieures présentent généralement des motifs simples et facilement interprétables, tandis que les filtres plus profonds illustrent des formes globales sur du contenu iconique moyenné. Tous les filtres obtenus sur la dernière couche de convolution sont alors moyennés et affichés. A cause de sa résolution différente de celle des images d’entrée, la carte de caractéristique moyenne n’est pas superposée à l’image d’entrée.

### IV.3.3 Comparaison des encodages

Nous avons vu précédemment qu’il existe différentes façons de représenter une transformation rigide et notamment le paramétrage de la rotation (voir section II.4.2). Parmi les différentes représentations envisagées dans la littérature, nous avons choisi de comparer 3 types d’encodage pour représenter les transformations de sortie  $T_t$  : les angles d’Euler, le couple sin/cos et les coordonnées de l’image. Chacun des ses encodages a été testé en adaptant le type d’architecture du réseau (principale, U-net) et en utilisant différentes fonctions d’optimisation (terme de pénalisation mathématique, terme de distance d’alignement).

La description des différentes expériences réalisées ainsi que la présentation des résultats et conclusions obtenus sont présentées dans l’Annexe B, tandis que pour le reste de ce chapitre, seul un encodage avec les angles d’Euler est présenté.

### IV.3.4 Jeux de données

Pour réaliser les différentes expériences de cette première proposition méthodologique, nous nous sommes appuyés sur la base de données issue du simulateur Biopsym, présentée dans le chapitre précédent (voir section III.1.3.1). Nous détaillons les caractéristiques associées à cette base de données ci-dessous.

#### IV.3.4.1 Caractéristiques de la base de données principale

Nous avons demandé à un praticien expérimenté de venir réaliser des trajectoires de biopsies standardisées sur le simulateur. L’expertise d’un praticien est définie selon le nombre d’examen de biopsie réalisés en clinique : les praticiens sont considérés comme débutants s’ils ont réalisé moins de 20 biopsies, et comme experts s’ils en ont réalisé plus de 100. Nous avons enregistré 7 trajectoires que nous avons appliquées sur une large base de données contenant 100 volumes panorama provenant de 100 patients différents. Appliquer chaque trajectoire à plusieurs volumes nous a permis d’augmenter la taille de notre jeu de données. Pour les jeux de données test, nous avons pris soin de sélectionner des trajectoires parfaitement nouvelles sur des patients complètement nouveaux. En effet, afin de favoriser la généralisation du modèle sur plusieurs anatomies, nous basons notre répartition entre les jeux d’entraînement et de test selon le nombre total de patients plutôt que sur le nombre total d’échantillons d’images. La description de la séparation des données est décrite dans la Table IV.1.

Nous rappelons que ces trajectoires respectent un schéma de procédure de biopsies standardisées (voir section III.1.3.1), où le praticien simule jusqu’au déclenchement de tir des biopsies. Les 12 tirs de biopsies sont alors parfaitement identifiés au cours des trajectoires et seront notamment utilisés pour établir la navigation par portion, entre deux « pseudo » sites de biopsies. La fréquence d’échantillonnage choisie pour ce jeu de donnée est fixée autour de 6.6 Hz, permettant d’obtenir des images toutes les 150 ms. Toutes les images sont recadrées en  $130 \times 250$  pixels.

Pour finir, en raison de la construction des données limitée par l’utilisation d’un simulateur (voir la section III.1.3.1), seuls les 3 paramètres de rotation ont été estimés dans cette première étape des travaux. Au final, les dimensions des sorties selon l’encodage choisi sont :  $T_t = (\theta_x, \theta_y, \theta_z)$  avec  $\# = 3$  pour les angles d’Euler.

TABLE IV.1 – Description du jeu de données issu du simulateur avec les pourcentages de répartition des échantillons d’images entre les jeux d’entraînement, de validation, et de test.

Base de données	Entraînement & Val (87 % & 5 %)	Test (8%)	Caractéristiques
"Biopsym"	5 traj x 80 patients $\approx 53\,000$ images	2 traj x 20 patients $\approx 4\,500$ images	Trajectoire de biopsie aléatoire Amplitude $\pm[30^\circ \text{ x } 180^\circ \text{ y } 20^\circ \text{ z}]$ Praticien : Expert ( $\geq 100$ examens) Taille des images : $130 \times 250$ Fréquence d’échantillonnage : 150 ms

#### IV.3.4.2 Autres expériences et caractéristiques

Afin d'évaluer la capacité de généralisation de notre modèle selon différents niveaux d'expertise du praticien, nous avons entraîné deux modèles différents que nous avons évalués sur deux bases de données différentes. L'objectif est de valider les capacités d'un modèle à généraliser ses performances de prédiction de recalage sur des trajectoires d'expertise différente de celle utilisée durant sa phase d'apprentissage. Ces bases de données ont été générées de la même manière que celle décrite précédemment (section IV.3.4.1), grâce à des trajectoires réalisées sur le simulateur, soit (i) par un praticien expert soit (ii) par un praticien débutant. Elles sont respectivement dénommées « *Manip Expert* » et « *Manip Débutant* ». En particulier, la base de données « *Manip Expert* » correspond à notre base de données principale décrite précédemment (Table IV.1). Les modèles, entraînés à partir de ces deux bases de données différentes, sont respectivement dénommés « *Modèle Expert* » et « *Modèle Débutant* ».

Enfin, plusieurs fréquences d'échantillonnage entre les images successives ont été comparées. La fréquence d'acquisition du simulateur étant de l'ordre de 100 Hz (10 ms), il est possible de sélectionner différents pas de temps lors de la génération des images. Nous avons ainsi généré 4 bases de données (présentées sur la Table IV.2), sur le même principe que celle décrite précédemment, avec différentes fréquences d'échantillonnage des images : 1 Hz (1000 ms), 2 Hz (500ms), 6.6 Hz (150 ms) et 20 Hz (50 ms). Ces bases de données résultant en un nombre d'images différent, nous avons choisi une répartition commune et réduite entre les jeux d'entraînement et de test pour avoir des résultats comparables : 16 000 images (Entraînement & Val) et 1300 images (Test).

TABLE IV.2 – Description du jeu de données réduit spécifique aux expériences de variations de la fréquences d'échantillonnage. Les nouveaux pourcentages de répartition des échantillons d'images entre les jeux d'entraînement, de validation, et de test sont présentés.

Base de données	Entraînement & Val (87 % & 5 %)	Test (8%)	Caractéristiques
<b>Biopsym réduit</b> <b>Freq <math>\nu</math></b>	5 traj x 80 patients $\approx$ 16 000 images	2 traj x 20 patients $\approx$ 1 300 images	Trajectoire de biopsie aléatoire Praticien : Expert ( $\geq$ 100 examens) Taille des images : $130 \times 250$ Fréquence d'échantillonnage : $\nu$ variable entre [1-20] Hz

## IV.4 Résultats

### IV.4.1 Étude d’ablation

#### IV.4.1.1 Analyse quantitative

Dans les résultats qui vont suivre, nous avons comparé les performances avant et après recalage selon les 4 scénarios envisagés, et selon l’ajout ou non des antécédents  $T_{prior}$ . La Table IV.3 résume l’ensemble des valeurs moyennes de TRE tandis que la Figure IV.2 illustre les distributions (et leur médiane) des différentes métriques utilisées.

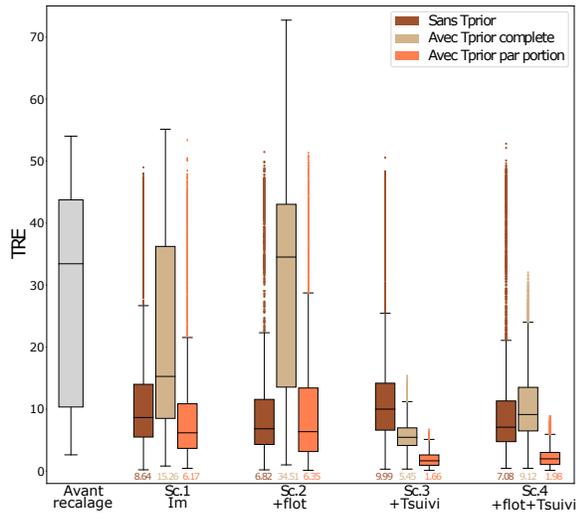
TABLE IV.3 – Comparaison des performances de recalage (valeurs moyennes de TRE) à travers une évaluation séquentielle, et pour différentes expériences. 4 scénarios sont comparés et plusieurs types d’ajout des antécédents sont évalués.

Évaluation séquentielle	TRE ( $\downarrow$ ) : $\mu \pm \sigma$ (mm)			
	TRE avant recalage : 27.47 ( $\pm$ 16,68)			
	Scenario 1 ( $I_t, I_{t-1}$ )	Scenario 2 $+I_{flot}$	Scenario 3 $+T_{suivi}$	Scenario 4 $+I_{flot} + T_{suivi}$
Sans $T_{prior}$	13.04 ( $\pm$ 12.02)	11.12 ( $\pm$ 11.56)	13.79 ( $\pm$ 11.66)	11.45 ( $\pm$ 11.56)
Théorique avec $T_{prior} = T_{t-1}$	3.86 ( $\pm$ 5.17)	3.88 ( $\pm$ 6.23)	0.51 ( $\pm$ 0.57)	0.53 ( $\pm$ 0.59)
Complète avec $T_{prior} = \widehat{T}_{t-1}$	21.64 ( $\pm$ 15.32)	29.91 ( $\pm$ 15.82)	5.72 ( $\pm$ 2.40)	10.75 ( $\pm$ 5.96)
Par portion avec $T_{prior} = \widehat{T}_{t-1}$ + réajustement	9.56 ( $\pm$ 9.60)	9.93 ( $\pm$ 10.17)	<b>1.91 (<math>\pm</math> 1.26)</b>	2.3 ( $\pm$ 1.60)
Contrôle Que $T_{prior}$ ou Que $T_{prior} * T_{suivi}$	10.75 ( $\pm$ 10.47)	11.35 ( $\pm$ 10.96)	2.11 ( $\pm$ 3.09)	2,33 ( $\pm$ 3.31)

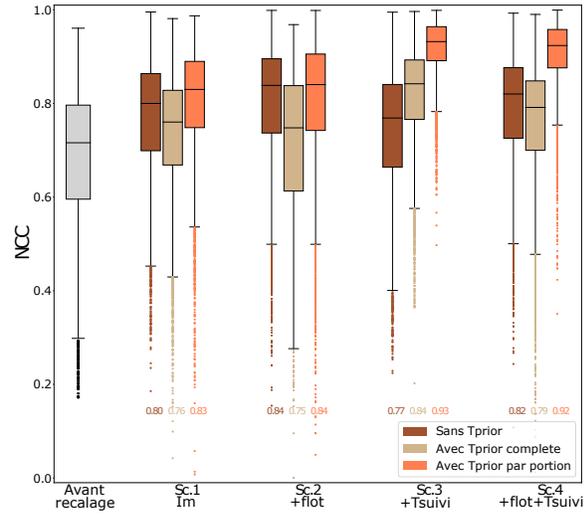
Tout d’abord, toutes les expériences et tous les scénarios montrent un bon apprentissage de la tâche de recalage, à travers une réduction plus ou moins importante de l’erreur initiale (avant recalage). Ceci est prometteur, d’autant plus que la valeur de TRE avant recalage révèle de très grandes erreurs initiales ( $\approx$  30 mm). Des tests statistiques de Wilcoxon (non-paramétriques, seuil = 0.05) ont été réalisés sur chaque couple d’expériences et de scénarios possibles. Ils ont tous mis en évidence des distributions significativement différentes, excepté entre les scénarios 1 et 2 de l’évaluation *complète* et entre les scénarios 3 et 4 de l’évaluation *théorique*.

L’expérience *sans*  $T_{prior}$  démontre un premier niveau d’apprentissage du réseau où l’erreur de recalage peut être corrigée jusqu’à 11 mm. Dans cette expérience, les scénarios combinant l’entrée de flot optique (scénario 2 et 4) semblent obtenir les meilleures performances.

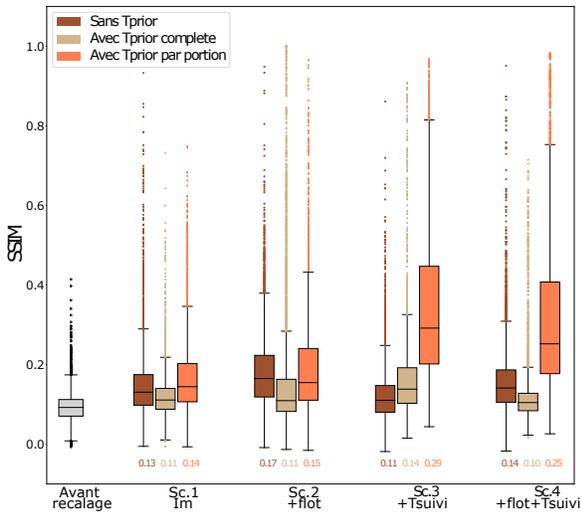
Les deux expériences *avec*  $T_{prior}$ , avec une évaluation *complète* ou *par portion*, obtiennent toutes deux de meilleurs résultats que les expériences *sans*  $T_{prior}$ , notamment pour les scénarios 3 et 4. En revanche, pour les scénarios 1 et 2, l’évaluation *complète* offre des performances bien inférieures à



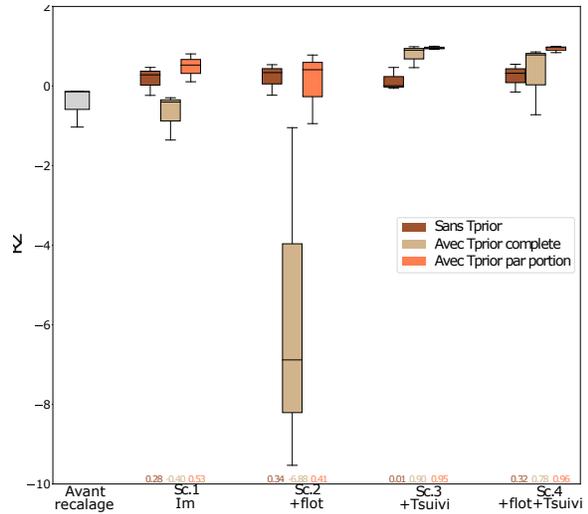
(a) Métrique TRE



(b) Métrique NCC



(c) Métrique SSIM



(d) Métrique  $R^2$

FIGURE IV.2 – Comparaison des distributions des métriques d'évaluation de recalage, selon les 4 scénarios testés et selon les 3 types d'ajout de  $T_{prior}$ . Les valeurs médianes sont reportées.

celles *sans*  $T_{prior}$  mais qui peuvent être suffisamment améliorées grâce à des réajustements ponctuels de  $T_{prior}$  (à chaque pseudo site de biopsie) dans l'évaluation *par portion*. Globalement, l'ajout de  $T_{prior}$ , avec un minimum de réajustements pour certains scénarios, semble permettre d'améliorer systématiquement les performances du modèle.

L'expérience évaluée *par portion* est notamment comparée à son évaluation *contrôle* associée, où un test statistique montre des distributions significativement différentes pour chacun des scénarios envisagés. La contribution à la fois du réseau et de toutes les entrées est donc vérifiée, et résulte en de meilleures performances que celles obtenues avec un traitement exclusif des entrées positionnelles (*contrôle*).

Enfin, l'évaluation séquentielle *théorique* reflète les résultats idéaux que le modèle pourrait espérer obtenir, avec des prédictions de  $T_{prior}$  idéales tout au long d'une trajectoire. Les valeurs minimales obtenues n'atteignent pas une valeur d'erreur parfaitement nulle et reflètent notamment l'impact de l'ajout d'un bruit aléatoire au niveau des entrées additionnelles de  $T_{prior}$  et  $T_{suivi}$ .

Au final, le meilleur résultat mesuré durant l'évaluation séquentielle par portion proposée, est obtenu pour le scénario 3 (+ $T_{suivi}$ ), avec une réduction de l'erreur de 27.47 mm ( $\pm 16,68$ ) à 1.91 mm ( $\pm 1.26$ ). La durée moyenne de prédiction de notre réseau est d'environ  $30 \pm 10$  ms.

#### IV.4.1.2 Analyse temporelle

Une analyse temporelle (sur une trajectoire complète) est complémentaire à l'analyse moyennée (sur toutes les données test) précédente. Les tendances et conclusions concernant les expériences et scénarios restent les mêmes que précédemment mais mettent mieux en évidence le comportement de chaque paramètre au cours de la trajectoire. Les Figures IV.3 et IV.4 présentent respectivement l'évolution des paramètres de transformation et l'évolution de l'erreur mesurée, au cours d'une trajectoire typique d'un examen de biopsie. Par ailleurs, la Table IV.4 illustre les mesures statistiques de  $R^2$  et de MAE, entre les transformations de vérités terrain et les transformations prédites, pour chaque angle au cours de la trajectoire.

TABLE IV.4 – Comparaison des performances de recalage (valeurs moyennes de  $R^2$  et MAE) à travers une évaluation séquentielle, et pour différentes expériences. 4 scénarios sont comparés et plusieurs types d'ajout des antécédents sont évalués.

Évaluation séquentielle	$R^2$ ( $\uparrow$ ) : Valeur moyenne pour $(\theta_x, \theta_y, \theta_z)$ en %			
	MAE ( $\downarrow$ ) : Valeur moyenne pour $[\theta_x, \theta_y, \theta_z]$ en $^\circ$			
	Scenario 1 $(I_t, I_{t-1})$	Scenario 2 $+I_{flot}$	Scenario 3 $+T_{suivi}$	Scenario 4 $+I_{flot} + T_{suivi}$
<b>Sans</b> $T_{prior}$	(-0.23, 0.47, 0.28) [9.99, 29.8, 5.80]	(-0.23, 0.54, 0.34) [9.58, 25.4, 4.8]	(-0.06, 0.46, 0.07) [9.75, 31.1, 7.63]	(-0.15, 0.55, 0.32) [9.15, 26.1, 5.67]
<b>Complète</b> avec $T_{prior} = \widehat{T}_{t-1}$	(-0.40, -0.29, -1.36) [10.7, 65.3, 9.88]	(-9.53, -1.04, -6.88) [24.8, 88.9, 16.7]	(0.90, 0.98, 0.46) [2.88, 8.49, 5.89]	(0.78, 0.85, -0.72) [4.11, 21.8, 11.1]
<b>Par portion</b> avec $T_{prior} = \widehat{T}_{t-1}$ + réajustement	(0.11, 0.81, 0.53) [8.10, 17.7, 4.81]	(-0.94, 0.77, 0.41) [11.3, 15.9, 4.75]	<b>(0.95, 0.99, 0.92)</b> <b>[1.93, 1.29, 1.96]</b>	(0.96, 0.99, 0.84) [1.49, 2.65, 2.97]

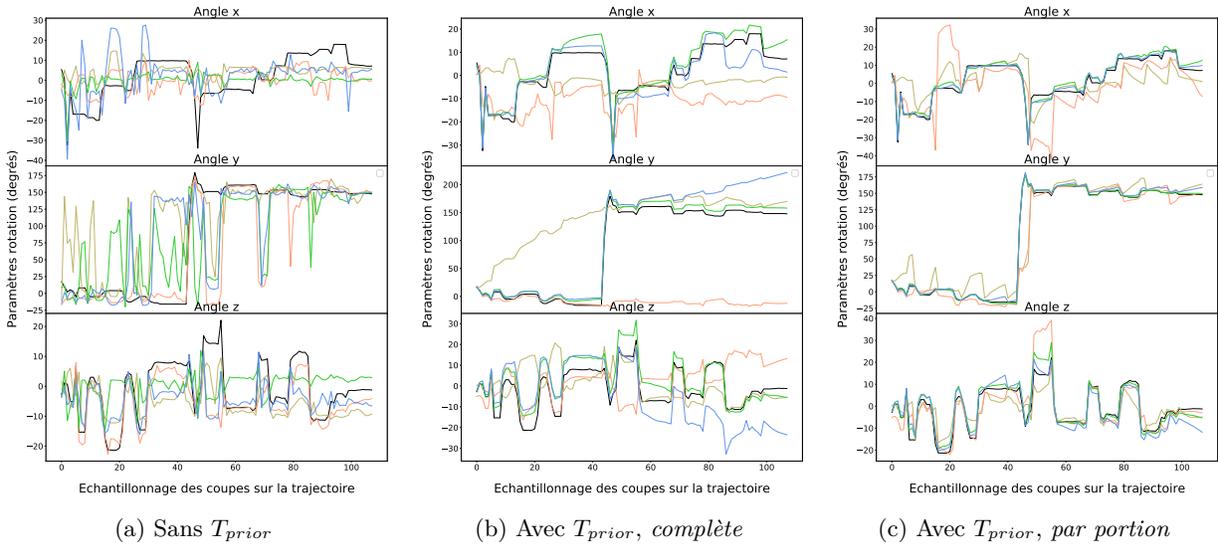


FIGURE IV.3 – Analyse temporelle de l'évolution des paramètres le long d'une trajectoire : Paramètres de vérités terrains (courbe noire) et paramètres prédits (courbes colorées). Scénario 1 (jaune), Scénario 2 (orange), Scénario 3 (vert), Scénario 4 (bleu).

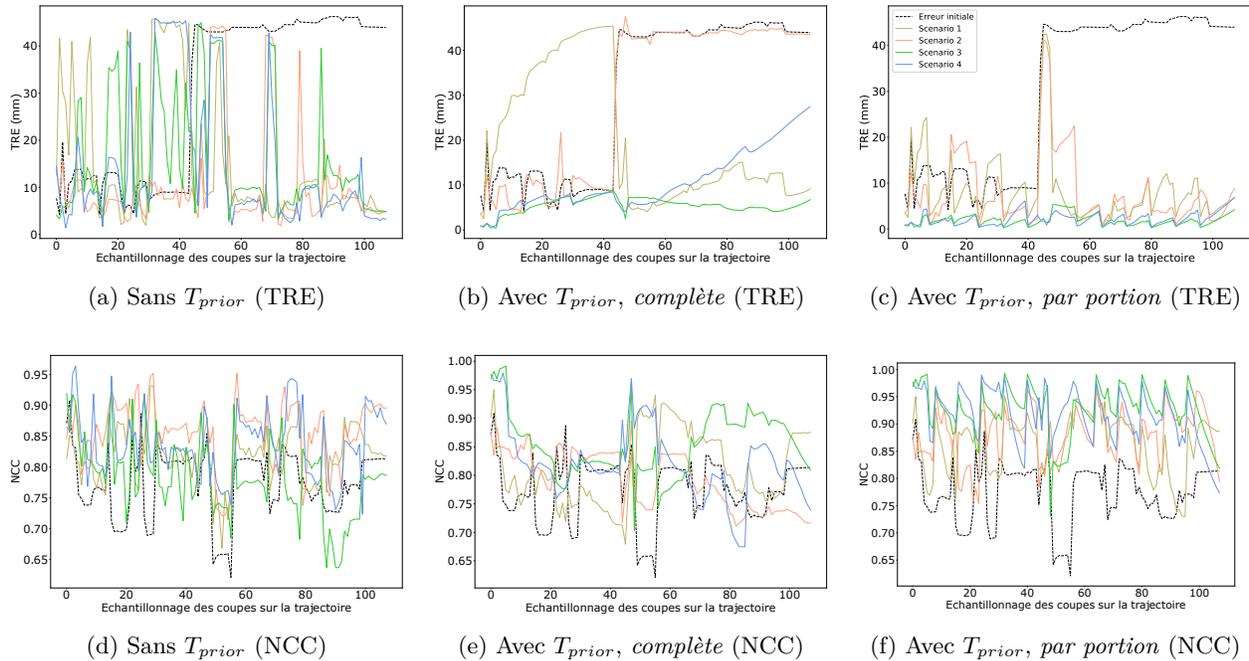


FIGURE IV.4 – Analyse temporelle de l'évolution des erreurs le long d'une trajectoire : erreurs initiales (courbe noire) et corrigées après application du recalage prédit selon les scénarios (courbes colorées). Scénario 1 (jaune), Scénario 2 (orange), Scénario 3 (vert), Scénario 4 (bleu).

L'expérience *Sans  $T_{prior}$*  illustre une reconstruction correcte de la trajectoire et une atténuation de l'erreur constante au cours de la trajectoire. La reconstruction de l'angle  $\theta_y$  (rotation de la sonde sur elle-même) semble plus difficile avec des oscillations fréquentes entre les deux valeurs 0 ou 160°. Cela correspond bien aux deux positions de sonde possibles pour la rotation hors-plan entre la moitié droite ou gauche de la prostate. Pour le scénario 2, le plus performant sur cette expérience, la différence moyenne angulaire est d'environ (9.58°, 25.4°, 4.8°) pour chaque angle. Le score du modèle est d'environ (-23%, 54%, 34%) et reflète une mauvaise cohérence de prédiction du modèle, notamment pour l'angle  $\theta_x$  (rotation base-apex).

L'évaluation séquentielle *complète* montre un phénomène de dérive au cours de la trajectoire, pour tous les scénarios. Cet impact est encore plus marqué pour les scénarios 1 et 2 avec une dérive angulaire observée jusqu'à (24.8°, 88.9°, 16.7°). Enfin, lors de l'évaluation séquentielle *par portion*, les 12 réajustements ponctuels (à chaque pseudo site de biopsie) sont bien visibles lors de la reconstruction de la trajectoire. Ils démontrent leur capacité à contenir le phénomène de dérive. Cette dérive reste cependant présente entre deux instants de correction. Le meilleur profil de trajectoire est obtenu pour le scénario 3, avec une différence moyenne angulaire d'environ (1.93°, 1.29°, 1.96°) pour chaque angle et avec un score du modèle de (95%, 99%, 92%).

#### IV.4.1.3 Analyse qualitative

Pour finir, les résultats qualitatifs sont également représentés dans La Figure IV.5 suivante. Cette visualisation des résultats est complémentaire aux analyses quantitatives précédentes et permet de mieux interpréter l'impact des erreurs directement sur l'organe.

L'exemple choisi présente un cas moyen lors d'une évaluation séquentielle par portion. Les résultats obtenus pour les scénarios 1 et 3 *avec  $T_{prior}$*  présentent des images très similaires à l'image courante et une superposition globale satisfaisante du maillage de prostate. En revanche, les images obtenues pour les expériences *sans  $T_{prior}$*  témoignent d'une mauvaise localisation de la coupe US par rapport au volume de référence.

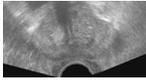
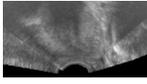
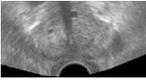
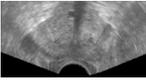
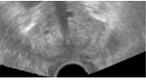
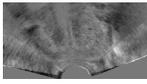
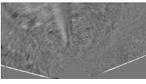
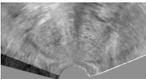
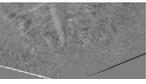
	Image courante $I_t$	Sans $T_{prior}$ (Sc.1)	Avec $T_{prior}$ (Sc.1)	Sans $T_{prior}$ (Sc.3)	Avec $T_{prior}$ (Sc.3)
Métriques		TRE = 17.25 NCC = 0.80 SSIM = 0.14	TRE = 2.21 NCC = 0.93 SSIM = 0.39	TRE = 14.6 NCC = 0.89 SSIM = 0.23	TRE = 2.17 NCC = 0.94 SSIM = 0.43
Images					
Différences de pixels					
Superposition de maillage					

FIGURE IV.5 – Exemple illustratif de la qualité du recalage des images US. Les images rééchantillonnées à la position prédite sont comparées à l'image courante par la différence de pixels. La superposition des maillages de la prostate (référence en vert, prédite en pointillés rouges) est également illustrée.

### IV.4.2 Étude d'interprétabilité : biais des zones noires

Plusieurs exemples de cartes d'activation sont donnés sur la Figure IV.6. Ces résultats ont été obtenus en considérant comme classe d'intérêt la prédiction de l'angle  $\theta x$ , et uniquement pour le scénario 1 de l'expérience *Sans  $T_{prior}$* . Les autres scénarios et les autres classes d'angles, résultent en les mêmes tendances et conclusions, mais ne seront pas davantage discutés.

La Figure IV.6 présente plusieurs exemples de coupes rééchantillonnées (issues de la base de données du simulateur), avec la présence de « zones noires » d'orientation variable selon le cône échographique intersecté. Globalement, le réseau semble se concentrer sur des régions de l'image cohérentes avec les régions présentant une différence de pixels importante. Sur cette base de données, rappelons que la fréquence d'échantillonnage est de 150 ms et permet d'obtenir des images successives proches. Nous remarquons cependant que les attentions se portent sur des zones vastes dont le côté varie selon l'orientation de l'image. Cela peut faussement contribuer à la prédiction de la localisation par rapport à l'espace de référence.

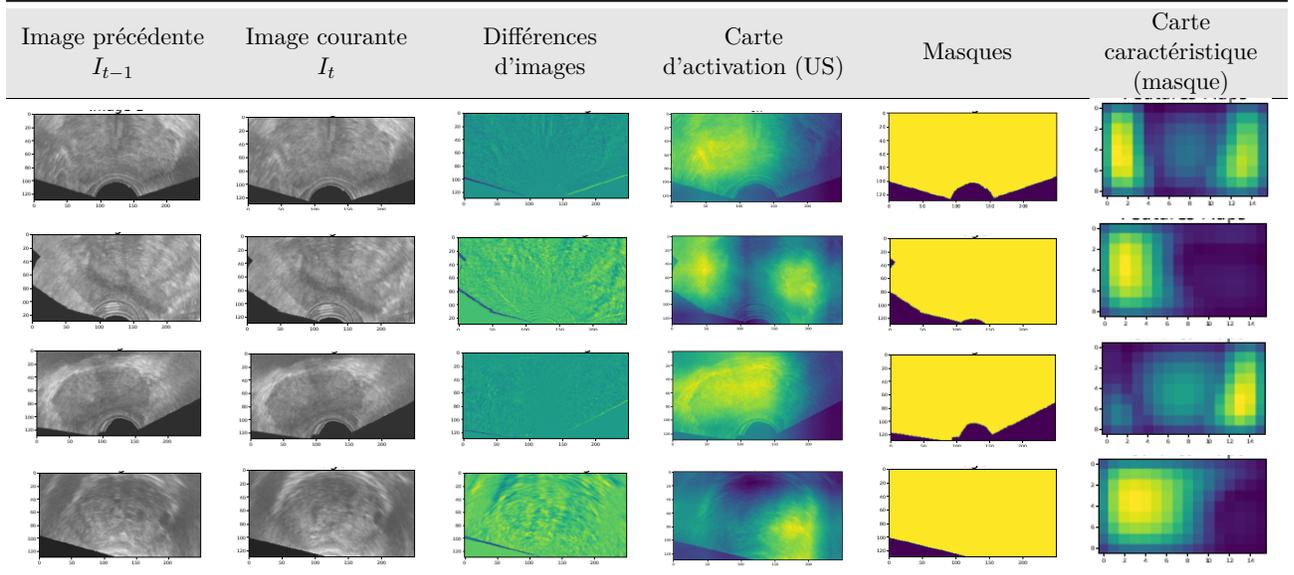


FIGURE IV.6 – Plusieurs exemples de visualisation des cartes d'activation et des cartes caractéristiques moyennes obtenues par le réseau : pour l'expérience *Sans  $T_{prior}$*  et le scénario 1 et considérant uniquement la prédiction de l'angle  $\theta x$  comme la classe d'intérêt.

Pour vérifier l'attention que le réseau porte à l'orientation variable des zones noires, nous avons également utilisé des images d'entrée sous forme de simple masque noir/blanc, sans aucun contenu échographique, comme illustré dans les dernières colonnes de la Figure IV.6. Les résultats associés à cette expérience sur les masques binaires sont détaillés dans la Table IV.5 et la Figure IV.7.

On constate, avec surprise, une amélioration des résultats après recalage. En effet, le scénario 1, qui n'utilise pourtant que la paire de simples masques dépourvus de toute information échographique, réduit l'erreur moyenne jusqu'à environ 17.76 mm. Même si ce scénario n'atteint pas d'aussi bons résultats qu'avec les images échographiques complètes initiales (13.04 mm), cela démontre l'importance de l'information contenue dans ces cônes échographiques d'orientation variables et non

réalistes. Cela se confirme notamment sur les cartes d’activation où l’intérêt accordé par le réseau se concentre essentiellement sur l’orientation du bord de rééchantillonnage, pouvant donner des indications sur la position absolue de la coupe au sein du volume de référence. On peut donc conclure que cette base de données, liée à la génération de données du simulateur, inclut un biais d’apprentissage.

Ces mêmes conclusions sont observées pour les scénarios avec  $T_{prior}$ , lors d’une évaluation séquentielle *par portion*. On remarque dans ces expériences, que le réseau testé sur les masques obtient des performances presque similaires à celles obtenues avec images. L’information absolue contenue dans l’entrée  $T_{prior}$  semble suffisamment riche, d’autant plus lorsque combinée avec  $T_{suivi}$  (Scénario 3), pour que le réseau ne soit pas perturbé par le changement d’images d’entrée. Cela confirme alors l’importance et l’utilité de l’ajout de  $T_{prior}$ , tout en alertant sur une potentielle pondération excessive de cette entrée au détriment des images.

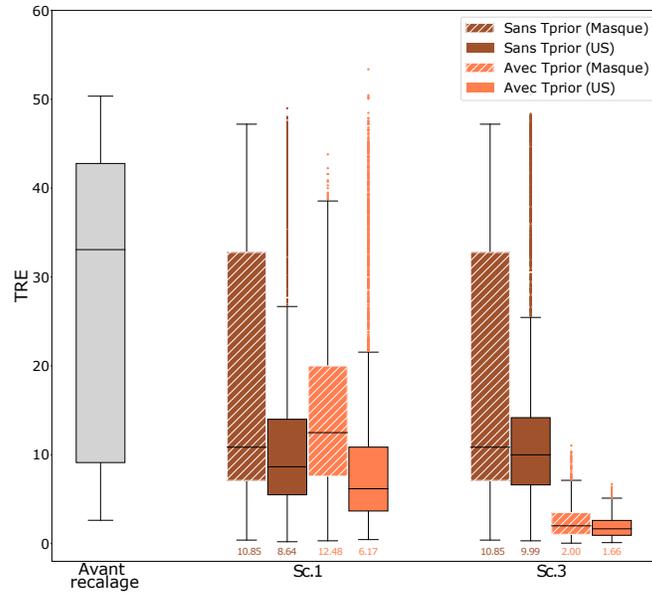


FIGURE IV.7 – Comparaison des distributions de TRE pour plusieurs expériences (Avec/Sans  $T_{prior}$ ) et selon les différentes images d’entrées : images échographiques initiales (US) ou images masquées (noir/blanc).

TABLE IV.5 – Comparaison des performances de recalage (TRE) sur des entrées d’images échographiques (US) ou images masquées (noir/blanc). Expériences sans  $T_{prior}$  et avec  $T_{prior}$  évalués *par portion*.

Évaluation séquentielle	TRE avant recalage : 26.1 ( $\pm 16,0$ )	
	Scenario 1 ( $I_t, I_{t-1}$ )	Scenario 3 ( $+T_{suivi}$ )
Sans $T_{prior}$		
Images (US)	13.04 ( $\pm 12.02$ )	13.79 ( $\pm 11.66$ )
Masques (noir/blanc)	18.16 ( $\pm 14.11$ )	18.25 ( $\pm 14.55$ )
Avec $T_{prior}$		
Images (US)	9.56 ( $\pm 9.60$ )	1.91 ( $\pm 1.26$ )
Masques (noir/blanc)	14.49 ( $\pm 8.69$ )	2.43 ( $\pm 1.78$ )

### IV.4.3 Variation des caractéristiques du jeu de données

#### IV.4.3.1 Fréquence d'échantillonnage

Plusieurs fréquences d'échantillonnage ont été comparées, résultant en des images espacées de 50, 150, 500 ou 1000 ms. La comparaison des distributions est illustrée dans la Figure IV.8.

Les meilleures performances sont obtenues pour les plus grandes fréquences d'échantillonnage : 50 ms (20 Hz) et 150 ms (6.6 Hz). Avec l'association progressive des informations de déplacements absolus de la coupe précédente (expérience *Avec  $T_{prior}$* ) et/ou de déplacements relatifs entre les deux images ( $T_{suivi}$  dans le Scénario 3), la différence de performance entre toutes les valeurs de fréquences testées est cependant de moins en moins importante. Les distributions des expériences entre 50 et 150 ms et entre 500 et 1000 ms ne sont d'ailleurs plus statistiquement différentes (p-value de 0.0508) pour le scénario 3 *Avec  $T_{prior}$* . Cela reflète l'importance et la complémentarité des informations d'a priori positionnelles ( $T_{prior}$  et  $T_{suivi}$ ) en plus de la paire d'images.

Seule l'expérience *Sans  $T_{prior}$* , pour le scénario 1, donne de meilleures performances pour un pas d'échantillonnage de 150 ms. L'explication que nous donnons est que ce scénario, ne disposant que du contenu des deux images d'entrée, a besoin d'images suffisamment distantes pour rendre compte d'un minimum de différences iconiques, plus facilement interprétable pour retrouver des informations positionnelles.

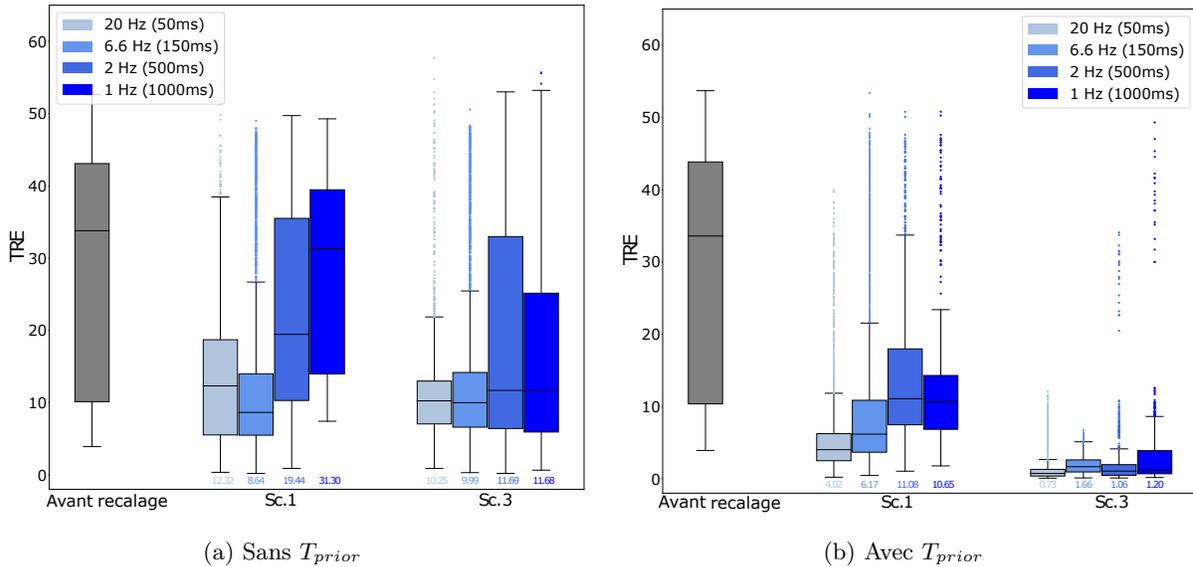


FIGURE IV.8 – Comparaison des distributions de TRE pour plusieurs fréquences d'échantillonnage entre les images successives : 50, 150, 500, ou 1000 ms. Le scénarios 1 ( $I_t, I_{t-1}$ ) et scénario 3 ( $+T_{suivi}$ ) sont testés sans et avec  $T_{prior}$  (*par portion*). Les valeurs médianes sont également reportées.

### IV.4.3.2 Niveaux d'expertises

Enfin, nous avons évalué la capacité de généralisation de notre modèle selon différents niveaux d'expertise du praticien. Les résultats sont résumés dans la Figure IV.9. On constate tout d'abord, que les erreurs après recalage sont toutes inférieures pour les *Manip Expert* que pour les *Manip Débutant*, alors qu'initialement (avant recalage) l'erreur est plus importante pour les *Manip Expert*. Il semble alors que ces deux types de trajectoires (réalisées selon l'expertise du praticien) ne présentent pas exactement le même niveau de difficulté pour la tâche de recalage alors qu'elles sont censées suivre le même schéma stéréotypé.

Pour tous les scénarios testés et sur les deux types de trajectoires, on peut observer des performances assez similaires entre les deux modèles. Les tendances et conclusions concernant les scénarios restent les mêmes que celles observées précédemment (voir section IV.4.1.1) et sont identiques pour les deux modèles. Néanmoins, on constate dans chaque cas, que c'est le modèle ayant été entraîné avec les mêmes types de trajectoires que celles évaluées qui performe le mieux : *Modèle Expert* pour *Manip Expert* et *Modèle Débutant* pour *Manip Débutant*. Les modèles semblent alors spécifiques à l'expertise du praticien et au profil des trajectoires utilisées durant l'entraînement.

Les deux modèles ne présentent pas les mêmes capacités de généralisation. On constate que le *Modèle Débutant* généralise très bien sur les *Manip Expert* où il obtient des performances similaires à celles du *Modèle Expert*. À l'inverse, le *Modèle Expert* généralise moins bien sur les *Manip Débutant*, avec des erreurs plus élevées que le *Modèle Débutant*. Cela reflète également du niveau de difficulté de recalage différents entre ces différents profils des trajectoires et de leur impact sur les capacités d'apprentissage et de généralisation des modèles.

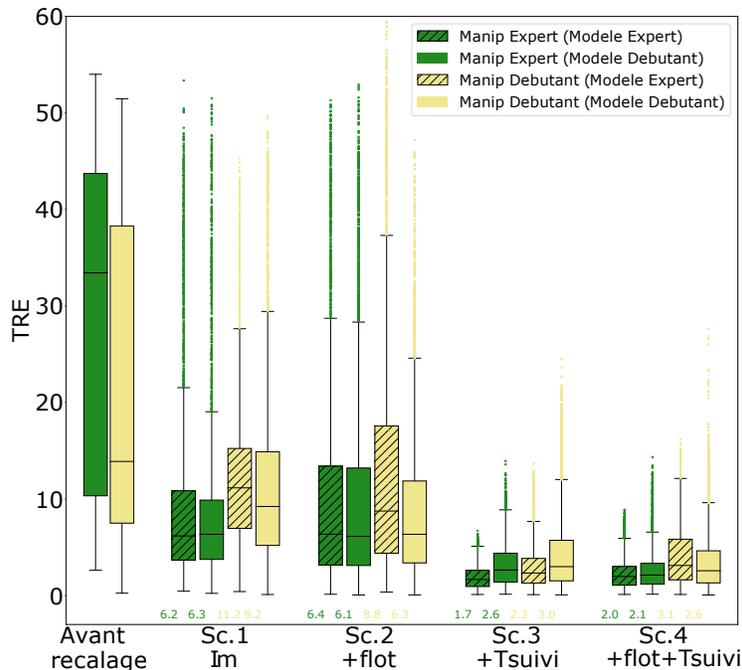


FIGURE IV.9 – Comparaison des distributions de TRE (et leur médiane) pour les 4 scénarios testés selon différents niveaux d'expertise du praticien participant à l'apprentissage du modèle ou à l'inférence.

## IV.5 Discussion

### IV.5.1 Discussion sur les contributions méthodologiques

#### IV.5.1.1 Différents scénarios testés

L'ajout du flot optique en données d'entrée, dans les scénarios 2 ou 4, semble favoriser l'apprentissage du réseau seulement dans les expériences *Sans  $T_{prior}$* , où on constate une faible amélioration des performances. Dans ces expériences, très peu d'information sur la localisation absolue de la coupe au sein du volume de référence est fournie, à l'exception dans le contenu iconique de la paire d'images d'entrée. Il est raisonnable de supposer que tout apport d'informations concernant le déplacement plan relatif entre les deux images, directement fournies et prétraitées par l'algorithme de flot optique, peut être plus facilement associé avec le contenu des images et améliorer la tâche de recalage.

En revanche, l'impact de l'ajout du flot optique n'est pas aussi bénéfique qu'espéré. Les visualisations des cartes d'attention ne révèlent pas de lien particulier entre les zones d'attention du modèle et les zones de la carte de flot de forte magnitude. Même si ce type d'entrée aide dans certains cas à la tâche de recalage, il est cependant difficile de conclure quant à sa bonne contribution générale, notamment lorsque il est couplé à d'autres types de données d'entrée, potentiellement plus bénéfiques. Par exemple, dans les expériences *Avec  $T_{prior}$* , l'ajout du flot optique ne semble plus pertinent, voir même au contraire pénalisant pour le bon apprentissage du réseau.

Cette mauvaise contribution générale du flot optique peut s'expliquer par le fait que les cartes du flot optique présentent des données de nature et de grandeur différentes des intensités de pixels. Elles peuvent alors ajouter de la complexité lors de la mise en relation avec les images d'entrées et lors de l'optimisation des poids. Ceci peut également s'expliquer par le fait que les déplacements simulés dans le plan de l'image sont très faibles, contrairement aux déplacements hors plan. Lors d'une acquisition de données avec des déformations anatomiques, avec notamment un écrasement de la prostate à la surface de la sonde dans le plan échographique, le flot optique pourrait alors s'avérer plus utile.

L'ajout du suivi de sonde ( $T_{suivi}$ ) en données d'entrée, dans les scénarios 3 ou 4, semble fortement favoriser l'apprentissage du réseau, notamment dans les expériences avec  $T_{prior}$ , où on constate une amélioration significative des performances de recalage. L'entrée  $T_{suivi}$ , donnant des informations sur le déplacement relatif qui a été réalisé depuis le précédent résultat de recalage (entrée  $T_{prior}$ ), semble être justement combinée et associée avec ce dernier par le réseau pour retrouver la position absolue de la coupe.

Plusieurs configurations d'ajout du  $T_{suivi}$  ont été testées et ont notamment suggéré de meilleures performances lorsque les entrées  $T_{suivi}$  et  $T_{prior}$  sont dissociées en deux transformations d'entrée différentes plutôt qu'en une seule composition d'entrée des transformées ( $T_{suivi} * T_{prior}$ ). Les informations de déplacements relatifs et absolus ayant leur propre nature, référentiel, et avantages respectifs, la séparation de celles-ci favorise leur association distincte avec le contenu des images. Sans ajout du  $T_{prior}$ , en revanche, les scénarios 1 et 3 se valent.

### IV.5.1.2 Ajout de $T_{prior}$

L'ajout des  $T_{prior}$  (issus des labels ou des prédictions) semble être la contribution la plus bénéfique permettant l'amélioration de la qualité du recalage. Ce bénéfice s'explique notamment par le fait que des informations concernant la localisation absolue sont nécessaires au réseau qui doit estimer la position d'une coupe dans un volume de référence dont il n'a aucune information. Or, excepté le contexte très limité présent dans les images 2D en entrée du réseau, le reste des informations d'entrée ne concerne que des déplacements relatifs autour de la coupe. L'ajout du  $T_{prior}$  semble faciliter l'apprentissage et se rapproche de l'idée d'une initialisation qui favorise la convergence dans un cadre de recalage classique. De plus, une telle information directement ajoutée sous une forme similaire à l'encodage du vecteur de sortie  $T_t$ , peut être facilement interprétée par le réseau.

Cependant, lors d'une évaluation séquentielle réaliste à partir d'antécédents  $T_{prior}$  successivement prédits par le réseau, une accumulation d'erreur est observée. La dérive observée vient de l'importance donnée par le réseau aux entrées  $T_{prior}$  comparativement aux autres données d'entrées. Les résultats obtenus, bien que comportant une dérive, restent généralement meilleurs que les prédictions des modèles sans  $T_{prior}$  et nous confortent dans l'idée que cette entrée est importante, malgré la dérive observée. De plus, cette accumulation d'erreur peut être corrigée à des moments clés, en simulant une correction à chaque nouvelle biopsie, qui suffit alors à contenir l'essentiel du phénomène de dérive.

Pour finir, la comparaison avec l'évaluation de *contrôle*, a permis de vérifier la bonne contribution à la fois du réseau et du reste des entrées, et d'exclure l'hypothèse d'une utilisation exclusive de l'entrée additionnelle  $T_{prior}$ . En effet, la transformation d'entrée  $T_{prior}$  peut être très proche de la transformation  $T_t$  à estimer (seulement 150 ms d'écart et peu de déplacement), notamment lorsque combinée avec le suivi relatif  $T_{suivi}$ . Cependant, il est important de remarquer que l'écart des performances entre ce scénario *contrôle* et le meilleur scénario (sc. 3) n'est pas si conséquent, ce qui nous encourage à veiller à la bonne pondération de cette entrée par la suite. Des expériences supplémentaires, avec différentes valeurs de bruit ajouté au  $T_{prior}$  ( $\pm 0^\circ$ ,  $\pm 2^\circ$ ,  $\pm 5^\circ$ ), n'ont cependant pas démontré de meilleure performance ou de meilleure généralisation au moment de l'analyse séquentielle.

## IV.5.2 Discussion sur les contributions expérimentales

### IV.5.2.1 Validité de la génération de données

Notre génération de données, à partir de l'utilisation du simulateur de biopsie, génère des zones noires, qui diffèrent selon l'orientation de la coupe rééchantillonnée au sein du volume de référence. Les images ainsi obtenues ne représentent pas la réalité du cône échographique fixe qui est obtenu sur des images issues du guidage échographique. Ces zones noires semblent faciliter la tâche de recalage, et biaisent ainsi nos données d'apprentissage. Parmi les études de la littérature produisant de telles zones noires, très peu font mention du biais associé ou ont cherché à quantifier son impact [69]. Nous démontrons cependant que ces zones entraînent un biais conséquent, pouvant favoriser l'apprentissage du réseau et empêcher la généralisation du réseau dans des conditions cliniques réalistes. C'est pour nous affranchir de ces zones noires que nous avons souhaité mettre en place de nouveaux procédés de simulation (déjà décrits en section III.1.3).

Enfin, les expériences ont révélé de meilleures performances du modèle, pour des bases de données générées à partir de grande fréquence d'échantillonnage et avec des trajectoires issues de plusieurs niveaux d'expertise. Ces conditions permettent une meilleure généralisation du modèle à des trajectoires représentatives d'autres niveaux d'expertise. Ces observations sont à garder en tête en vue d'une généralisation clinique même si ces simulations n'ont pas été exploitées davantage du fait du biais des zones noires.

### IV.5.2.2 Évaluation séquentielle sur des trajectoires complètes

Pour établir une évaluation réaliste du déroulement clinique, nous avons mesuré les erreurs de recalage à travers une série de recalage successifs, réalisés tout au long d'une trajectoire complète. Cette évaluation est d'autant plus importante que certaines entrées, notamment  $T_{prior}$ , représentent les prédictions de résultats précédents et peuvent alors propager l'erreur au cours du temps.

À la différence de certains travaux de la littérature partageant des problématiques similaires [56], nous avons évalué l'impact d'une potentielle accumulation d'erreur et nous avons démontré qu'il était significatif sur les performances d'un modèle. Une telle dérive peut être corrigée en utilisant plusieurs réajustements à des moments clés (les 12 pseudo biopsies), où les valeurs de  $T_{prior}$  sont alors réinitialisées à la valeur vérité terrain (voir section III.2.2.2). Ces réajustements sont inspirés des recalages 3D/3D intégrés aux plateformes cliniques [17] dont l'accès et l'utilisation semblent tout à fait envisageable dans une assistance à la navigation par portion telle que la nôtre.

### IV.5.2.3 Réalisme et faisabilité clinique

Notre méthode est facilement reproductible grâce à une architecture de réseau simple et facile à entraîner, et grâce à des entrées simples. Cependant, certaines de ces entrées peuvent être plus spécifiques, comme le suivi des sondes et les réajustements de recalage.

Alors que la plupart des études ont tendance à privilégier des travaux sans utilisation de capteur inertiel supplémentaire, nous concluons cependant qu'un tel ajout peut être très utile. Ces dispositifs de suivi sont petits et bon marché et devraient pouvoir alors facilement être intégrés dans une application clinique. Pour ce qui concerne les réajustements de dérive, ils reposent sur un algorithme de recalage 3D/3D, disponible sur plusieurs plateformes commerciales cliniques ou

pouvant être estimés à partir de la méthode détaillée dans [17] ou à partir d'autres méthodes de recalage classiques.

### IV.5.3 Conclusions sur ce premier modèle

#### IV.5.3.1 Conclusions sur les performances obtenues

Il semble que la meilleure configuration soit obtenue en combinant les entrées d'images ( $I_t, I_{t-1}$ ), de prédiction de la position précédente ( $T_{prior}$ ), et de suivi de la sonde ( $T_{suivi}$ ). Cela démontre qu'en considérant des informations liées à la trajectoire réalisée, via des informations positionnelles notamment, on peut faciliter la tâche de recalage durant la navigation.

Cette configuration permet de corriger des erreurs de recalage partant de 26 mm jusqu'à 1.91 mm ( $\pm 1.26$  mm). Les résultats ainsi obtenus sont prometteurs et respectent nos attentes cliniques (TRE  $\leq 2.5$  mm et calcul  $\leq 50$ ms). Les résultats sont d'autant plus prometteurs que la valeur de TRE avant recalage révèle de très grandes erreurs initiales ( $\approx 30$  mm), de l'ordre du diamètre de la prostate, qui sont rarement testées dans la littérature et à partir desquelles les méthodes classiques de recalage ne parviennent généralement pas à converger.

Nous rappelons également que les données de test reposent sur : (i) des trajectoires réalistes et complexes de biopsie simulées par des praticiens experts, (ii) des données non prétraitées et non initialisées (contrairement à beaucoup d'études de la littérature), et (iii) un protocole d'évaluation rigoureux et réaliste cliniquement. Toutes ces considérations renforcent la qualité des résultats obtenus et démontrent une bonne capacité de généralisation de notre modèle.

#### IV.5.3.2 Comparaison avec la littérature

Parmi les études référencées autour des problématiques de recalage 2D(US)/3D(US) pour le guidage de procédure autour de la prostate (voir section II.2.2), rappelons que très peu [20,52,54–57] partagent des objectifs, des méthodes ou des attentes clinique proches des nôtres. A l'heure de cette première méthode, nous ne pouvons en réalité nous comparer qu'à certains de ces travaux [20,52,54]. La Table IV.6 établit un descriptif plus détaillé de cette comparaison avec la méthode proposée dans ce chapitre (notée Dupuy et. al 2021). Les éléments en rouge reflètent les critères les plus limitants ne permettant pas une comparaison rigoureuse avec les autres études.

Pour une interprétation et une comparaison rigoureuse de notre étude, certains constats sont à garder à tête concernant notamment la qualité de l'évaluation, des données en jeu, ou encore des applications des études. Notre simulation des données et des trajectoires de biopsie via le simulateur, permet une représentation plus réaliste et plus complexe que la plupart des études référencées dans la littérature : les erreurs initiales sont plus importantes ( $\approx 30$  mm), les mouvements sont plus complexes avec de plus grandes amplitudes (3 rotations, entre 30-180°), la validation plus juste (TRE sur maillage vs. quelques fiducials), et la base de données clinique plus conséquente (100 patients). Bien que notre méthode utilise des protocoles d'évaluation souvent plus complexes que ceux de l'état de l'art, nos performances (précision et rapidité) sont supérieures.

Pour finir, même si nous nous sommes inspirés de certains éléments de l'étude [76], celle-ci s'intéresse à une problématique de reconstruction de volume, où le recalage est estimé entre deux coupes successives et présente donc des objectifs et des ordres de grandeurs différents des nôtres.

### IV.5.3.3 Conclusions sur les objectifs

Les conclusions en réponse aux trois objectifs principaux de cette première méthode proposée sont :

- Le réseau proposé est capable de prédire en temps-réel les paramètres de transformation pour une tâche de recalage 2D/3D. Compte tenu des résultats trouvés dans l'état-de-l'art, nos résultats sont comparables voire supérieurs. Ils respectent les spécifications cliniques attendues et démontrent qu'une telle méthode peut être utilisée dans le cadre d'un guidage continu durant une procédure de biopsie de prostate.
- Nos résultats soulignent la pertinence et l'efficacité d'exploiter les informations liées à la trajectoire, notamment à travers l'ajout des résultats de recalage précédents et du suivi de la sonde. Cette piste est donc à privilégier davantage en étudiant d'autres types d'informations liées au concept de « trajectoire ».
- Les protocoles d'évaluation proposés, avec l'utilisation de plusieurs métriques différentes et une évaluation séquentielle tout au long d'une trajectoire, semblent pertinents et adaptés pour la validation de la tâche de recalage 2D/3D. En revanche, le réalisme de notre premier procédé de génération de données (issu du simulateur Biopsym) est remis en question avec notamment la détection de biais de zones noires. Cette observation est la cause principale ayant motivé l'évolution et la mise en place de nouveaux procédés de génération des données (voir section III.1.3).

Il convient désormais d'étudier les autres types d'informations contextuelles de trajectoire, définis dans notre positionnement général de cette thèse. Le chapitre suivant s'intéresse notamment à l'étude de l'ajout d'informations « spatiales ». Il s'agira alors de combiner ces informations spatiales aux informations positionnelles déjà retenues ( $T_{prior}$ ,  $T_{suivi}$ ), pour déterminer la configuration permettant d'obtenir les performances de recalage optimales.

TABLE IV.6 – Comparaison de notre première méthode (notée Dupuy et. al, 2021) avec certaines études de recalage « 2D/3D » de la littérature établie au Chapitre II. Les éléments en rouge reflètent les critères limitants ne permettant pas une comparaison rigoureuse avec les autres études.

Auteur Année	Recalage	Organe	Transfo	Architecture & Apprentissage	Renvois aux classes d'analyse	Résultats	Évaluation Données & GT
Dupuy 2021 [103]	2D US / 3D US	Prostate	Rigide	CNN $\Rightarrow$ (doF) SL : MSE(doF)	II.4.1.1-Positionnelle II.4.2.1-Paramétrage II.4.2.2-Coordonnées II.4.5.1-Génération II.4.5.4-Déroulement	<b>TRE</b> : 1.91±1.26 mm initialement à 27.5 mm <b>MAE</b> : 1.93°/1.3°/1.9° <b>TIME</b> : 30 ± 10 ms	Maillage rigide 2D extraites de traj biopsie simulée par expert $\pm(30^\circ, 180^\circ, 20^\circ)$ : 57 000 images (100 patients)
Prevost 2018 [76]	2D US / 2D US	Bras Jambe (Reconstr.)	Rigide	CNN $\Rightarrow \delta$ (doF) SL : MSE(doF)	II.4.1.1-Positionnelle II.4.1.3-Séquentielle	<b>Final drift</b> : 11mm (5% de 200) <b>MAE</b> : 5mm, 1-3° (relatif)	2D extraites de 600 balayages rectilignes : 300 000 images (15 patients) GT localisateur optique
Gillies 2017 [20]	2D US / 3D US	Prostate	Rigide	Tradi. iconique Powell(NCC)	II.4.1.1-Positionnelle II.4.5.1-Génération	<b>TRE</b> : 1,4 ± 1.17mm plan 0.2 ± 0.3 mm hors plan 0.7±0.4 mm roll 0.8 ± 1.0° <b>TIME</b> : 57 ± 20 ms	Fiduciels (7 patients) validation sur fantômes
Beitone 2021 [52]	2D US / 3D US	Prostate	Rigide	(i) init. Prédiction (ii) Detection feature (iii) Match - Rejet (iv) Tradi. hybride ICP( dist(f, I) )	II.4.1.1-Positionnelle II.4.5.1-Métriques II.4.5.1-Génération II.4.5.4-Déroulement	<b>TRE</b> : 3,80±0.40 mm initialement à 8.56 mm <b>TIME</b> : 60-200 ms	Maillage rigide sur 900 volumes issus de 80 patients 2D extraites de simulation sur volume de $\pm 5^\circ$ et $\pm 7,5$ mm
Bhardwaj 2020 [54]	2D US / 3D IRM	Prostate	Non-rigide	(i) init. manuelle 3D IRM/3D US (ii) 2D-Unet $\Rightarrow$ Seg (iii) Tradi. iconique SGD( MSE(seg) ) (iv) 2x 2D-Unet $\Rightarrow$ DDF SL :DSC/ MSE(DDF)	II.4.1.1-Positionnelle II.4.1.4-Segmentation II.4.3.1-Hiérarchique II.4.5.2-Génération GT II.4.5.4-Déroulement	<b>TRE</b> : 2.99±1.51 mm initialement à 8.9±5.1 <b>TIME</b> : 112 ms (46+32+34 ms)	Fiduciels (12 patients) GT(DDF) calculée par le logiciel ITK 2D US (2000 coupes) enregistrées (vidéo, 10 Hz) et localisées

**Points clé**

- ❖ Nous avons développé un réseau de type CNN supervisé permettant d'estimer les paramètres de transformation, pour une tâche de recalage 2D US/ 3D US en temps-réel. Parmi différentes configurations testées, nous privilégions un encodage sous la forme d'angles d'Euler et une fonction de coût basée sur la MSE.
- ❖ En incorporant des informations liées à la trajectoire réalisée, à travers notamment les résultats du recalage précédent et du suivi de sonde, nous observons une amélioration significative de la qualité de notre recalage. Cela démontre le bénéfice des informations positionnelles proposées dans le cadre de ce travail.
- ❖ Nous obtenons des résultats meilleurs que ceux observés dans la littérature, tout en respectant les spécifications cliniques établies. Avec un protocole d'évaluation adapté, sur des tâches complexes et fidèles aux difficultés cliniques, une telle méthode semble pouvoir être utilisée dans le cadre d'un guidage continu durant une procédure de biopsie de prostate.
- ❖ Ces travaux ont fait l'objet d'une publication et de la présentation d'un poster à la conférence SPIE Medical Imaging 2021 [103].

---

## Recalage 2D/3D avec des informations spatiales

---

*Ce* chapitre présente une seconde évolution de méthode pour la réalisation du recalage 2D/3D durant la navigation de biopsie prostatique, qui s'intéresse particulièrement aux types d'informations « spatiales » liées à la trajectoire réalisée.

Les principaux objectifs de ce chapitre sont d'étudier l'apport d'informations spatiales et de les combiner aux informations positionnelles précédemment établies ( $T_{prior}$ ,  $T_{suivi}$ ). Plusieurs expériences ont été réalisées : selon l'ajout et la combinaison des données d'entrée ; selon les fonctions de coûts testées ; et selon les architectures utilisées. En nous comparant à des méthodes similaires de la littérature, nous vérifions que l'approche retenue permet d'obtenir des résultats cohérents et comparables tout en respectant les attentes cliniques. Nous concluons quant à nos contributions face à notre première méthode et face aux méthodes de la littérature.

---

## Sommaire

---

<b>V.1 Objectifs et contributions</b>	<b>113</b>
<b>V.2 Matériel &amp; Méthodes</b>	<b>114</b>
V.2.1 Approche proposée	114
V.2.1.1 Définitions et notations	114
V.2.1.2 Architecture du réseau	116
V.2.2 Apprentissage	117
V.2.2.1 Terme d'auto-pénalisation de similarité iconique (SSIM)	117
V.2.2.2 Combinaison et pondération de plusieurs termes	118
<b>V.3 Expériences et configurations</b>	<b>118</b>
V.3.1 Différentes formes d'informations spatiales	118
V.3.2 Différentes architectures	119
V.3.3 Jeux de données	120
V.3.4 Comparaison avec l'état-de-l'art	121
V.3.5 Paramètres d'apprentissage	121
<b>V.4 Résultats</b>	<b>122</b>
V.4.1 Différentes formes d'informations spatiales	122
V.4.2 Différentes architectures	125
V.4.3 Comparaison avec l'état-de-l'art	126
V.4.3.1 Réimplémentation et étude d'ablation	126
V.4.3.2 Comparaison quantitative	126
V.4.3.3 Comparaison qualitative	128
V.4.3.4 Comparaison temporelle	129
<b>V.5 Discussion</b>	<b>131</b>
V.5.1 Discussion sur les contributions méthodologiques	131
V.5.1.1 Formes d'informations spatiales	131
V.5.1.2 Confrontation entre les informations positionnelles et spatiales	132
V.5.1.3 Différentes architectures testées	133
V.5.2 Discussion sur les contributions expérimentales	134
V.5.2.1 Généralisation sur plusieurs niveaux de difficultés	134
V.5.2.2 Importance de l'évaluation séquentielle	134
V.5.3 Conclusions sur ce deuxième modèle	135
V.5.3.1 Conclusions sur les performances obtenues	135
V.5.3.2 Comparaison avec la littérature	135
V.5.3.3 Conclusions sur les objectifs	135

---

## V.1 Objectifs et contributions

Déterminer la localisation optimale d'une image courante  $I_t$  par rapport à un espace de référence, pour lequel aucune information n'est donnée au réseau, s'est avéré être une tâche complexe. Nous avons démontré précédemment qu'apporter un a priori positionnel ( $T_{prior}$ ) concernant la localisation absolue de l'image précédente vers ce volume de référence favorise grandement la tâche de localisation 2D/3D. Afin d'améliorer notre méthode de recalage 2D/3D en temps-réel, l'objectif de ce chapitre est d'exploiter davantage les informations contextuelles spatiales de l'espace de référence.

Alors que ces informations spatiales (coupes voisines, informations extraites du volume de référence) sont souvent utilisées dans des méthodes de recalage classiques, où elles permettent de comparer la ressemblance entre les deux espaces itérativement alignés, rappelons que très peu d'études par apprentissage profond ont réussi à exploiter ces concepts dans un cadre de recalage 2D/3D tout en s'affranchissant des problématiques occasionnées (voir section II.3.2). En effet, fournir les informations spatiales (2D et 3D) sous forme d'entrée entraîne généralement une mauvaise combinaison de ces entrées de dimensions différentes et une représentation déséquilibrée de leurs caractéristiques au sein du réseau (voir section II.4.1.2). De même, l'utilisation de pénalité spatiale dans la fonction de coût, n'est étudiée que dans très peu d'études et notamment lorsque l'espace de référence est présent en entrée et est de même dimension (voir section II.4.4.1).

Fortement motivés par l'idée de considérer des informations contextuelles spatiales permettant de mieux considérer la trajectoire réalisée au sein de l'espace de référence, les trois principaux objectifs de cette seconde méthode sont :

- D'étudier et de comparer plusieurs formes d'informations spatiales. Pour cela, nous considérons l'ajout d'entrées spatiales sous la forme : (i) d'empilement d'une ou plusieurs coupes voisines, (ii) du volume de référence entier, et (iii) d'une sous-partie de ce volume de référence. Nous proposons de modifier l'architecture du réseau convolutif précédent, afin de mieux intégrer et combiner les différentes entrées spatiales étudiées. L'ajout d'un terme de pénalité spatiale durant l'apprentissage sera également étudié. Parmi les combinaisons testées, nous établirons la configuration optimale permettant d'obtenir de meilleures performances de recalage.
- De comparer les résultats obtenus à ceux de notre méthode précédente et à ceux des études de la littérature. Cela permet d'une part de vérifier l'apport des informations spatiales combinées et/ou confrontées aux informations positionnelles (méthode précédente) et d'autre part, de comparer nos contributions méthodologiques et expérimentales par rapport à d'autres études similaires de la littérature.
- De valider la bonne généralisation de notre réseau aux tâches de recalage présentant des niveaux de difficultés et de réalisme croissants. Pour cela, nous nous concentrons sur la base de données provenant des simulations de balayage base-apex (décrite en section III.1.3.3). Cela permettra à la fois de valider le réalisme de ce procédé de génération de données et de mieux pouvoir comparer les résultats obtenus à ceux de la littérature.

La section suivante présente les différentes stratégies d'ajout d'informations d'entrées spatiales et de pénalisation spatiale. La nouvelle architecture proposée ainsi que l'apprentissage du réseau seront également décrits.

## V.2 Matériel & Méthodes

### V.2.1 Approche proposée

La Figure V.1 illustre la nouvelle méthode proposée, qui se compose des éléments suivants :

- une branche principale avec une architecture inspirée de l'étude Guo et. al [56] (bloc 1) ;
- des entrées positionnelles retenues de la méthode précédente en section IV.5.3 (bloc 2) ;
- différentes formes d'entrées spatiales à tester (bloc 3) ;
- et d'un terme de pénalisation spatiale dans la fonction de coût (bloc 4).

La transformation à estimer à l'instant  $t$  est désignée par  $\hat{T}_t$  et se compose des  $\#$  paramètres de transformation selon l'encodage choisi dans nos jeux de données (voir section V.3.5). La partie suivante décrit plus en détails ces différents blocs de travail.

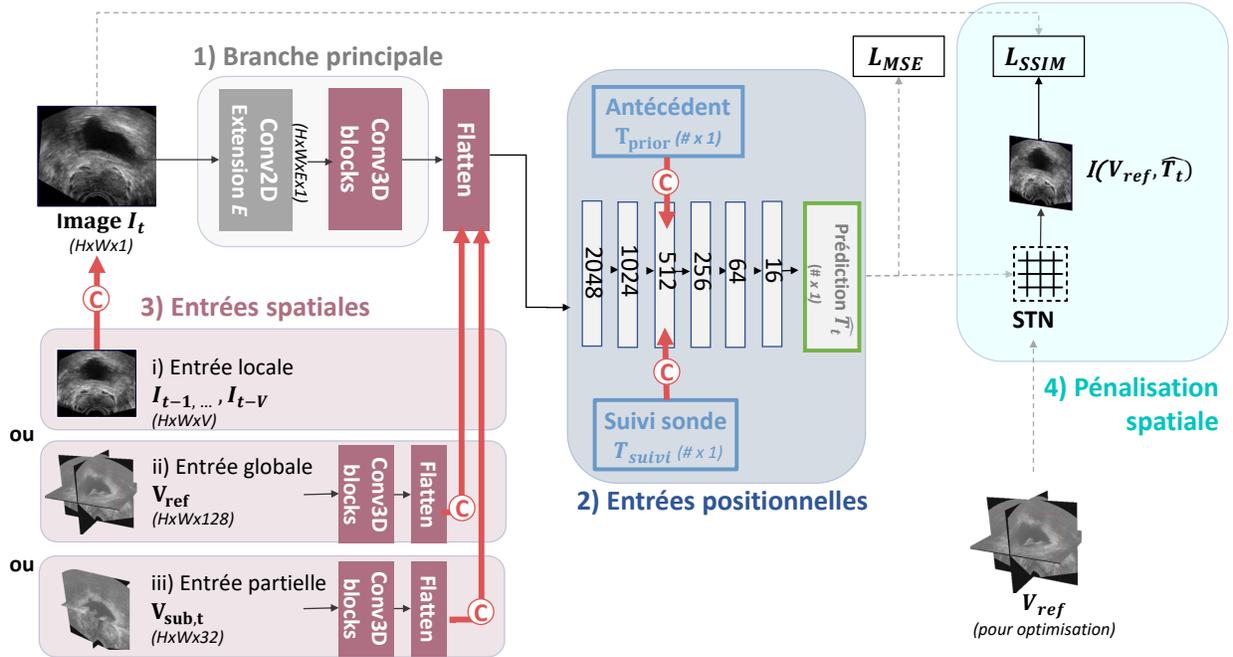


FIGURE V.1 – Architecture de réseau proposée avec la combinaison des différents blocs : Branche principale avec une nouvelle architecture 3D (bloc 1), les entrées positionnelles retenues (bloc 2), les différentes entrées spatiales à tester (bloc 3), et un terme de pénalisation spatiale dans la fonction de coût (bloc 4).

#### V.2.1.1 Définitions et notations

- La branche principale : L'entrée principale de ce réseau est l'image courante  $I_t$ , recadrée à la dimension  $H \times W \times 1$ . Cette entrée est directement reliée à une couche de convolution 2D, caractérisée par un nombre de filtres égal à  $E$ , défini comme étant notre « paramètre d'extension ».

Ce paramètre est choisi selon les différentes formes d’entrées spatiales afin de favoriser leur intégration dans le réseau. Son rôle et sa valeur seront explicités dans la section V.2.1.2. La sortie de cette couche (de dimension  $H \times W \times E \times 1$ ), est ensuite traitée par une succession de couches de convolution 3D (intitulées « *conv3D blocks* »), décrites également en section V.2.1.2.

- Les informations d’entrées positionnelles : Elles consistent en l’ajout des entrées  $T_{prior}$  et  $T_{suivi}$ , précédemment étudiées. Les intégrations sont similaires à notre méthode précédente : via une concaténation séparée, de deux vecteurs de dimension  $(\# \times 1)$  et sur la même couche du réseau (FC-512). La succession des couches entièrement connectées (FC) jusqu’à la prédiction finale  $\hat{T}_t$ , est également conservée de notre précédente méthode (voir section IV.2.1.2).

- Les informations d’entrées spatiales : Nous avons évalué différentes formes d’entrées spatiales qui seront ajoutées séparément, par le biais de différentes expériences (décrites en section V.3.1). Pour rappel, une description plus détaillée des méthodes de simulation de ces entrées spatiales est fournie en section III.1.4.2.

- L’entrée spatiale *locale* correspond à la concaténation d’une ou plusieurs coupes voisines  $I_{t-1}$  jusqu’à  $I_{t-V}$ . On définit  $V$  comme le nombre de coupes voisines considérées. Cette entrée est directement concaténée à l’image d’entrée principale  $I_t$ , de façon à former une image à plusieurs canaux  $H \times W \times (V + 1)$ . Cette entrée est ensuite envoyée à la branche principale du réseau, composée des couches de convolution 3D, qui permettent d’analyser et de dissocier le contenu hors-plan de toutes les images voisines.
- L’entrée spatiale *globale* correspond à l’ajout du volume de référence  $V_{ref}$ . Cette entrée est traitée à travers une branche parallèle à la branche principale, composée de la même succession de couches de convolution 3D (*conv3D block*). Cela permet d’obtenir un encodage similaire et équilibré, entre les deux branches, jusqu’à la concaténation de celles-ci.
- L’entrée spatiale *partielle* correspond à une sous-partie du volume de référence, définie comme  $V_{sub,t}$ . Nous rappelons que ce sous-volume est déterminé à partir de l’orientation estimée pour la coupe précédente  $T_{prior}$  (voir Figure III.18). Cela permet de ne considérer qu’un espace restreint, centré et recadré autour d’une position de recherche proche de celle recherchée pour l’image courante. Dans ce cas-ci, le reste du réseau subit quelques modifications : les informations positionnelles ( $T_{prior}$  et  $T_{suivi}$ ) étant d’ores et déjà utilisées pour orienter le sous-volume, ces entrées ne seront pas ajoutées sur la couche FC-512 du bloc n°2.

- La pénalisation spatiale : Afin de considérer des informations sur le contenu iconique des images localisées par le réseau, nous avons considéré un terme de similarité ( $L_{SSIM}$ ). Celui-ci se base sur la ressemblance des pixels entre l’image rééchantillonnée à la position prédite  $I(V_{ref}, \hat{T}_t)$  et l’image courante  $I_t$  (voir section V.2.2.1). La mesure de similarité est calculée entre les pixels de la coupe ainsi extraite et la coupe d’entrée. Sa combinaison avec la fonction de coût précédemment étudiée ( $L_{MSE}$ ) sera explicitée dans la section V.2.2.2.

### V.2.1.2 Architecture du réseau

Nous avons modifié l'architecture de la branche principale de notre réseau précédent, passant d'un réseau convolutif 2D (voir section IV.2.1.2) à un réseau convolutif 3D. Les modifications du réseau ont notamment été inspirées de l'étude de Guo et. al [56].

Cette branche principale se compose d'une première couche de convolution 2D, où le nombre de filtres  $E$  définit la dimension de sortie de cette couche (sortie :  $H \times W \times E \times 1$ ). Cela correspond à une image 3D (à  $E$  canaux), composée de toutes les cartes caractéristiques concaténées obtenues suite à cette étape de convolution 2D. Pour faire correspondre la taille de la sortie ainsi obtenue à la taille de l'entrée spatiale (*locale, partielle* ou *globale*) de la branche parallèle, le paramètre  $E$  doit être adapté à la taille de la profondeur d'entrée :  $E=V$  pour une entrée de  $V$  coupes locales,  $E = 128$  pour une entrée en volume complet ou  $E = 32$  pour une entrée en sous-volume.

Ensuite, deux blocs principaux de convolution 3D (« conv3D blocks ») avec un nombre croissant de filtres sont appliqués. Chacun d'eux contient deux couches successives de convolution 3D (noyau= $5 \times 5 \times 5$  puis noyau  $3 \times 3 \times 3$ ), suivies de couches de "pooling" (noyau=2) et d'une fonction d'activation ReLU. Cette succession de couches de convolution est commune à la branche principale du réseau et à la branche parallèle. Le vecteur ainsi encodé est transformé en un vecteur 1D ("flatten"), avant d'être concaténé à la branche secondaire et/ou traité par la suite du réseau.

Enfin, le réseau consiste en 6 couches entièrement connectées, avec un nombre décroissant de neurones jusqu'à la couche de sortie finale.

Une description couche par couche est illustrée dans la Figure V.2 et est disponible en Annexe A (Table 3). Le nombre de paramètres de ce réseau dépend de la configuration envisagée selon la combinaison des entrées spatiales. Ceux-ci sont décrits en Annexe A (Table 3) et totalisent entre 71 et 141 M de paramètres sur l'ensemble des couches du modèle.

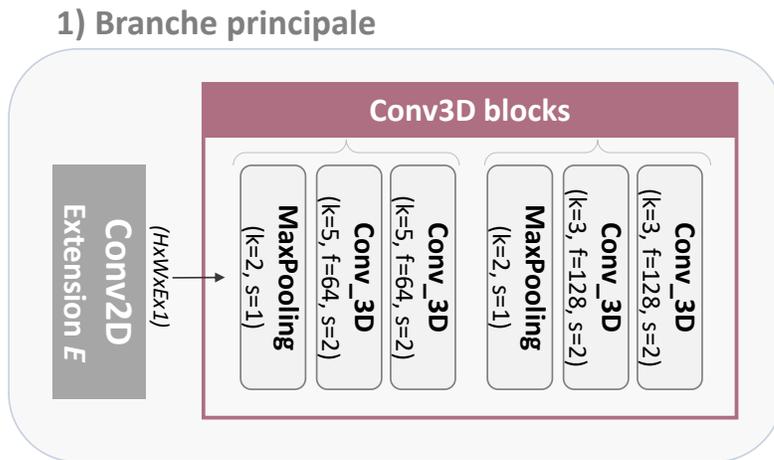


FIGURE V.2 – Description de l'architecture de la branche principale, couche par couche.

## V.2.2 Apprentissage

Afin de créer une fonction de coût qui soit révélatrice de la qualité de nos prédictions de recalage, tout en renforçant l’apport des informations spatiales, nous avons implémenté deux fonctions : (i) une fonction supervisée basée sur le calcul de MSE (utilisée également dans la méthode précédente voir section IV.2.2.1), et (ii) une fonction auto-supervisée basée sur le calcul de SSIM. Ces différentes fonctions de coût, notées  $L_{MSE}$  et  $L_{SSIM}$ , ont été évaluées séparément puis combinées ensembles par le biais de différentes expériences.

### V.2.2.1 Terme d’auto-pénalisation de similarité iconique (SSIM)

La mesure de similarité est appliquée entre les pixels de l’image rééchantillonnée à la position prédite  $I(V_{ref}, \widehat{T}_t)$  et l’image courante  $I_t$ . Cela nécessite d’avoir le volume de rééchantillonnage et l’image de comparaison en entrée du module de pénalisation. L’image  $I_t$ , fournie comme entrée du réseau, est toujours accessible lors de l’optimisation, alors que le volume  $V_{ref}$  n’est pas toujours fourni en entrée spatiale selon les expériences envisagées. Lorsque  $V_{ref}$  n’est pas fourni en entrée globale du réseau, il sera fourni directement en entrée du module de pénalisation spatiale.

La mesure SSIM est bornée entre 0 et 1 où 1 traduit le meilleur score. L’optimisation devra minimiser le calcul défini dans l’équation V.1a, pour chaque élément  $k$  de l’entraînement. Pour cela, nous avons implémenté un module de rééchantillonnage différentiable STN [92], qui permet d’extraire la coupe correspondant à la position prédite au sein du volume de référence. L’algorithme se compose des plusieurs étapes. D’abord, une création d’une grille des coordonnées des points physiques correspondant à la taille de l’image que nous voulons obtenir, dans notre cas 2D ( $H \times W$ ). Ces coordonnées physiques sont ensuite transformées selon la transformation prédite inverse  $\widehat{T}_t^{-1}$ , par un calcul de multiplication matricielle. Les points obtenus doivent maintenant être convertis en coordonnées matricielles de voxels, en utilisant la mise à l’échelle correspondante. Pour finir, une interpolation trilineaire est effectuée entre les valeurs de voxels du volume de rééchantillonnage les plus proches des points transformés de la grille. Tout point intersecté à l’extérieur du volume de rééchantillonnage est associé à la valeur zéro (voxel noir).

$$L_{SSIM} = \frac{1}{K} \sum_{k=1}^K SSIM_k \quad (\text{V.1a})$$

$$\text{avec } SSIM_k = \left\| I_k - I(V_{ref}, \widehat{T}_k) \right\|^2 \quad (\text{V.1b})$$

Cette mesure de similarité comme fonction de coût auto-supervisée permet de tenir compte des structures anatomiques présentes dans l’espace de référence. Cependant cette mesure a un profil de convergence plus contraignant, puisque la différence de pixels n’est pertinente qu’entre deux images suffisamment proches. De plus, afin de ne pas considérer les zones noires issues de l’échantillonnage, la mesure de similarité est réalisée sur l’intersection commune (et non vide) des deux images. En fonction de l’étendue des zones noires, cela peut fortement contraindre la valeur de la mesure et biaiser l’interprétation lors de l’optimisation.

### V.2.2.2 Combinaison et pondération de plusieurs termes

Il est également possible de combiner les termes  $L_{MSE}$  et  $L_{SSIM}$  au sein d'une même fonction de coût, comme défini en équation V.2. La bonne pondération de ces termes dans la même fonction ( $L_{tot}$ ) reste cependant à déterminer pour la rendre facilement convergente.

Nous avons envisagé deux cas de pondération des différents termes : adaptative ou fixe. Dans le premier cas, les coefficients de pondération utilisés ( $\alpha, \beta$ ) sont ajustés et normalisés en fonction du rapport des différents termes à combiner. Ce ratio est obtenu après chaque itération d'apprentissage ("epoch") afin d'avoir constamment un rapport équilibré entre les deux termes, quel que soit leur profil de convergence. Dans le second cas, les coefficients de pondération sont fixés à différentes valeurs selon le profil de convergence respectif (rapide ou lent). Cela permet de se concentrer sur les métriques les plus pénalisantes dans un premier temps, puis sur des métriques plus précises pour affiner le résultat par la suite. Les différentes pondérations testées sont décrites en section V.3.1.

$$L_{tot} = \alpha * L_{MSE} + \beta * L_{SSIM} \quad (\text{V.2})$$

## V.3 Expériences et configurations

### V.3.1 Différentes formes d'informations spatiales

Afin de déterminer indépendamment quelle entrée spatiale permet de faciliter la tâche de recalage, nous avons réalisé différentes expériences d'ablation, où les entrées sont ajoutées séparément. Les différentes expériences envisagées selon les combinaisons possibles de ces données d'entrée sont dénommées respectivement :

- Expérience « *locale* » : avec l'ajout des  $V$  coupes voisines de  $I_{t-1}$  jusqu'à  $I_{t-V}$ . Par défaut, on utilise  $V = 2$ , sinon la mention *locale-V* sera ajoutée selon le nombre de coupes considérées.
- Expérience « *globale* » : avec l'ajout du volume de référence  $V_{ref}$ .
- Expérience « *partielle* » : avec l'ajout d'une sous-partie orientée du volume de référence  $V_{sub,t}$ .

Concernant les autres entrées de ce réseau, et notamment les entrées positionnelles étudiées dans le chapitre précédent, nous avons choisi de tester les performances avec/sans  $T_{prior}$  et avec/sans ajout du  $T_{suivi}$  (respectivement Scénario  $Im + T_{suivi}$  et Scénario  $Im$ ). Alors que les contributions de ces deux entrées positionnelles ( $T_{suivi}$  et  $T_{prior}$ ) ont déjà été démontrées, l'intérêt est d'étudier les contributions des entrées spatiales de manière indépendante. Cela permet d'obtenir des conclusions non-biaisées en confrontant ces deux types d'informations ou en les combinant.

De plus, seule l'évaluation séquentielle « *par portion* » (résultats successifs  $T_{prior} = \widehat{T}_{t-1}$  avec réajustement, voir section III.2.2.2), et son évaluation « *contrôle* » associée (sans utilisation ni du réseau ni des images d'entrée, voir IV.3.1) seront présentées dans les résultats suivants.

Enfin, différentes fonctions de coût ont été envisagées lors de ces expériences. Nous avons fait le choix de tester chacun des termes d'optimisation  $L_{MSE}$  et  $L_{SSIM}$  séparément dans un premier temps (où le coefficient associé au bon terme est fixé à 1 et les autres à 0), puis en les combinant avec différentes pondérations. Nous avons choisi de n'illustrer dans ce manuscrit que les résultats avec

la configuration ( $\alpha = 1$  et  $\beta = 50$ ), qui nous a permis d'obtenir les meilleurs profils de convergence et les meilleurs résultats. Cette expérience est intitulée  $L_{MSE+SSIM}$ . Par défaut, si aucune mention n'est spécifiée, la fonction de base utilisée est  $L_{MSE}$ .

### V.3.2 Différentes architectures

Afin de valider la nouvelle architecture proposée pour la branche principale (bloc 1, Figure V.1), nous avons comparé 2 configurations différentes représentées sur la Figure V.3.

- La première architecture (Figure V.3-A) est inspirée de notre méthode précédente (voir section IV.2.1.2). Elle utilise des convolutions 2D, habituellement adaptées aux informations d'entrée 2D. Les entrées *locales* (concaténées) sont alors traitées par ces mêmes couches de convolution 2D tandis que les entrées *globales* ou *partielles* sont traitées sur une branche parallèle avec des convolutions 3D. Chaque entrée est donc traitée par des convolutions adaptées à sa nature et à sa dimension. Cette architecture sera intitulée « *Main-2D* » et est décrite en Annexe A (Table 4).
- La deuxième architecture (Figure V.3-B) est la nouvelle architecture proposée dans cette étude (voir section V.2.1.2). Contrairement à l'architecture précédente, elle permet de maintenir un équilibre entre les branches, quelle que soit l'entrée spatiale ajoutée. Cette architecture sera intitulée « *Main-3D* ».

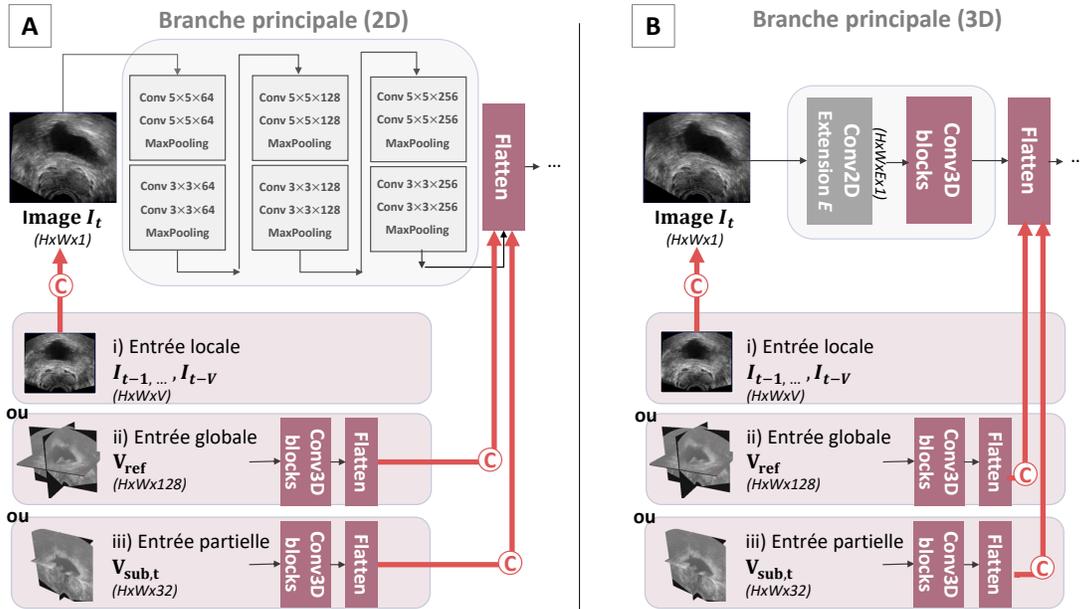


FIGURE V.3 – Différentes architectures utilisées pour la branche principale (bloc 1) : convolution 2D en A ou convolution 3D en B. Le reste du réseau est gardé à l'identique.

### V.3.3 Jeux de données

Pour réaliser les différentes expériences proposées, nous nous sommes appuyés sur la base de données issue du balayage base-apex, présentée précédemment (voir section III.1.3.3). Nous détaillons ci-dessous les caractéristiques associées à cette base de données.

Nous avons simulé des trajectoires de balayage base-apex, que nous avons appliquées à une grande base de données composée de 600 volumes de biopsie TRUS, provenant des sessions de biopsie sur plus de 100 patients différents. Deux niveaux de difficulté sont simulés et dénommés «  $I_{i,t}$  à  $V_i$  » ou «  $I_{i,t}$  à  $V_j$  ». Les données ainsi simulées sont réparties dans les jeux d’entraînement, validation et test selon la description de la Table V.1. Nous avons pris soin de garder des patients inédits dans l’ensemble de données test, afin d’évaluer les bonnes capacités de généralisation du modèle sur de nouvelles anatomies.

TABLE V.1 – Description du jeu de données issu du balayage base-apex avec les pourcentages de répartition des échantillons d’images entre les jeux d’entraînement, de validation, et de test.

Base de données	Entraînement & Val (75 % & 8 %)	Test (17%)	Caractéristiques
« $I_{i,t}$ à $V_i$ »	5 $V_i$ x 80 patients $\approx$ 30 000 images	5 $V_i$ x 20 patients $\approx$ 6 000 images	Trajectoire de balayage ( $\pm 40^\circ$ x) Taille des images : $256 \times 256$ ( $T_t = \theta x, \theta y, \theta z, t_x, t_y, t_z$ ) avec ( $\# = 6$ ) Fréquence d’échantillonnage : $\approx 150$ ms
« $I_{i,t}$ à $V_j$ »	5 $V_{i,j}$ x 80 patients $\approx$ 16 000 images	5 $V_{i,j}$ x 20 patients $\approx$ 3 000 images	Difficulté supplémentaire avec image indépendante et anatomie différente ( $T_t = \theta x, \theta y, \theta z, t_x, t_y, t_z$ ) avec ( $\# = 6$ ) Fréquence d’échantillonnage : $\approx 150$ ms

Nous rappelons que ces trajectoires simulent des mouvements d’allers-retour entre la base et l’apex de la prostate. Pour reproduire un pseudo schéma de biopsie, nous avons divisé arbitrairement nos trajectoires en 12 portions égales. Ces 12 « pseudo » biopsies seront notamment utilisées pour établir la navigation par portion et les réajustements nécessaires.

De plus, nous avons ajusté notre fréquence de rééchantillonnage pour reproduire au mieux une temporalité réaliste par rapport à celle observée pour des mouvements de balayage base-apex, réalisés par des cliniciens sur le simulateur pour lesquels la fréquence d’acquisition est bien connue. Pour respecter certaines limites de mémoire numérique, nous avons simulé une fréquence d’acquisition échographique d’environ 6.7 Hz (150 ms), inférieure à celle disponible en clinique. Cette fréquence d’acquisition simulée est à distinguer de la fréquence de prédiction que l’on vise pour respecter les spécificités d’applications cliniques (temps-réel  $\leq 50$  ms).

Pour finir, toutes les images sont redimensionnées à la taille  $256 \times 256$ . Pour la taille du sous-volume nous avons utilisé une taille de  $256 \times 256 \times 32$  pour respecter les mêmes conditions expérimentales que la méthode de la littérature [56], tandis que pour le volume entier nous avons choisi une taille de  $256 \times 256 \times 128$  en accord avec les capacités de calcul des cartes graphiques utilisées.

### V.3.4 Comparaison avec l'état-de-l'art

Parmi les études référencées dans notre état-de-l'art et décrites précédemment (voir Chapitre II), l'étude de Guo [56] s'intéresse en particulier à l'intégration d'informations spatiales. Les auteurs démontrent qu'en associant des entrées spatiales 2D (image courante) et 3D (sous-volumes) dans une architecture 3D équilibrée avec un terme de pénalisation, ils obtiennent de meilleures performances. Leur réseau développé s'intitule « *FVR-Net* » et est illustré sur la Figure II.10.

Afin de pouvoir nous comparer à cette étude, nous avons réimplémenté leur architecture (à partir du code accessible sur Github<sup>1</sup>) et évalué leur réseau dans les mêmes conditions que le nôtre (évaluation séquentielle et base de données). Contrairement à l'étude originale [56], les sous-volumes d'entrées ne sont pas initialement orientés à partir de la position de la vérité terrain bruitée, mais orientés à partir des résultats des prédictions précédentes ( $T_{prior}$ ) pour se conformer à notre approche d'évaluation séquentielle. De plus, les expériences sont évaluées sur nos bases de données issues de la simulation par balayage base-apex avec deux niveaux de difficultés «  $I_{i,t}$  à  $V_i$  » et «  $I_{i,t}$  à  $V_j$  », alors que l'étude d'origine ne présente des résultats que sur l'équivalent du premier cas.

Nous nous comparons également à notre première méthode, décrite dans le Chapitre IV et associée à l'étude [103]. D'un point de vue méthodologique, celle-ci correspond à l'expérience « *locale* » en utilisant l'architecture « *Main-2D* ». La ré-évaluation des travaux précédents dans les conditions expérimentales établies dans ce chapitre répondra à l'appellation « *Pre-Net* ».

### V.3.5 Paramètres d'apprentissage

Toutes les expériences ont été réalisées en entraînant notre réseau à partir de l'optimiseur Nadam [99] et avec un taux d'apprentissage initial de  $1e-4$ , en utilisant des GPU A100 (NVIDIA) via une implémentation sur TensorFlow [100]. Le réseau a été entraîné sur plus de 30 epochs, avec une taille de lot  $K$  variable selon les expériences d'entrée spatiales testées :  $K=50$ , 30 ou 16 pour les expériences d'entrée spatiale *locale*, *partielle* ou *globale* respectivement.

Deux types d'apprentissage ont été utilisés : soit un apprentissage à partir de zéro, soit un apprentissage par transfert. Sur les expériences  $I_{i,t}$  à  $V_i$ , les poids du réseau ont été initialisés avec une distribution Gaussienne (moy=0, std=0.01) et optimisés durant l'entraînement. Sur les expériences  $I_{i,t}$  à  $V_j$ , les poids du réseau sont initialisés à partir des valeurs des poids du même modèle pré-entraîné sur les expériences  $I_{i,t}$  à  $V_i$ .

---

1. <https://github.com/DIAL-RPI/FVR-Net>

## V.4 Résultats

### V.4.1 Différentes formes d'informations spatiales

Dans les résultats qui vont suivre, nous avons comparé les performances avant et après recalage pour toutes les expériences envisagées (3 formes d'entrées spatiales avec/sans pénalisation spatiale), selon les deux scénarios positionnels (Scénario  $Im$  ou  $Im + T_{suivi}$ ), et selon les deux bases de données proposées («  $I_{i,t}$  à  $V_i$  » ou «  $I_{i,t}$  à  $V_j$  »). La Table V.2 résume l'ensemble des valeurs moyennes de TRE et la Figure V.4 récapitule l'ensemble des distributions obtenues. Ayant obtenus des tendances et conclusions similaires pour les expériences avec/sans  $T_{prior}$ , seuls les résultats avec  $T_{prior}$  sont présentés ici. Des tests statistiques de Wilcoxon (non-paramétriques, seuil = 0.05) ont été réalisés sur chaque couple d'expériences et de scénarios possibles. Ils ont tous mis en évidence des distributions significativement différentes excepté entre les expériences *part* et *part+SSIM*.

TABLE V.2 – Comparaison des performances de recalage (valeurs moyennes de TRE) à travers une évaluation séquentielle par portion, et pour différentes expériences. Différentes formes d'entrées spatiales sont comparées : *locale/globale/partielle*. La notation "+SSIM" fait référence à l'utilisation de la fonction de coût  $L_{MSE+SSIM}$  dans le module de pénalisation. Les Scénarios  $Im + T_{suivi}$  et  $Im$  représentent respectivement les scénarios avec les images d'entrée avec/sans l'entrée de suivi de sonde. Les résultats surlignés en vert reflètent les meilleures performances pour chaque base de données.

Évaluation séquentielle Avec $T_{prior}$ $I_{i,t}$ à $V_i$	TRE ( $\downarrow$ ) : $\mu \pm \sigma$ (mm)					
	TRE avant recalage : 9.52 ( $\pm$ 6,49)					
	local	glob	part	local+SSIM	glob+SSIM	part+SSIM
Scenario $Im$						
<b>Par portion</b> : $T_{prior} = \widehat{T}_{t-1}$	2.74 $\pm$ 6.04	3.83 $\pm$ 7.11	4.57 $\pm$ 6.42	2.91 $\pm$ 6.00	3.43 $\pm$ 6.71	4.69 $\pm$ 6.43
<b>Contrôle</b> : Que $T_{prior}$	5.30 $\pm$ 6.32	6.17 $\pm$ 7.23	5.43 $\pm$ 5.89	5.37 $\pm$ 6.28	5.88 $\pm$ 6.88	5.55 $\pm$ 5.75
Scenario $Im + T_{suivi}$						
<b>Par portion</b> : $T_{prior} = \widehat{T}_{t-1}$	0.21 $\pm$ 0.28	0.21 $\pm$ 0.19	6.43 $\pm$ 6.30	0.62 $\pm$ 0.95	0.52 $\pm$ 0.55	7.13 $\pm$ 7.19
<b>Contrôle</b> : Que $T_{prior} * T_{suivi}$	1.01 $\pm$ 2.23	1.00 $\pm$ 2.24	6.42 $\pm$ 6.32	1.33 $\pm$ 2.30	1.24 $\pm$ 2.23	7.09 $\pm$ 7.22
Évaluation séquentielle Avec $T_{prior}$ $I_{i,t}$ à $V_j$	TRE ( $\downarrow$ ) : $\mu \pm \sigma$ (mm)					
	TRE avant recalage : 11.72 ( $\pm$ 6,28)					
	local	glob	part	local+SSIM	glob+SSIM	part+SSIM
Scenario $Im$						
<b>Par portion</b> : $T_{prior} = \widehat{T}_{t-1}$	5.06 $\pm$ 5.70	8.87 $\pm$ 5.82	5.01 $\pm$ 7.82	5.80 $\pm$ 7.00	7.30 $\pm$ 7.25	7.28 $\pm$ 8.50
<b>Contrôle</b> : Que $T_{prior}$	7.70 $\pm$ 6.24	9.65 $\pm$ 6.11	6.35 $\pm$ 6.88	5.91 $\pm$ 6.77	9.39 $\pm$ 7.65	7.63 $\pm$ 7.67
Scenario $Im + T_{suivi}$						
<b>Par portion</b> : $T_{prior} = \widehat{T}_{t-1}$	2.68 $\pm$ 1.49	2.81 $\pm$ 2.00	5.28 $\pm$ 6.30	3.97 $\pm$ 2.03	3.15 $\pm$ 2.17	5.91 $\pm$ 7.19
<b>Contrôle</b> : Que $T_{prior} * T_{suivi}$	3.11 $\pm$ 2.67	3.15 $\pm$ 2.94	5.23 $\pm$ 6.32	4.28 $\pm$ 2.93	3.53 $\pm$ 3.10	5.84 $\pm$ 7.22

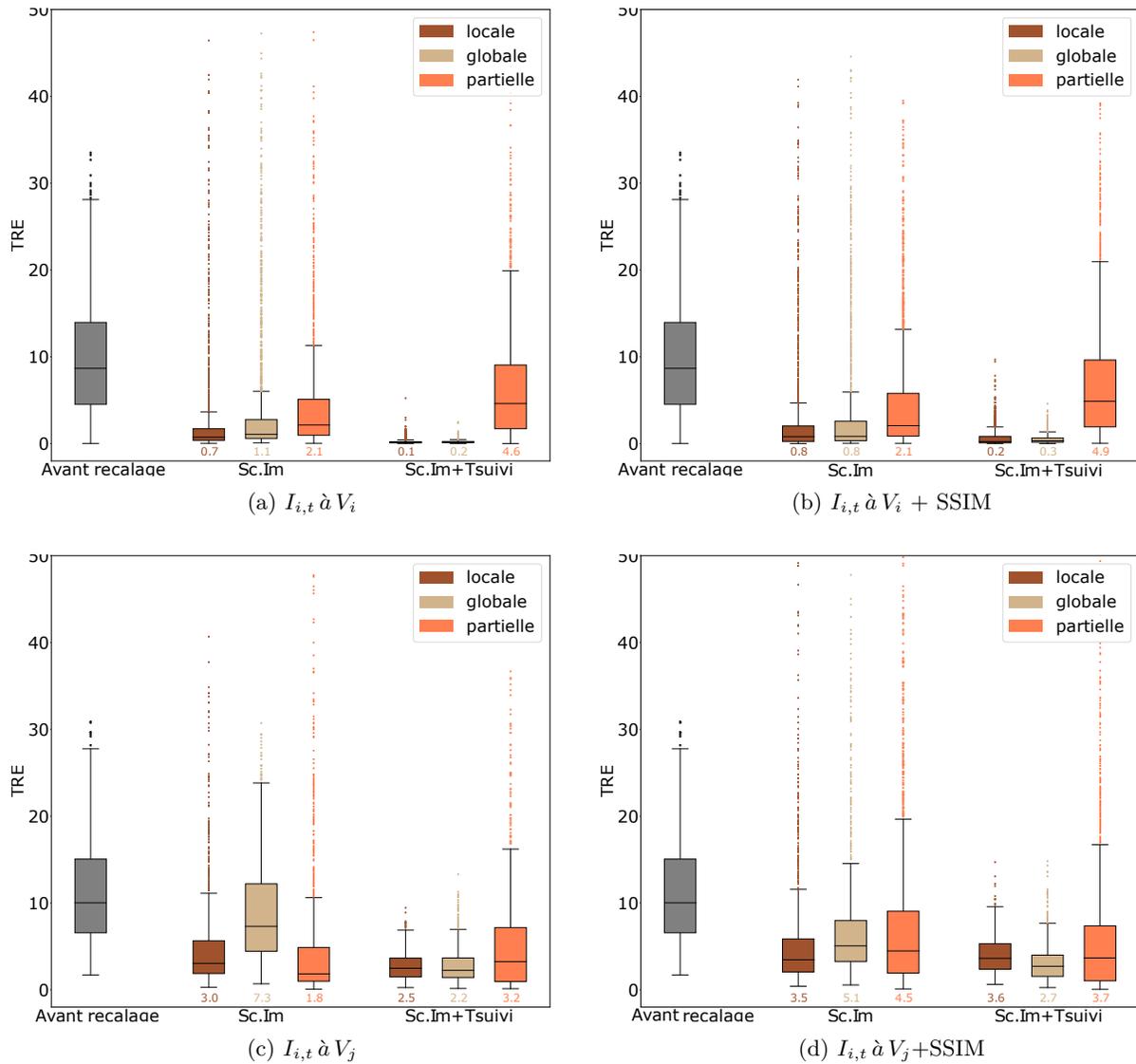


FIGURE V.4 – Comparaison des distributions des métriques d’évaluation de recalage, selon les différentes expériences testées (3 formes d’entrées spatiales avec/sans pénalisation spatiale, avec/sans  $T_{suivi}$ ) et sur les deux jeux de données («  $I_{i,t}$  à  $V_i$  » ou «  $I_{i,t}$  à  $V_j$  »). Les valeurs médianes sont également reportées.

Tout d’abord toutes les expériences sont comparées à leur évaluation de *contrôle* associée (traitement exclusif des entrées positionnelles), où elles obtiennent toutes de meilleurs résultats (et tous significativement différents). Cela démontre la bonne contribution du réseau et le bon traitement entre tous les types d’entrées.

Sur les deux bases de données proposées («  $I_{i,t}$  à  $V_i$  » ou «  $I_{i,t}$  à  $V_j$  »), on obtient les mêmes tendances et conclusions.

- Le Scénario  $Im + T_{suivi}$  démontre systématiquement de meilleures performances que le Scénario  $Im$ . Ce constat avait déjà pu être observé lors des résultats du chapitre précédent (voir section IV.4.1) et se confirme quel que soit l’entrée spatiale ajoutée. Cela démontre la contribution systématique des informations positionnelles, même lorsque celles-ci sont combinées aux informations spatiales.
- L’expérience avec les informations spatiales *globales* permet dans certains cas d’améliorer les performances de recalage, comparativement aux expériences d’entrées *locales*, notamment pour le scénario  $Im + T_{suivi}$ . Ces différences de performances sont assez peu marquées et ne permettent pas de conclure quant à la supériorité de ce scénario, également plus coûteux en temps d’entraînement.
- L’expérience avec les informations spatiales *partielles* ne permet pas d’améliorer les performances de recalage, comparativement aux deux autres entrées *locales* ou *globales*. Elle résulte en une qualité de recalage bien inférieure aux deux autres expériences.
- L’utilisation de la fonction de coût  $L_{MSE+SSIM}$  ajoutée dans le module de pénalisation spatiale ne semble pas améliorer la qualité du recalage, quelle que soit l’entrée spatiale testée. Les résultats obtenus sont similaires à ceux obtenus dans la configuration par défaut ( $L_{MSE}$  sans ajout de pénalité). Pour ces cas-ci, ce sont les expériences *globales* qui semblent le mieux bénéficier de l’ajout de la pénalisation  $+SSIM$ , même si elles n’égale pas les meilleures performances sans pénalisation.

Au final, la meilleure configuration évaluée (pour les deux bases de données) est obtenue pour le scénario  $Im + T_{suivi}$  et avec  $T_{prior}$ , de l’expérience *locale*. Pour la base de données «  $I_{i,t}$  à  $V_i$  », on mesure une réduction de l’erreur partant de 9.52 mm ( $\pm 6,49$ ) à 0.21 mm ( $\pm 0,28$ ). Pour la base de données «  $I_{i,t}$  à  $V_j$  », on mesure une réduction de l’erreur partant de 11.72 mm à 2.68 ( $\pm 1,49$  mm). L’erreur finale est cette fois-ci plus importante, reflétant une complexité supplémentaire liée aux déformations de prostate.

### V.4.2 Différentes architectures

Nous avons comparé 2 architectures différentes pour la branche principale du réseau : *Main-2D* et *Main-3D* (voir section V.3.2). La Table V.3 et la Figure V.5 résument les résultats obtenus pour différentes entrées spatiales et pour les deux jeux de données testés. Seul le scénario  $Im + T_{suivi}$  avec  $T_{prior}$  est présenté. Pour toutes les expériences réalisées, les performances obtenues avec l'architecture *Main-3D* sont supérieures à celles de l'architecture *Main-2D*. Il semble donc, que quelle soit la forme d'entrée spatiale ajoutée (empilement 2D ou 3D), un traitement par convolution 3D soit plus adapté et favorise l'apprentissage du modèle. De plus, il semble que plus nous ajoutons de l'information spatiale au réseau (entrée spatiale avec de plus en plus de coupes précédentes), moins les performances s'améliorent. Cette tendance est valable globalement pour les deux architectures et les deux jeux de données, malgré la présence de certaines exceptions.

TABLE V.3 – Comparaison des performances de recalage (TRE moyennes) selon les architectures testées.

Évaluation séquentielle Sc. $Im + T_{suivi}$ pour $I_{i,t}$ à $V_i$	TRE avant recalage : 9.52 ( $\pm$ 6,49)				
	local-2	local-4	local-8	local-16	global
Avec $T_{prior}$					
<b>Main-2D</b>	0.25 $\pm$ 0.47	0.42 $\pm$ 1.01	0.75 $\pm$ 1.05	0.42 $\pm$ 0.8	0.30 $\pm$ 0.29
<b>Main-3D</b>	0.21 $\pm$ 0.28	0.25 $\pm$ 0.97	0.45 $\pm$ 1.01	0.48 $\pm$ 0.61	0.21 $\pm$ 0.19
Évaluation séquentielle Sc. $Im + T_{suivi}$ pour $I_{i,t}$ à $V_j$	TRE avant recalage : 11.72 ( $\pm$ 6,28)				
	local-2	local-4	local-8	local-16	global
Avec $T_{prior}$					
<b>Main-2D</b>	2.89 $\pm$ 2.50	2.34 $\pm$ 2.03	3.11 $\pm$ 2.88	2.82 $\pm$ 2.40	N/A
<b>Main-3D</b>	2.68 $\pm$ 1.49	1.77 $\pm$ 1.47	2.65 $\pm$ 1.99	1.39 $\pm$ 1.28	2.81 $\pm$ 2.00

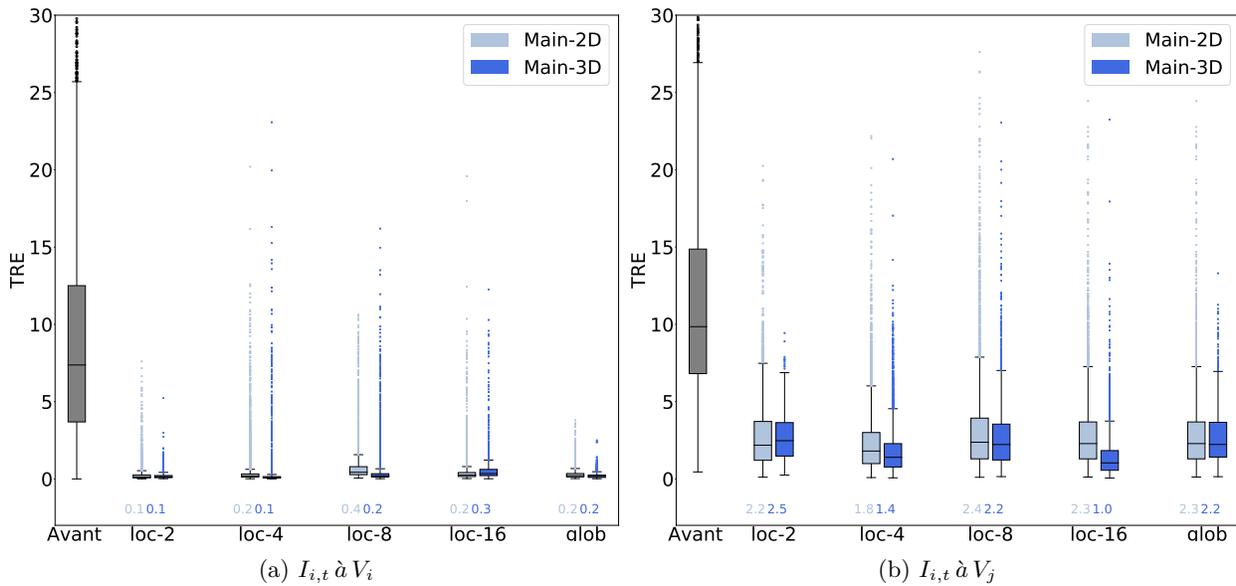


FIGURE V.5 – Variations de la taille des entrées spatiales testées selon les 2 architectures utilisées.

### V.4.3 Comparaison avec l'état-de-l'art

#### V.4.3.1 Réimplémentation et étude d'ablation

Afin de nous comparer à la principale méthode concurrente de la littérature *FVR-Net* [56], nous avons réimplémenté leur méthode dans les mêmes conditions expérimentales que l'étude originale (paramètres d'apprentissage, initialisation des sous-volumes à partir de la vérité terrain bruitée). Cette réimplémentation équivaut à l'expérience notée *part+SSIM* à travers une évaluation dite théorique. Dans un deuxième temps, nous avons mené des expériences complémentaires en étendant leur étude d'ablation. La Table V.4 présente ces résultats sur notre base de données.

TABLE V.4 – Performances de recalage de l'étude publiée *FVR-Net* à travers une étude étendue des formes d'informations spatiales. Les résultats de l'étude initiale sont reportés ainsi qu'une ré-évaluation sur notre base de données en conservant leurs conditions expérimentales.

Évaluation théorique équivalent $I_{i,t}$ à $V_i$ pour FVR-Net [56]	TRE ( $\downarrow$ ) : $\mu \pm \sigma$ (mm)					
	TRE avant recalage : 7.68 (Publication) et $9.52 \pm 6,49$ (Ré-évaluation)					
	local	glob	part	local+SSIM	glob+SSIM	part+SSIM
Publication initiale	N/A	N/A	3.17	N/A	N/A	2.73
Ré-évaluation	N/A	$2.85 \pm 3.49$	$4.75 \pm 9.39$	N/A	$2.53 \pm 4.95$	$2.57 \pm 6.03$

Tout d'abord, nous observons la même tendance que dans l'article original avec une amélioration significative des performances lors de la combinaison des sous-volumes d'entrée et de la pénalisation spatiale (passage de 4.75 pour *part* à 2.57 pour *part+SSIM*). Cela vérifie la bonne implémentation et la bonne contribution des modules ajoutés.

On peut cependant noter que l'étude conclut quant à l'efficacité des structures spatiales *partielles* et de pénalisation sans présenter d'autre étude d'ablation ou de comparaison avec l'état-de-l'art. En étendant leur étude aux expériences d'entrées spatiales *globales*, nous remarquons que les nouvelles expériences *glob* et *glob+SSIM* obtiennent de meilleurs résultats que leurs expériences *part* et *part+SSIM* proposées. Cela suggère que d'autres entrées peuvent être plus bénéfiques que les structures suggérées et coïncide avec les conclusions de notre étude d'ablation (voir Table V.2).

#### V.4.3.2 Comparaison quantitative

Pour aller au delà de l'évaluation théorique des deux méthodes de la littérature *FVR-Net* et *Pre-Net* présentées en section V.3.3, nous avons également comparé notre approche et ces deux méthodes à travers une ré-évaluation complète : sur nos deux ensembles de données et en utilisant notre protocole d'évaluation séquentielle. La Table V.5 résume la comparaison avec ces méthodes. La notation *SpT-Net* fait référence à notre meilleure configuration obtenue dans les résultats précédents (architecture *Main-3D*, *local*, scénario  $Im + T_{suivi}$  avec  $T_{prior}$  de la Table V.2).

Pour cette étude (avec un nombre d'expériences plus restreint), nous avons utilisé une stratégie de validation croisée (sur 5 groupes), afin d'augmenter la fiabilité des résultats. Cette nouvelle répartition des données reste similaire à celle présentée en Table V.1, où les jeux d'entraînement sont divisés en 4 groupes tandis que le jeu de test correspond au dernier groupe. À chaque étape, un groupe différent est utilisé pour le test tandis que les 4 autres sont utilisés pour l'apprentissage. Les

TABLE V.5 – Comparaison des performances de recalage avec les méthodes de la littérature ré-évaluées à travers notre protocole d'évaluation séquentielle par portion et sur nos deux jeux de données. Les p-valeurs entre ces méthodes et la méthode proposée dans ce chapitre sont également reportées (valables pour toutes les métriques). Le temps de calcul moyen par prédiction ainsi que le nombre de paramètres des réseaux sont également indiqués. N.B : Pour rappel, la métrique  $R^2$  ne peut pas être calculée pour des variables constantes (notamment  $tx, \theta y, \theta z$  du jeu " $I_{i,t}$  à  $V_i$ ").

Méthodes ( $I_{i,t}$ à $V_i$ )	TRE (mm) $9.07 \pm 6.55$	NCC $0.84 \pm 0.12$	$R^2$ ( $tx, ty, tz, \theta x, \theta y, \theta z$ )	Calcul (s) $0.06 \pm 0.02$	Param. (M) 71	p-valeur (vs SpT)
SpT-Net (proposée)	<b><math>0.19 \pm 0.77</math></b>	<b><math>0.99 \pm 0.04</math></b>	( $\emptyset, 0.99, 0.99,$ $0.99, \emptyset, \emptyset$ )	$0.06 \pm 0.02$	71	
FVR-Net [56]	$2.28 \pm 8.52$	$0.95 \pm 0.11$	( $\emptyset, 0.84, 0.80,$ $0.81, \emptyset, \emptyset$ )	$0.17 \pm 0.06$	68	$p \leq 0.001$
Pre-Net [103]	$0.26 \pm 0.76$	$0.98 \pm 0.03$	( $\emptyset, 0.99, 0.99,$ $0.99, \emptyset, \emptyset$ )	$0.03 \pm 0.01$	41	$p \leq 0.001$
Méthodes ( $I_{i,t}$ à $V_j$ )	TRE (mm) $11.14 \pm 6.13$	NCC $0.81 \pm 0.11$	$R^2$ ( $tx, ty, tz, \theta x, \theta y, \theta z$ )	Calcul (s) $0.06 \pm 0.02$	Param. (M) 71	p-valeur (vs SpT)
SpT-Net (proposée)	<b><math>2.21 \pm 2.42</math></b>	<b><math>0.93 \pm 0.09</math></b>	( $0.98, 0.98, 0.99,$ $0.99, 0.98, 0.94$ )	$0.06 \pm 0.02$	71	
FVR-Net [56]	$4.36 \pm 12.18$	$0.85 \pm 0.16$	( $0.83, -0.78, 0.59,$ $0.50, -2.99, -10$ )	$0.17 \pm 0.06$	68	$p \leq 0.001$
Pre-Net [103]	$2.26 \pm 3.16$	$0.81 \pm 0.12$	( $0.94, 0.98, 0.99,$ $0.99, 0.97, 0.89$ )	$0.03 \pm 0.01$	41	$p \leq 0.001$

résultats sont ensuite moyennés pour tous les groupes. Quelques variations dans les résultats sont alors obtenus pour le modèle *SpT-Net* comparativement à la Table V.2 (sans validation croisée).

On remarque que les deux études *Pre-Net* (travaux préliminaires du chapitre IV) et *SpT-Net* (méthode proposée), semblent plus précises que l'étude *FVR-Net*. Exceptées les différences liées aux architectures, ces deux études bénéficient des informations positionnelles (de type  $T_{prior}$  et  $T_{suivi}$ ) alors que l'expérience *FVR-Net* repose sur l'entrée spatiale *partielle* (sous-volume) et la pénalisation spatiale ( $L_{MSE+SSIM}$ ). Il semble donc que l'ajout d'informations positionnelles permette un meilleur apprentissage que l'ajout d'informations spatiales. En outre, on remarque que même notre scénario *Im* (*local* sans  $T_{suivi}$ , Table V.2) obtient des performances similaires, voire meilleures, que celles de la méthode *FVR-Net*. Ceci démontre des résultats compétitifs de notre méthode, même incomplète.

Par rapport à l'étude préliminaire *Pre-Net* (Chapitre IV), notre modèle a bénéficié d'une nouvelle architecture basée sur des convolutions 3D. Nous obtenons de meilleurs résultats, dans les deux jeux de données proposés, ce qui suggère une architecture de réseau plus appropriée. Ces conclusions sont similaires à celles obtenues lors de la comparaison des architectures (voir section V.4.2).

Pour finir, notre approche *SpT-Net* surpasse de loin les autres méthodes de la littérature. Pour les deux ensembles de données, les résultats du *FVR-Net* sont loin des exigences cliniques attendues en termes de précision ( $TRE \geq 2.5$  mm) et même en temps de calcul ( $\geq 50$ ms). Cela peut être dû aux opérations coûteuses liées à l'orientation et au rééchantillonnage des sous-volumes à chaque nouvelle prédiction et à leur architecture complexe de réseau. Notre approche *SpT-Net*, quant à elle, répond à ces deux conditions cliniques et pour les deux jeux de données.

### V.4.3.3 Comparaison qualitative

La Figure V.6 présente les résultats qualitatifs, complémentaires de l’analyse quantitative précédente, permettant de mieux interpréter l’impact des erreurs directement sur l’organe.

L’exemple choisi illustre un cas typique avec une valeur de TRE proche des résultats moyens que nous avons obtenus (Table V.5). Nous observons des résultats d’images très similaires pour *SpT-Net* par rapport aux autres méthodes, avec une superposition globale du maillage satisfaisante.

Les erreurs sur l’ensemble de données  $I_{i,t}$  à  $V_j$  sont plus importantes en raison de la difficulté à inclure les déformations de la prostate entre les images. Les métriques de similarité (NCC et SSIM) présentent des valeurs inférieures, qui se justifient pas la différence de contenu iconique entre les deux images issues d’anatomies différentes. On peut remarquer sur cet exemple, que les valeurs de SSIM pour la méthode *FVR-Net* sont meilleures que les autres méthodes, même si l’alignement général est moins bon. Cette méthode utilisant un terme de pénalisation spatiale basé sur la SSIM, on peut supposer qu’elle a appris à mieux pénaliser ce terme même s’il n’est pas suffisant pour garantir un meilleur alignement.

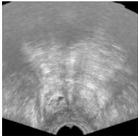
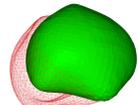
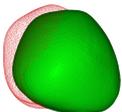
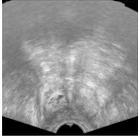
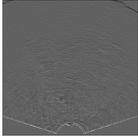
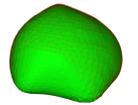
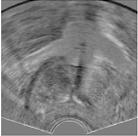
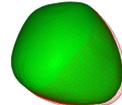
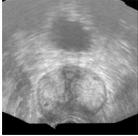
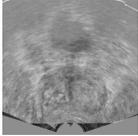
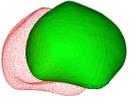
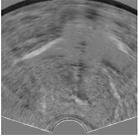
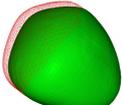
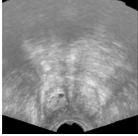
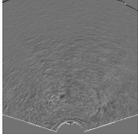
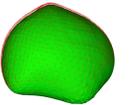
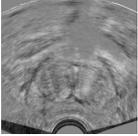
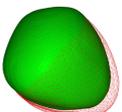
	$I_{i,t}$ à $V_i$			$I_{i,t}$ à $V_j$		
	Images	Différences de pixels	Maillage superposés	Images	Différences pixels	Maillage superposés
<b>Image courante</b>				<b>Image courante</b>		
<b>SpT-Net</b> TRE = 0.19 NCC = 0.99 SSIM = 0.90				<b>SpT-Net</b> TRE = 2.25 NCC = 0.84 SSIM = 0.15		
<b>FVR-Net</b> TRE = 11.5 NCC = 0.89 SSIM = 0.12				<b>FVR-Net</b> TRE = 3.67 NCC = 0.88 SSIM = 0.18		
<b>Pre-Net</b> TRE = 0.90 NCC = 0.98				<b>Pre-Net</b> TRE = 4.89 NCC = 0.82		

FIGURE V.6 – Exemple illustratif de la qualité du recalage des images US pour les jeux de données  $I_{i,t}$  à  $V_i$  et  $I_{i,t}$  à  $V_j$ . Comparaison de la superposition des maillages de la prostate de référence (vert) et prédit (rouge).

### V.4.3.4 Comparaison temporelle

Pour compléter notre analyse, nous avons comparé les évolutions temporelles de la qualité des prédictions, pour les différentes méthodes de la littérature, au cours d'une trajectoire type de balayage base-apex de nos deux jeux de données. Les Figures V.7 illustrent l'évolution des paramètres de transformation et l'évolution de l'erreur avant/après recalage sur une trajectoire. Nous rappelons que les allures des trajectoires obtenues ont été discutées lors du protocole de génération des données (voir section III.1.3.3).

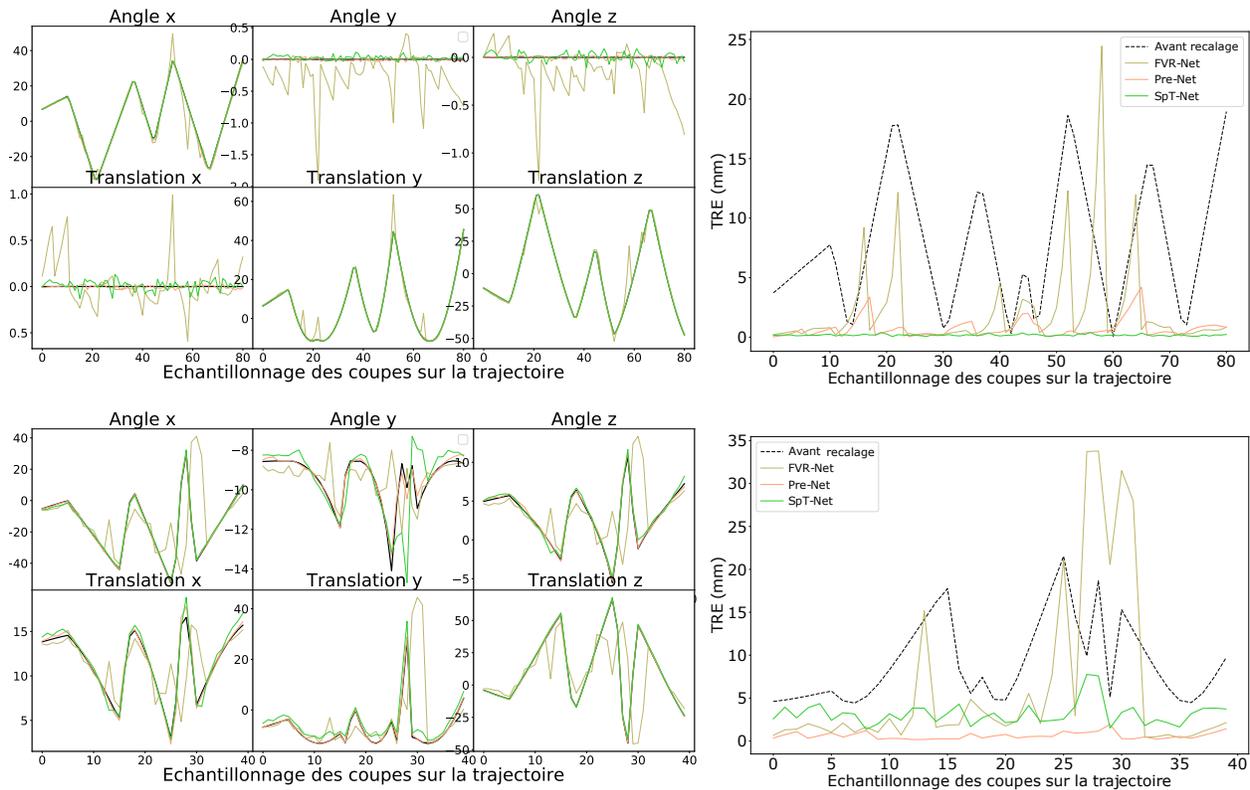


FIGURE V.7 – Comparaison temporelle le long d'une trajectoire entre différentes méthodes de la littérature : *FVR-Net* (jaune), *Pre-Net* (orange) et *SpT-Net* (vert). Évolution des paramètres de vérités terrains (courbe noire) et des erreurs avant recalage (pointillé noir). Résultats sur  $I_{i,t} \rightarrow V_i$  en haut et sur  $I_{i,t} \rightarrow V_j$  en bas.

Les résultats sur le jeu de données " $I_{i,t} \rightarrow V_i$ " montrent une bonne reconstruction de la trajectoire sans dérive et une bonne correction des erreurs. Les amplitudes de chaque valeur de paramètre sont bien respectées, même pour les paramètres hors plan ( $\theta_x, \theta_y, t_z$ ) et pour les paramètres sans variation attendue ( $\theta_y, \theta_z, t_x$ ). On remarque que le modèle *FVR-Net* obtient les moins bonnes performances, notamment sur les paramètres sans variation.

Sur le jeu de données " $I_{i,t} \rightarrow V_j$ ", une dérive est plus visible entre l'évolution prédite et l'évolution réelle, notamment pour la méthode *FVR-Net*. Le profil d'évolution des prédictions montre parfois de fortes erreurs, qui sont réajustées plusieurs fois au cours de la trajectoire. Cela reflète un moins bon contrôle de la dérive au cours d'une trajectoire complète. Les deux autres études montrent

un meilleur suivi de la trajectoire, même si les corrections d'erreur sont moins efficaces que sur le premier jeu de données. Ceci peut être expliqué par le niveau de difficulté plus important lié aux anatomies différentes entre les images.

Afin d'avoir une analyse temporelle plus complète, nous avons exploité plusieurs variations de notre protocole d'évaluation séquentielle, présentés sur la Figure V.8. Dans les résultats qui suivent nous avons adapté le nombre de ré-ajustements du  $T_{prior}$ , utilisé soit comme entrée positionnelle pour les expériences *SpT-Net*, soit comme initialisation de l'orientation du sous-volume pour l'expérience *FVR-Net*. Ce nombre de ré-ajustements varie entre  $[0-12, \infty]$ , où :

- 0 représente une expérience séquentielle *complète* ( $T_{prior} = \widehat{T}_{t-1}$  sur la totalité de la trajectoire)
- 12 représente une expérience séquentielle *par portion* standard (schéma de 12 pseudo biopsies)
- et où  $\infty$  représente une expérience séquentielle *théorique* ( $T_{prior} = T_{t-1}$ ).

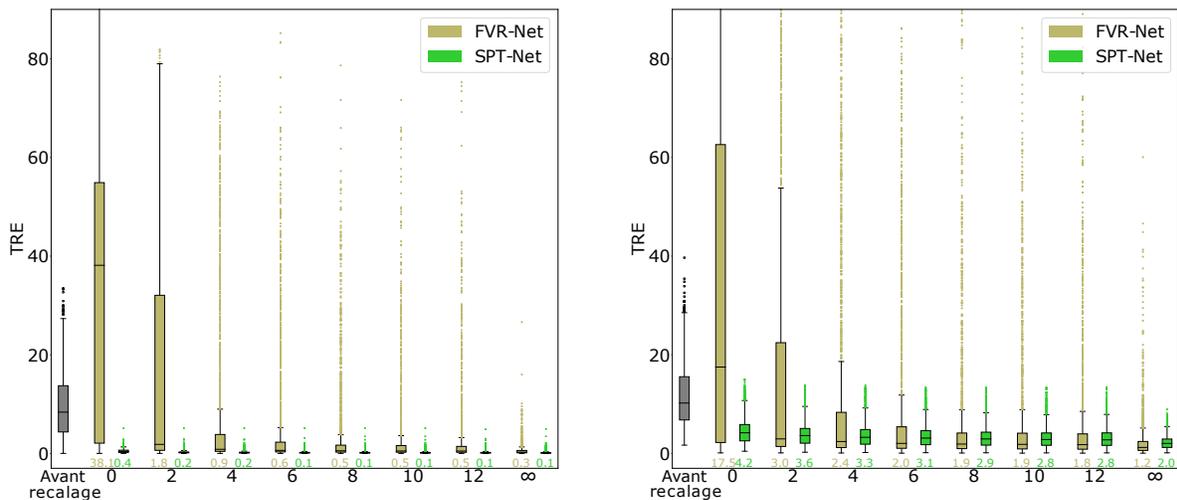


FIGURE V.8 – Variations du nombre de réajustements durant le protocole d'évaluation séquentielle. Comparaison des effets sur les différentes études de la littérature et sur les deux jeux de données. Résultats sur  $I_{i,t}$  à  $V_i$  à gauche et sur  $I_{i,t}$  à  $V_j$  à droite.

On peut constater que les deux approches ne réagissent pas de la même façon face au phénomène de dérive observé : la méthode *FVR-Net* semble beaucoup plus impactée par le changement du nombre de réajustement, pour les deux jeux de données testés. Sans aucun réajustement, la méthode présente des erreurs jusqu'à 38 mm, pires que l'alignement avant recalage. En revanche, la méthode proposée *SpT-Net* présente des performances assez constantes et toujours exploitables même sans aucun réajustement (médianes de 0.4 mm et 4.2 mm, respectivement sur les deux jeux testés).

Il est important de noter que sur une évaluation séquentielle *théorique* (notée " $\infty$ "), les meilleures performances sont obtenues pour la méthode *FVR-Net*. Ces résultats utilisent cependant les valeurs de vérités terrain pour initialiser  $T_{prior}$ , alors que celles-ci ne sont pas disponibles lors d'une pratique réelle clinique. Cela constitue donc un cas d'utilisation idéal mais non réaliste. Ces observations démontrent donc l'importance d'une analyse temporelle et d'un protocole adapté afin d'obtenir les bonnes conclusions.

## V.5 Discussion

### V.5.1 Discussion sur les contributions méthodologiques

#### V.5.1.1 Formes d'informations spatiales

- Les entrées spatiales :

L'ajout d'informations spatiales sous la forme d'entrées additionnelles au réseau (empilement de coupes voisines, volume de référence entier ou sous-volume), ne semble pas faciliter la tâche de recalage à réaliser. Le réseau n'apprend pas d'informations supplémentaires à partir de ces nouvelles entrées, voire au contraire voit son apprentissage perturbé.

Les expériences avec ajout d'une ou plusieurs coupes voisines (expériences *locale-V* avec  $V \geq 2$ ) n'ont pas permis d'obtenir de meilleures performances. Il semble même que plus nous ajoutons des coupes (i.e. des informations contextuelles supplémentaires), moins les performances sont correctes, et ce, quelle que soit l'architecture utilisée (*Main-2D* ou *Main-3D*). Nous pensons qu'un traitement par convolution n'est peut-être pas forcément adapté aux entrées de plusieurs coupes. Il pourrait être envisagé d'utiliser par la suite un réseau spécifique aux séquences de coupes, qui pourrait également combiner d'autres séquences d'entrées (de type  $T_{prior}$ ,  $T_{suivi}$ , etc.).

L'ajout des sous-volumes d'entrée (expérience *partielle*) semble très pénalisante pour le réseau, notamment lors d'une évaluation séquentielle. Par construction, ce type d'entrée n'est pas robuste face aux phénomènes de dérive pouvant se produire dans un cas réel. Nous rappelons que ces sous-volumes sont préalablement centrés et orientés autour de la position de la coupe précédente (initialisée grâce à  $T_{prior}$ ) et peuvent donc présenter un contenu iconique très différent de ce qu'attendu en cas d'erreur de la prédiction. En effet, un phénomène de dérive important peut alors impliquer des sous-volumes loin, voire hors (zones noires) du contenu échographique, impactant la totalité des poids du réseau ayant appris à fortement considérer cette entrée durant l'apprentissage. Cette hypothèse se vérifie lors d'une évaluation séquentielle théorique (voir Figure V.8), où les expériences *partielles* présentent des entrées parfaitement orientées (comme l'apprentissage) et surpassent toutes les autres méthodes.

Relativement à l'entrée *partielle*, l'ajout du volume de référence entier (expérience *globale*), semble une entrée spatiale plus pertinente. En effet, l'utilisation du volume en entier semble beaucoup plus robuste face au phénomène de dérive observé dans les évaluations séquentielles. Même sur des évaluations théoriques (sans dérive), des études d'ablation étendues ont démontré de meilleures performances avec l'entrée *globale* que l'entrée *partielle* (voir Table V.4).

Contrairement à l'entrée *locale*, l'ajout du volume de référence ne permet pas d'améliorer systématiquement et significativement les performances de recalage. Malgré les efforts concernant son intégration à travers une architecture équilibrée, nous pensons que l'association des caractéristiques de nature différente reste difficile. Le paramètre d'extension ne modifie que la dimension des caractéristiques mais ne modifie pas la nature de ce qu'elles représentent (surfaccique/volumique). Un espace de recherche plus grand, avec notamment plus de paramètres à optimiser (par rapport à un réseau à une seule entrée) peut également entraver la convergence rapide et efficace du réseau.

- La pénalisation spatiale :

L'ajout du terme de pénalisation spatiale dans la fonction de coût ( $L_{MSE+SSIM}$ ) n'a, lui non plus, pas entraîné une quelconque amélioration de la qualité du recalage par rapport à la fonction initiale ( $L_{MSE}$ ). Une des raisons possibles est que l'évolution de l'erreur de similarité n'est pas assez marquée et ralentit la bonne convergence du modèle. Nous attribuons ces effets à (i) des mesures entre deux images trop différentes (d'autant plus avec un phénomène de dérive) et à (ii) des mesures sur un contenu iconique trop limité (lié à la problématique d'apparition des zones noires durant le rééchantillonnage des volumes). D'autres études de la littérature mentionnent également l'instabilité d'une convergence basée uniquement sur le terme de similarité [56].

Parmi les différentes pondérations testées pour combiner les deux termes dans la fonction totale, nous avons obtenu les meilleures convergences avec une pondération de 1 pour la MSE et 50 pour la SSIM. Contrairement à la pondération adaptative au cours de l'apprentissage, ce ratio fixe permet de se concentrer sur le terme de similarité d'image seulement après plusieurs itérations, une fois que le terme de MSE a corrigé les plus grandes erreurs et permis d'obtenir deux images suffisamment proches. En effet, la MSE est une fonction qui converge très rapidement et qui obtient des valeurs  $\leq 50$  en moins de 5 itérations. Cela permet donc de privilégier l'optimisation de la similarité seulement dans un deuxième temps, avec un poids plus important tandis que les valeurs de MSE sont déjà bien diminuées.

### V.5.1.2 Confrontation entre les informations positionnelles et spatiales

- Positionnelle VS Spatiale :

Bien que les informations contextuelles spatiales et positionnelles soient redondantes concernant la localisation absolue par rapport à l'espace de référence, il semble que l'ajout plus direct des positions soit plus favorable et notamment plus robuste face aux accumulations d'erreurs. Nous pensons que l'utilisation de  $T_{prior}$  directement sous une forme vectorielle et similaire à l'encodage de sortie attendu ( $T_t$ ), facilite l'apprentissage. Cela peut être comparé, plus ou moins, à l'idée d'une initialisation lors d'un recalage classique. À l'inverse, nous pensons que l'information  $T_{prior}$  une fois utilisée pour orienter le volume, n'est alors plus dissociable de l'image d'entrée, ce qui peut limiter son intérêt.

Le phénomène de dérive semble, quant à lui, impacté par la façon dont l'information  $T_{prior}$  est intégrée dans le réseau. Réorienter et rééchantillonner un sous-volume d'entrée avec  $T_{prior}$  ( $V_{sub,t}$ ) semble être plus sujet à l'accumulation d'erreurs que d'ajouter l'entrée  $T_{prior}$  directement sous forme de vecteur en fin de réseau.

- Sans les informations positionnelles :

Afin d'étudier les contributions des entrées spatiales de manière indépendante et sans aucune interférence avec les informations positionnelles, nous avons comparé des scénarios avec/sans  $T_{prior}$  et avec/sans  $T_{suivi}$ .

Lors des expériences sans  $T_{prior}$ , les résultats démontrent globalement les mêmes tendances et conclusions concernant les informations spatiales : à savoir de moins bonnes performances lors de l'ajout de formes spatiales d'entrées ou de pénalisation. En revanche, les performances obtenues sont encore plus faibles et démontrent, dans ce chapitre encore, l'importance du rôle du  $T_{prior}$ .

Pour toutes les formes d'informations spatiales testées (entrées ou pénalisation), le Scénario  $Im$  obtient les mêmes tendances que précédemment, avec systématiquement de moins bonnes performances que le Scénario  $Im+T_{suivi}$ . Ce constat, déjà observé lors des résultats du chapitre précédent, prouve également l'importance du rôle du  $T_{suivi}$ .

Finalement, les contributions concernant les informations spatiales restent non concluantes, avec/sans informations positionnelles. De même, les contributions des informations positionnelles restent bénéfiques, même lorsque combinées aux informations spatiales. Même en cas d'égalité des performances, la méthode basée sur l'ajout d'informations spatiales requiert des réseaux plus gourmands en paramètres (branches parallèles vs. branche simple), et plus complexes en calculs (rééchantillonnage et pénalisation spatiale) qui peuvent ralentir la convergence et l'apprentissage. Pour toutes ces raisons, nous privilégions une méthode basée sur les informations positionnelles uniquement.

### V.5.1.3 Différentes architectures testées

L'utilisation d'une architecture équilibrée lors de l'exploitation de caractéristiques spatiales de natures différentes (2D ou 3D) est un élément indispensable pour un bon apprentissage du réseau. Des expériences préliminaires avec des architectures déséquilibrées ne permettaient pas d'obtenir la moindre convergence durant l'apprentissage. Avec les deux méthodes envisagées pour équilibrer les représentations, nous avons bien obtenu des convergences satisfaisantes.

Parmi les deux architectures équilibrées testées, il semble que l'utilisation de convolutions 2D puis étendues en convolutions 3D permet d'obtenir un traitement des informations plus pertinent, et ce sur toutes les entrées spatiales envisagées (même *locales-V*). Alors qu'un traitement par convolution 2D moyenne tous les filtres en ne considérant que les déplacements plans, nous pensons qu'un traitement par convolution 3D permet de mieux capturer les déplacements hors-plan à travers une analyse plus complexe et sur plusieurs filtres 3D. Ce type d'approche semble mieux convenir à l'exploitation de la navigation de trajectoire entre les deux images fournies en entrée, dans notre problématique de localisation 2D/3D.

## V.5.2 Discussion sur les contributions expérimentales

### V.5.2.1 Généralisation sur plusieurs niveaux de difficultés

Notre nouvelle génération de données permet de générer deux niveaux de difficultés où les images, rééchantillonnées à partir de balayage base-apex, peuvent être localisées soit dans le même volume ( $I_{i,t}$  à  $V_i$ ), soit dans un volume voisin ( $I_{i,t}$  à  $V_j$ ). Ce deuxième jeu de données permet notamment d'inclure plus de complexité avec des coupes d'entrée indépendantes du volume de référence par rapport auquel nous voulons les localiser et présentant des déformations anatomiques.

Chaque expérience a été réalisée sur chacun de ces jeux de données. Nous avons obtenu un apprentissage plus performant sur l'ensemble de données «  $I_{i,t}$  à  $V_j$  », par transfert d'apprentissage, i.e en réutilisant le modèle pré-entraîné des expériences «  $I_{i,t}$  à  $V_i$  ». À chaque fois, nous avons obtenu les mêmes tendances et les mêmes conclusions, démontrant la robustesse du modèle aux deux niveaux de difficulté. Les erreurs obtenues sur le deuxième jeu de données, sont en revanche plus élevées, ce qui reflète une tâche de recalage plus difficile à réaliser.

Globalement, notre méthode démontre des capacités de généralisation robustes, même sur des tâches plus complexes et plus réalistes.

### V.5.2.2 Importance de l'évaluation séquentielle

Un phénomène de dérive (accumulation d'erreur) peut avoir un impact significatif sur les performances d'un modèle, notamment lorsque celui-ci utilise des résultats précédents (a priori, initialisation d'orientation, etc.). Ne pas évaluer son impact peut conduire à de mauvaises conclusions concernant des choix méthodologiques.

Nous avons montré que les conclusions du modèle *FVR-Net* [56] sont très différentes selon le protocole d'évaluation séquentielle utilisé pour valider les résultats. Lors d'une évaluation séquentielle théorique ( $T_{prior} = T_{t-1}$ ), les sous-volumes sont centrés et orientés autour de la localisation de vérité terrain précédente. Dans ces conditions, ce modèle obtient les meilleures performances par rapport aux autres méthodes proposées, notamment grâce au contexte spatial du sous-volume très proche de l'image à localiser. Nous rappelons que ce type d'évaluation n'est pas représentatif d'une utilisation clinique, où les vérités terrains ne sont pas accessibles. Nous pouvons même noter que l'étude initiale [56] utilise les localisations de vérité terrain bruitées de l'instant  $t$  (que l'on cherche à estimer), ce qui biaise d'autant plus l'évaluation. Par conséquent, lorsque ce modèle est évalué à travers une évaluation séquentielle réaliste, les performances sont très largement impactées par le phénomène de dérive observé. Même des réajustements à des moments clés (séquentielle *par portion*) ne suffisent pas pour contenir la dérive.

Les modèles proposés dans nos études (*SpT-Net* ou *Pre-Net*) sont, quant à eux, plus robustes face au phénomène de dérive. En effet, ils ont été construits et évalués à travers ces évaluations séquentielles, qui ont permis d'établir la meilleure configuration possible.

### V.5.3 Conclusions sur ce deuxième modèle

#### V.5.3.1 Conclusions sur les performances obtenues

La meilleure configuration que nous avons obtenue consiste en la combinaison : d'une entrée spatiale locale (ajout de l'image précédente  $I_{t-1}$ ) associée à l'image courante  $I_t$  ; des entrées positionnelles (ajout des vecteurs  $T_{prior}$  et  $T_{suivi}$ ) ; d'une fonction de coût supervisée basée sur le calcul de MSE. Le tout à travers un réseau convolutif 3D.

Grâce à nos deux jeux de données ( $I_{i,t}$  à  $V_i$  et  $I_{i,t}$  à  $V_j$ ) et à notre protocole de validation séquentielle, nous avons pu évaluer la robustesse de notre modèle sur des tâches de recalage plus complexes et sur des trajectoires complètes, même en présence de phénomènes de dérive.

Les meilleurs résultats obtenus, dans les deux cas, répondent aux exigences cliniques pour la navigation pour la biopsie de la prostate : avec un temps de réponse d'environ 60 ms par image et avec une TRE moyenne inférieure aux 2,5 mm attendus.

#### V.5.3.2 Comparaison avec la littérature

En plus de l'utilisation des a priori positionnels ( $T_{prior}$ ,  $T_{suivi}$ ), spécifique de notre méthode, nous avons montré que la mise en place et l'utilisation de protocoles d'analyse séquentielle adaptés permet de tirer de meilleures conclusions quant aux résultats obtenus et d'établir une configuration plus robuste face aux défis cliniques rencontrés.

Grâce à notre protocole de génération de données sur plusieurs niveaux de difficultés, nous avons pu à la fois nous comparer à des tâches de recalage similaires de la littérature et également évaluer les capacités de généralisation de notre modèle sur des tâches plus complexes et réalistes. Sur les deux ensembles de données évalués ( $I_{i,t}$  à  $V_i$  et  $I_{i,t}$  à  $V_j$ ), notre modèle surpasse de loin les méthodes similaires de la littérature [56] et généralise mieux sur des situations d'assistance à la navigation réalistes, malgré la présence d'un phénomène d'accumulation d'erreur.

#### V.5.3.3 Conclusions sur les objectifs

Les conclusions en réponse aux trois objectifs principaux de cette seconde méthode sont :

- Les expériences menées suggèrent que l'ajout de nouveaux types d'entrées spatiales (empilement de coupes voisines/ volume/sous-volume), ou de pénalisation spatiale (SSIM) ne participent pas à l'amélioration de la tâche de recalage, et ce, malgré l'utilisation d'une nouvelle architecture plus adaptée et équilibrée. La configuration optimale retenue est la combinaison la plus simple des entrées proposées ( $T_{prior}$ ,  $T_{suivi}$ ,  $I_{t-1}$ , CNN 3D).
- Il semble que les informations positionnelles sont plus robustes que les informations spatiales, notamment durant une évaluation séquentielle réaliste présentant un phénomène de dérive. Notre méthode et nos procédés de validation surpassent ceux utilisés dans d'autres études de la littérature et permettent d'obtenir des résultats supérieurs.
- Nous avons vérifié la robustesse de notre méthode sur deux bases de données présentant différents niveaux de difficulté. Cette évaluation est, à notre connaissance, plus robuste que toute autre approche de validation proposée dans la littérature. Cela fait de notre approche un outil prometteur pour l'aide à la navigation.

Il convient désormais d'étudier les autres types d'informations contextuelles de trajectoire, définis dans notre positionnement général de cette thèse. Le chapitre suivant s'intéresse notamment à l'étude de l'ajout d'informations « structurelles ». Il s'agira de trouver de nouvelles manières d'intégrer des informations durant l'apprentissage, pour améliorer les performances de recalage.

#### Points clé

- ❖ Bien que les informations spatiales informent sur le contexte de l'espace de référence, les expériences menées suggèrent que ni l'ajout d'entrées spatiales ni l'ajout de pénalisation spatiale ne participent pas à l'amélioration de la tâche de recalage, et ce, malgré l'utilisation d'une nouvelle architecture plus adaptée et équilibrée.
- ❖ L'ajout d'informations spatiales implique des réseaux plus gourmands et plus complexes en calculs qui peuvent ralentir la convergence et l'apprentissage. De plus, leur forme et leur intégration sont moins adaptées et moins robustes face aux potentiels phénomènes d'accumulation d'erreurs. Nous privilégions donc une méthode basée uniquement sur les informations positionnelles.
- ❖ Notre méthode a été construite et évaluée à travers des évaluations séquentielles réalistes et adaptées aux défis cliniques actuels, et sur des jeux de données présentant plusieurs niveaux de difficultés. Toutes ces conditions nous ont permis de tirer les meilleures conclusions possibles et d'établir une configuration robuste et meilleure que celles observées dans la littérature.
- ❖ Ces travaux ont fait l'objet d'une publication dans le journal "IEEE, Transactions on Bio-medical Engineering" [104].

---

## Recalage 2D/3D avec des informations structurelles

---

*Ce* chapitre propose une troisième évolution de méthode visant à l'amélioration de la tâche de recalage 2D/3D durant la navigation de biopsie prostatique. Cette méthode s'intéresse particulièrement aux informations « structurelles », plus précises et plus spécifiques que les informations spatiales testées dans le chapitre précédent, et qui permettent de considérer des a priori sur les contours de la prostate au sein des images.

Les objectifs principaux de ce chapitre sont d'évaluer et de quantifier l'apport des informations de segmentation sur les performances de recalage 2D/3D en temps-réel. Plusieurs expériences ont été réalisées : selon l'ajout et la combinaison des données d'entrée (étude d'ablation) ; selon le type d'apprentissage envisagé (mono-tâche ou multi-tâche) ; et selon différentes architectures testées. Nous présentons les résultats de ces différentes expériences, avant de conclure quant aux contributions méthodologiques et expérimentales face à nos méthodes précédentes.

---

## Sommaire

---

<b>VI.1 Objectifs et contributions</b> . . . . .	<b>139</b>
<b>VI.2 Matériel &amp; Méthodes</b> . . . . .	<b>140</b>
VI.2.1 Approche proposée . . . . .	140
VI.2.1.1 Définitions et notations . . . . .	141
VI.2.1.2 Architecture complète du réseau . . . . .	141
VI.2.2 Apprentissage . . . . .	142
VI.2.2.1 Apprentissage multi-tâche . . . . .	142
VI.2.2.2 Combinaison et pondération de plusieurs termes . . . . .	143
<b>VI.3 Expériences et configurations</b> . . . . .	<b>144</b>
VI.3.1 Étude d’ablation . . . . .	144
VI.3.2 Différentes architectures testées . . . . .	145
VI.3.3 Jeux de données . . . . .	146
VI.3.4 Paramètres d’apprentissage . . . . .	148
VI.3.5 Métriques de segmentation . . . . .	148
<b>VI.4 Résultats</b> . . . . .	<b>149</b>
VI.4.1 Analyse quantitative . . . . .	149
VI.4.2 Analyse temporelle . . . . .	151
VI.4.3 Analyse qualitative . . . . .	154
VI.4.4 Pondération et apprentissage . . . . .	154
<b>VI.5 Discussion</b> . . . . .	<b>157</b>
VI.5.1 Discussion sur les contributions méthodologiques . . . . .	157
VI.5.1.1 Entrée auxiliaire de segmentation . . . . .	157
VI.5.1.2 Tâche conjointe de segmentation . . . . .	157
VI.5.1.3 Combinaison de deux formes de segmentation . . . . .	158
VI.5.2 Conclusions sur ce troisième modèle . . . . .	159
VI.5.2.1 Conclusions sur les performances obtenues . . . . .	159
VI.5.2.2 Comparaison avec la littérature . . . . .	159
VI.5.2.3 Conclusions sur les objectifs . . . . .	159

---

## VI.1 Objectifs et contributions

Bien que ni l’apport d’entrées spatiales ni l’ajout de pénalisation spatiale, étudiés précédemment, n’aient permis d’améliorer les performances de notre méthode, nous restons motivés par l’idée d’exploiter les informations contextuelles de l’espace de référence. Afin d’explorer d’autres pistes de types d’entrée et de type d’optimisation du réseau, nous avons décidé d’étendre notre démarche aux informations structurelles de segmentation, complémentaires et parfois plus précises que des informations spatiales (voir section II.4.1.4). En particulier, nous pensons qu’apporter des informations sur les contours de la prostate pourrait aider le recalage à se focaliser sur la localisation de l’organe, qui reste notre objectif principal.

La plupart des travaux ayant étudié ces informations de segmentation se concentrent sur des problématiques de recalage multimodal (voir section II.2.1.2), où l’utilisation de caractéristiques de formes anatomiques est plus pertinente et plus robuste que l’utilisation de caractéristiques d’intensité, souvent variables entre différentes modalités d’imagerie. Cependant, à notre connaissance, aucune étude n’a investigué le cas monomodal ou multi-dimensionnel correspondant à notre problématique. Nous souhaitons donc évaluer ces méthodes pour notre cas particulier de recalage 2D(US)/3D(US). De plus, les récentes avancées concernant l’apprentissage multi-tâche semblent encourageantes quant à la faisabilité et l’intérêt d’exploiter de telles informations structurelles (voir section II.4.4.2). Nous souhaitons étudier ce type d’apprentissage et voir s’il peut aller au-delà des limites de nos méthodes précédentes.

Dans ce contexte, les trois principaux objectifs de ce travail sont :

- De vérifier que l’utilisation d’informations structurelles de segmentation (à travers un apprentissage multi-tâche et/ou à travers l’ajout de données d’entrées) répond aux limites des informations spatiales de notre méthode précédente et permet une amélioration du recalage 2D/3D.
- D’établir la configuration optimale de notre nouvelle méthode en choisissant une architecture de réseau adaptée aux tâches conjointes, une technique d’optimisation pertinente pour un apprentissage multi-tâche, et une bonne combinaison des données d’entrées (positionnelles et/ou structurelles).
- De valider la robustesse de notre méthode avec nos procédés de simulation réalistes et nos protocoles d’évaluation séquentielle. Pour cela, nous nous sommes concentrés sur les bases de données provenant des simulations de balayage base-apex (section III.1.3.3).

La section suivante présente la méthode proposée avec les différentes stratégies d’ajout d’informations structurelles à travers un apprentissage et une architecture multi-tâche.

## VI.2 Matériel & Méthodes

### VI.2.1 Approche proposée

Afin d'introduire des informations liées aux contours de la prostate, nous proposons une nouvelle approche de réseau multi-tâche séquentiel illustré sur la Figure VI.1.

Ce réseau *sJSR-Net* (pour "sequential Joint-Segmentation-Registration Network") estime simultanément la segmentation  $\widehat{S}_t$  et la localisation de l'image courante  $\widehat{T}_t$  à partir de l'image d'entrée ( $I_t$ ) et à partir des informations de segmentation et de localisation précédentes (respectivement  $S_{prior}$  et  $T_{prior}$ ). En plus de la branche de recalage principale (bloc A), cette méthode permet de combiner les informations de segmentation sous deux formes : comme tâche conjointe (bloc B) permettant un encodage mutuel et une optimisation partagée; et comme entrée auxiliaire du réseau (bloc C) permettant un a priori de forme. La partie suivante décrit plus en détails ces différents blocs de travail.

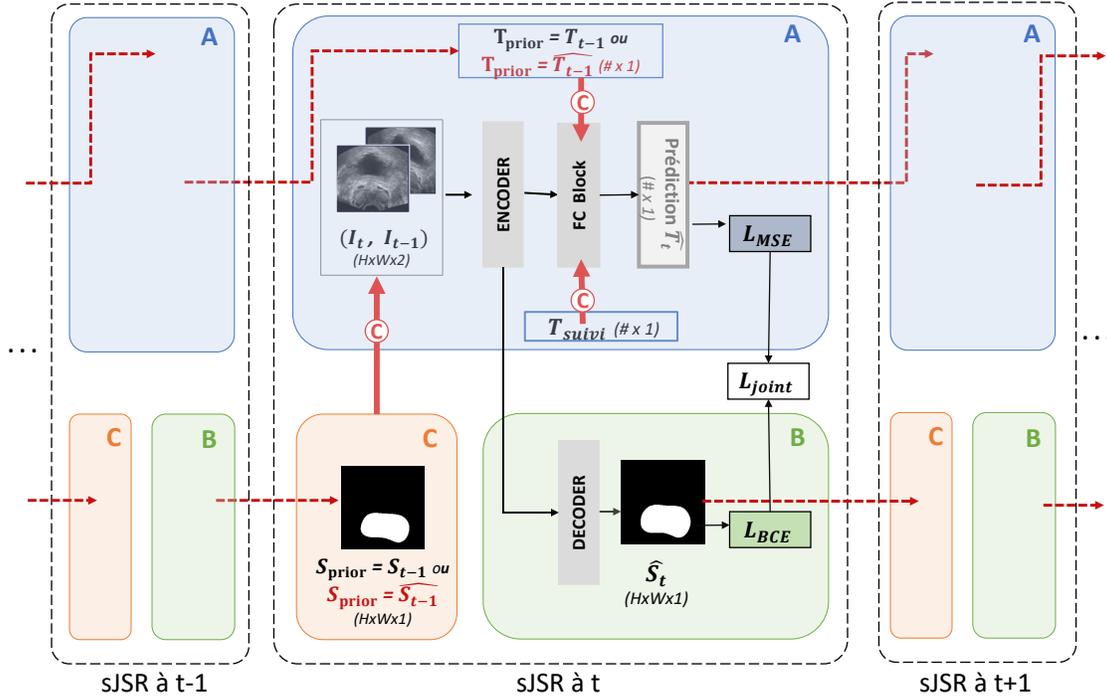


FIGURE VI.1 – Troisième approche proposée : apprentissage multi-tâche de segmentation et recalage d'images. Le réseau proposé, nommé sJSR-Net se compose des différents blocs : branche principale de recalage (bloc A), branche conjointe de segmentation (bloc B), et entrée auxiliaire de segmentation (bloc C). Les connexions séquentielles, valables seulement pour l'inférence, sont illustrées en pointillés rouges entre les prédictions du temps précédent et les entrées du temps suivant ( $S_{prior}$  et  $T_{prior}$ ).

### VI.2.1.1 Définitions et notations

- La branche principale de recalage : Cette partie du réseau correspond à notre méthode précédente, et possède une architecture identique à celle du réseau proposé *SpT-Net* (voir Figure V.1). Nous rappelons que ce réseau *SpT-Net* fait référence à notre meilleure configuration obtenue dans les résultats précédents (entrée spatiale locale  $I_{t-1}$ ,  $T_{prior}$  et  $T_{suivi}$ ). L'architecture d'encodage (rappelée en section VI.2.1.2) se compose de convolutions 2D suivies de convolutions 3D et d'une succession des couches entièrement connectées (« *FC block* ») jusqu'à la prédiction finale  $\widehat{T}_t$ . L'ajout des entrées  $T_{prior}$  et  $T_{suivi}$  reste identique à nos méthodes précédentes, via une concaténation séparée de deux vecteurs de même dimension et de même encodage que la sortie du réseau ( $\# \times 1$ ).

- La branche conjointe de segmentation : Suite à un encodage commun avec la branche principale, le réseau se sépare en une deuxième branche pour réaliser une tâche conjointe de segmentation. Cette seconde branche se compose d'un bloc de décodage (voir section VI.2.1.2) où la dernière fonction d'activation consiste en une fonction "Sigmoid" permettant d'estimer la carte de probabilité de la segmentation prédite  $\widehat{S}_t$ , de même dimension que les images d'entrée ( $H \times W \times 1$ ).

- L'entrée auxiliaire de segmentation : La segmentation 2D associée à la coupe précédente  $I_{t-1}$ , et notée  $S_{prior}$ , est concaténée à la paire d'images de la branche principale. De manière similaire à l'entrée  $T_{prior}$ , l'entrée  $S_{prior}$  peut être fournie grâce aux labels de la coupe précédente ( $S_{prior} = S_{t-1}$ ) dans un cadre séquentiel théorique lors de l'apprentissage ou dans un cadre séquentiel expérimental en utilisant les résultats des prédictions précédentes ( $S_{prior} = \widehat{S}_{t-1}$ ). Cet a priori séquentiel permet à notre réseau d'utiliser les données de segmentation qu'il a lui-même générées au cours de l'inférence précédente, sans nécessiter un réseau supplémentaire de segmentation. Nous rappelons que la manière dont les données de segmentation sont simulées est décrite dans la section III.1.4.1.

### VI.2.1.2 Architecture complète du réseau

L'architecture complète du réseau proposé (sJSR-Net) se compose des différents blocs A+B+C définis précédents et dont une description détaillée est disponible en Annexe A (Table 5).

Parmi les éléments composant le réseau, les blocs « *Encoder* » et « *FC* » ont été gardés de notre méthode précédente (voir section V.2.1.2). Les architectures sont rappelées dans la Figure VI.2.

Pour le bloc « *Decoder* » ajouté dans la branche conjointe de segmentation, nous avons proposé une architecture symétrique à celle de l'encodage de façon à reproduire un réseau de type U-net [105] très fréquent pour les applications de segmentation. Cela permet d'obtenir une architecture complète de type encodeur-décodeur traitant le problème « image-vers-image » en préservant leur format. La dernière fonction d'activation du décodeur consiste en une fonction "Sigmoid" permettant d'estimer la carte de probabilité de la segmentation.

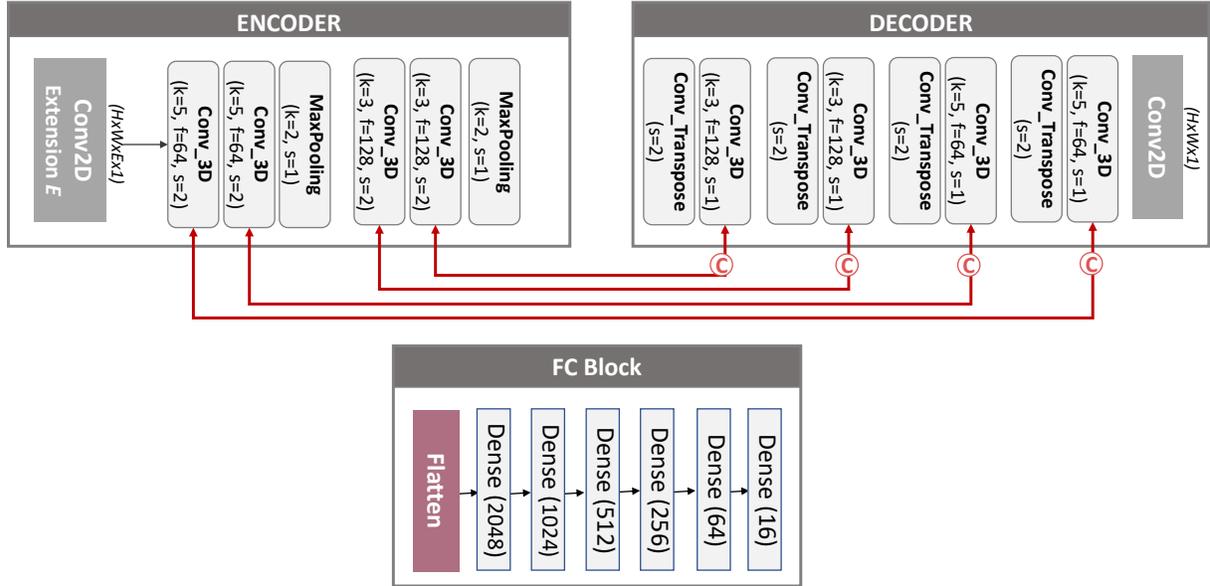


FIGURE VI.2 – Description couche par couche des architectures associées à chaque bloc du réseau proposé (« Encoder », « FC block », « Decoder »).

## VI.2.2 Apprentissage

### VI.2.2.1 Apprentissage multi-tâche

Pour entraîner le réseau multi-tâche proposé, il faut déterminer les fonctions de coût de chaque tâche et la manière de les combiner pour une optimisation commune et partagée.

- Optimisation de la tâche de recalage : Pour l'apprentissage de la tâche de recalage, nous avons utilisé la même fonction supervisée que dans nos méthodes précédentes : basée sur le calcul de MSE (voir l'équation IV.1a en section IV.2.2.1).

- Optimisation de la tâche de segmentation : Pour l'apprentissage de la tâche de segmentation, nous avons utilisé une fonction basée sur le calcul de BCE ("Binary Cross Entropy") couramment utilisée pour les problématiques de classification binaire et particulièrement pour la segmentation d'images médicales [105]. Cette fonction mesure la différence entre deux distributions de probabilités : les probabilités de segmentation prédites par le réseau  $\widehat{S}_k$  et les segmentations de vérité terrain  $S_k$ , pour chaque élément  $k$  de l'entraînement. Le calcul est défini dans l'équation VI.1a. La nature logarithmique favorise la pénalisation des grandes différences de distributions.

Cependant la fonction BCE, généralement basée sur la distribution probabiliste de Bernoulli, fonctionne mieux sur une distribution de données indépendantes et équilibrées entre les deux classe [106]. En effet, en présence de classes déséquilibrées, la fonction a tendance à donner plus d'importance à la classe dominante, ceci pouvant entraîner une mauvaise segmentation des classes minoritaires. Dans notre cas, la proportion de prostate contenue dans les images varie entre 30 et

80% selon le moment du balayage. De plus, la BCE suppose que chaque pixel est indépendant, ce qui n'est pas toujours le cas dans les images médicales avec des corrélations spatiales du contenu iconique voisin. Ces limites sont à garder en tête.

$$L_{BCE} = \frac{1}{K} \sum_{k=1}^K BCE_k \quad (\text{VI.1a})$$

$$\text{avec } BCE_k = S_k * \log(\widehat{S}_k) + (1 - S_k) * \log(1 - \widehat{S}_k) \quad (\text{VI.1b})$$

### VI.2.2.2 Combinaison et pondération de plusieurs termes

Afin de combiner ces termes au sein d'une même fonction de coût pour créer une optimisation conjointe sur les deux tâches apprises, il faut établir la pondération de ces termes dans la même fonction ( $L_{joint}$ ), définie en équation VI.2.

Nous avons envisagé deux cas de pondération des différents termes : adaptative ou fixe. Dans le premier cas, les coefficients de pondération utilisés ( $\alpha, \beta$ ) sont ajustés et normalisés en fonction du rapport des différents termes à combiner. Ce ratio est obtenu après chaque itération d'apprentissage (batch) afin d'avoir constamment un rapport équilibré entre les deux termes et ainsi promouvoir chacune des deux tâches de la même façon.

Dans le second cas, les coefficients de pondération sont fixés à différentes valeurs selon le profil de convergence respectif (rapide ou lent). Cela permet de se concentrer sur la tâche présentant les métriques les plus pénalisantes dans un premier temps, avant d'affiner les deux tâches par la suite.

$$L_{joint} = \alpha * L_{MSE} + \beta * L_{BCE} \quad (\text{VI.2})$$

## VI.3 Expériences et configurations

### VI.3.1 Étude d’ablation

Afin de déterminer indépendamment si les informations de segmentation, sous forme d’une d’entrée auxiliaire (bloc C) ou sous forme d’un apprentissage multi-tâche (bloc B), favorisent ou non la tâche de recalage (bloc A), nous avons réalisé différentes expériences d’ablation, où les contributions de chaque bloc sont ajoutées séparément. Les différentes expériences envisagées selon les combinaisons possibles de ces blocs sont dénommées respectivement :

- Expérience « *SpT* » : avec seulement l’intégration du bloc principal de recalage (bloc A). Cette expérience correspond à notre méthode précédente, sans aucune considération des informations structurelles. Elle sert notamment de contrôle pour vérifier la bonne contribution des formes de segmentation.
- Expérience « *Joint* » : avec l’intégration d’une branche conjointe de segmentation (bloc A+B). Cette expérience permet d’évaluer la contribution d’un apprentissage multi-tâche.
- Expérience « *Aux* » : avec l’intégration d’une entrée auxiliaire de segmentation (bloc A+C). Pour ce cas-ci, les segmentations doivent être estimées par un réseau supplémentaire de segmentation, noté *PreSeg*, inspiré de l’architecture U-net [105]. Ce réseau a été entraîné, au préalable, à partir de jeux de données spécifiques décrits en section VI.3.3.
- Expérience « *sJSR* » : avec l’intégration complète de tous les blocs proposés (bloc A+B+C). Le réseau se suffit à lui-même puisque toutes les entrées proviennent des prédictions précédentes.

Selon les expériences réalisées, celles-ci sont alors soit mono-tâche (*SpT* et *Aux*) soit multi-tâche (*Joint* et *sJSR*). Pour le cas mono-tâche, seul le terme associé à la tâche principale de recalage est considéré ( $\alpha = 1, \beta = 0$ ), tandis que pour le cas multi-tâche nous avons évalué plusieurs pondérations possibles. Par défaut, les résultats sont présentés pour la pondération ( $\alpha = 1$  et  $\beta = 50$ ) ; sinon une mention particulière sera précisée en description.

Concernant les autres entrées de ce réseau, notamment les entrées positionnelles étudiées dans le chapitre précédent, nous avons décidé de nous concentrer uniquement sur le scénario  $Im + T_{suivi}$  avec  $T_{prior}$ , ayant largement démontré sa supériorité dans les résultats précédents.

Pour finir, le protocole d’évaluation utilisé par défaut dans toutes nos expériences, est le protocole d’évaluation séquentielle *par portion* (voir section III.2.2.2). Pour l’entrée  $T_{prior}$ , des réajustements de prédictions sont effectués à chaque biopsie, grâce aux recalages intégrés au sein des plateformes cliniques [17]. Par contre, aucun réajustement de l’entrée  $S_{prior}$  n’est effectué, puisque la segmentation vérité terrain n’est pas connue en cours d’intervention.

### VI.3.2 Différentes architectures testées

Pour chacune des contributions de segmentations envisagées, nous avons comparé différentes architectures possibles inspirées de la littérature.

- Entrée auxiliaire de segmentation : Parmi les études référencées dans notre état-de-l'art utilisant des informations structurelles en entrée de réseau (voir section II.4.1.4), nous nous sommes inspirés d'intégration par concaténation [43] ou par branche parallèle [38, 48]. Ces deux types d'intégrations ont été adaptés à la configuration de notre réseau et sont présentées sur la Figure VI.3. L'intégration par concaténation, intitulée « *Aux-Concat* » (Figure VI.3-A), correspond à celle déjà présentée précédemment, tandis qu'une intégration par branche parallèle « *Aux-Para* » (Figure VI.3-B) a également été implémentée en comparaison.

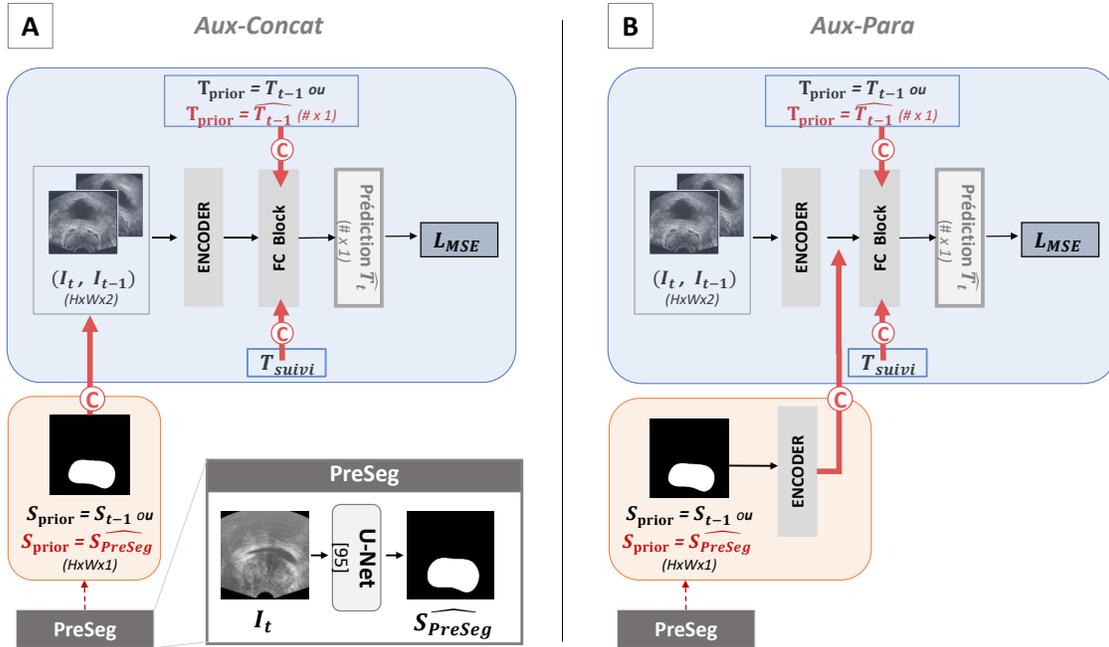
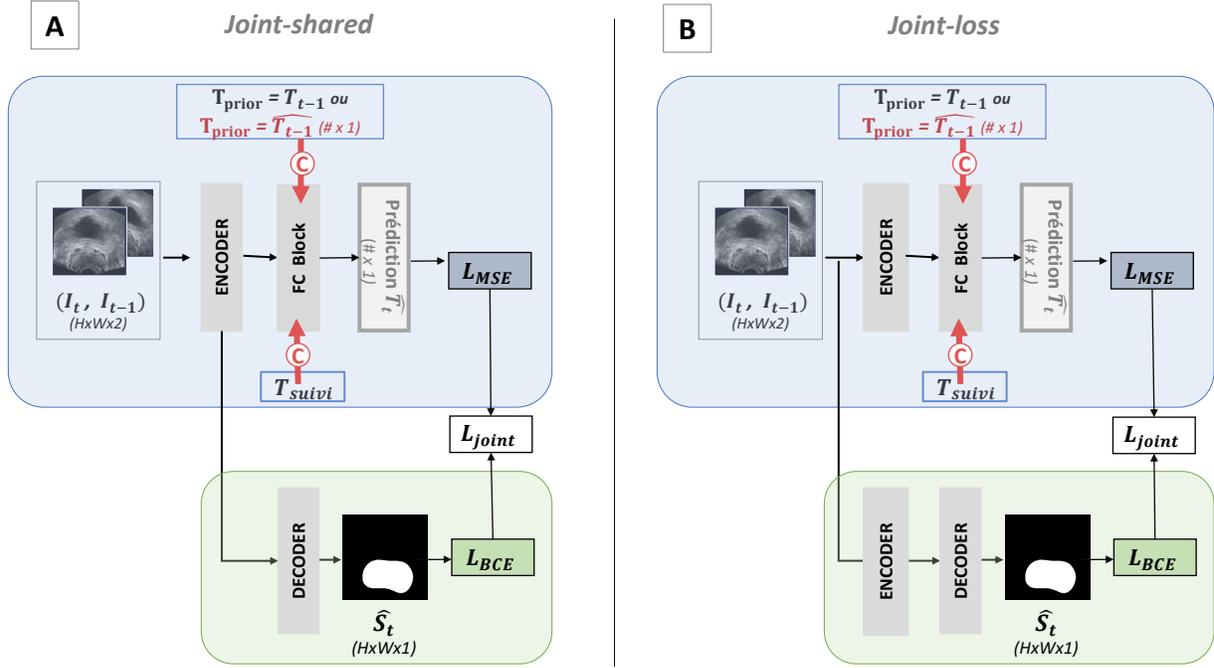


FIGURE VI.3 – Différentes architectures utilisées pour l'expérience *Aux* : *Aux-Concat* (A) *Aux-Para* (B).

- Tâche de segmentation conjointe : Parmi les études référencées dans notre état-de-l'art traitant des problématiques multi-tâches (voir section II.4.2.2), nous nous sommes inspirés d'architectures conjointes au moyen d'un encodage partagé [84] ou du seul partage au niveau de la fonction de coût [39]. Ces deux types d'architectures ont été adaptées à la configuration de notre réseau et sont présentées sur la Figure VI.4. L'expérience intitulée « *Joint-shared* » (Figure VI.4-A) permet un encodage commun, où les poids sont partagés et optimisés de manière similaire considérant les deux tâches tandis que l'expérience « *Joint-loss* » (Figure VI.4-B) présente une architecture parallèle et indépendante, où le seul partage est réalisé par le calcul commun de la fonction de coût. Par ces comparaisons, on peut alors quantifier et dissocier le rôle de la fonction de coût conjointe et/ou de l'architecture partagée lors de l'apprentissage multi-tâche.

FIGURE VI.4 – Différentes architectures utilisées pour l'expérience *Joint* : *Joint-shared* (A) et *Joint-loss* (B).

### VI.3.3 Jeux de données

Pour réaliser les différentes expériences proposées, nous nous sommes appuyés sur la base de données issue du balayage base-apex, dont les caractéristiques sont identiques à celles utilisées lors du chapitre précédent (voir Table V.1). Entre temps, notre base de données a pu être étendue à plus de 1500 volumes de biopsie TRUS provenant de 277 patients, dont 221 ont été utilisés pour les jeux d'entraînement et 56 pour le jeu de test. Pour les jeux de données tests, seuls les volumes  $V_{i=pano}$  ont été utilisés, pour garantir des données de segmentation établies manuellement par des experts (voir section III.1.2). Rappelons que les segmentations des autres volumes sont obtenues par transformation rigide du maillage estimé dans le panorama. Les deux jeux de données ainsi obtenus ( $I_{i,t}$  à  $V_i$  ou  $I_{i,t}$  à  $V_j$ ) avec leur répartition totale d'images sont présentés dans la Table VI.1.

TABLE VI.1 – Description du jeu de données issu du balayage base-apex.

Base de données	Entraînement & Val (82 % & 8 %)	Test (10%)	Caractéristiques
« $I_{i,t}$ à $V_i$ »	5 $V_i$ x 221 patients $\approx$ 55 000 images	$V_{pano}$ x 56 patients $\approx$ 4 700 images	Trajectoire de balayage ( $\pm 40^\circ$ x) Taille des images : $256 \times 256$
« $I_{i,t}$ à $V_j$ »	5 $V_{i,j}$ x 221 patients $\approx$ 52 000 images	5 $V_{pano,j}$ x 56 patients $\approx$ 7 000 images	Difficulté supplémentaire avec image indépendante et anatomie différente

Concernant l'entraînement du réseau supplémentaire de segmentation *PreSeg* utilisé notamment dans les expériences *Aux*, nous avons dû établir un jeu de données spécifique pour cet apprentissage de segmentation 2D. Pour cela, nous avons utilisé uniquement les volumes 3D US  $V_{pano}$  de notre base de données pour lesquels les segmentations 3D  $S_{pano}$  sont disponibles (voir section III.1.2). Sur ces deux volumes (US ou binaire), nous avons échantillonné plusieurs coupes le long des 3 axes des volumes (voir Figure VI.5). Cela permet d'obtenir le couple  $(I_t, S_t)$  représentant respectivement l'image d'entrée et la segmentation de sortie, nécessaire pour l'entraînement du réseau *PreSeg*. La répartition du jeu de données est détaillée dans la Table VI.2.

TABLE VI.2 – Description du jeu de données spécifique pour l'entraînement du réseau *PreSeg*.

Base de données	Entraînement & Val (78 % & 8 %)	Test (17%)	Caractéristiques
« PreSeg »	$V_{pano} \times 80$ patients $\approx 16\,000$ images	$V_{pano} \times 20$ patients $\approx 3\,700$ images	Echantillonnages sur de coupes sur les 3 axes

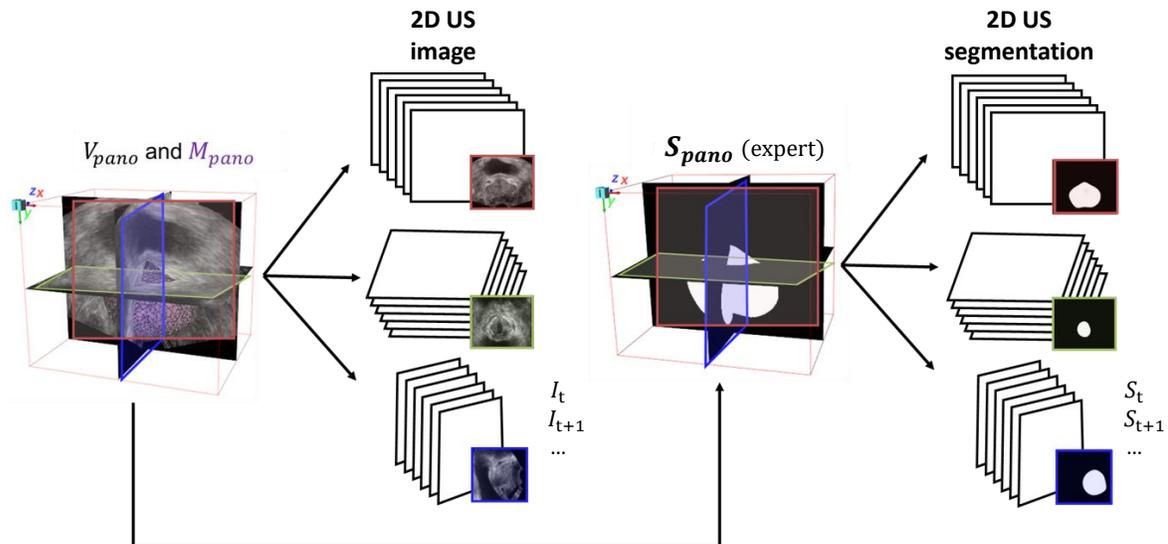


FIGURE VI.5 – Base de données spécifiques pour l'entraînement du réseau *PreSeg*.

### VI.3.4 Paramètres d'apprentissage

Toutes les expériences ont été réalisées en entraînant notre réseau à partir de l'optimiseur Nadam [99] et avec un taux d'apprentissage initial de 1e-4, en utilisant des GPU A100 (NVIDIA) via une implémentation sur TensorFlow [100]. Le réseau a été entraîné sur plus de 30 epochs, avec une taille de lot  $K=10$ . Ces paramètres sont constants entre toutes les expériences réalisées.

Pour ces expériences, les poids du réseau ont tous été initialisés avec une distribution Gaussienne (moy=0, std=0.01). Un apprentissage par transfert du modèle  $I_{i,t}$  à  $V_i$  au modèle  $I_{i,t}$  à  $V_j$  ou encore du modèle  $SpT$  (bloc A) aux modèles  $Aux$  (bloc A+C) et  $Joint$  (bloc A+B) n'a pas pu être réalisé à temps pour être présenté dans ce chapitre.

### VI.3.5 Métriques de segmentation

En plus des métriques définies dans la section III.2.1 permettant d'évaluer la qualité du recalage, dans ce chapitre il convient également d'évaluer la qualité de la segmentation prédite. Parmi les stratégies d'évaluation, il est important de quantifier et de combiner à la fois des mesures de recouvrement, de taille, et de forme entre les segmentations [107, 108]. Deux des mesures d'évaluation les plus utilisées pour la segmentation d'images médicales sont : le coefficient de similarité Dice (DSC) et la distance de Hausdorff (HD). Ces métriques sont complémentaires et possèdent leurs propres avantages et limites. Nous avons donc utilisé ces deux mesures pour retranscrire la qualité des segmentations prédites.

Le coefficient DSC est une mesure de similarité qui mesure la superposition des surfaces entre la segmentation prédite  $\widehat{S}_t$  et la vérité de terrain  $S_t$ . Cette métrique est définie en équation VI.3, et sa mesure est comprise entre 0 et 1 où la valeur 1 indique une superposition parfaite entre les résultats de segmentation. Cependant, cette mesure ne permet pas d'évaluer les différences de forme ni l'emplacement de la segmentation prédite.

$$DSC = \frac{2 * |\widehat{S}_t \cap S_t|}{|\widehat{S}_t| + |S_t|} \quad (VI.3)$$

La distance de Hausdorff mesure la distance maximale (en millimètres) entre deux ensembles de points (voir équation VI.4). Cette mesure est plus sensible à la localisation spatiale des erreurs de segmentation et permet de mieux pénaliser la forme des contours. Cependant, elle peut être coûteuse en terme de calcul, en particulier pour des images de grande taille ou des tâches de segmentation complexes.

$$HD = \max \left( \sup_{a \in S_t} \inf_{b \in \widehat{S}_t} d(a, b), \sup_{b \in \widehat{S}_t} \inf_{a \in S_t} d(a, b) \right) \quad (VI.4)$$

## VI.4 Résultats

### VI.4.1 Analyse quantitative

Nous présentons dans un premier temps les résultats quantitatifs obtenus pour les différentes expériences testées : selon les deux formes de segmentation comparées (entrée auxiliaire ou sortie multi-tâche) et selon les architectures utilisées. Les Tables VI.3 et VI.4 résumant l'ensemble des performances de recalage et de segmentation sur les deux jeux de données ( $I_{i,t}$  à  $V_i$  et  $I_{i,t}$  à  $V_j$ , respectivement). La Figure VI.6 illustre les distributions des différentes métriques utilisées.

Pour les résultats sur  $I_{i,t}$  à  $V_i$ , on remarque tout d'abord que toutes les configurations testées présentant diverses formes de segmentation obtiennent de meilleures performances que la méthode précédente *SpT* (sans aucune information de segmentation). Cela démontre la bonne contribution des informations de segmentation, sous forme d'entrée ou de sortie. Ces résultats sont d'ailleurs statistiquement significatifs (voir avant-dernière colonne Table VI.3).

Les expériences *Aux* ne produisent qu'une faible amélioration des performances de recalage. Entre les deux architectures testées, il semble que l'architecture *Aux-Concat*, avec moitié moins de paramètres à optimiser (71M vs 139M), soit à privilégier par rapport à l'architecture *Aux-Para*.

Les autres expériences (*Joint-shared*, *Joint-loss* et *sJSR*) présentent les plus grandes améliorations de recalage, avec jusqu'à 97% des résultats sous le seuil d'erreur à atteindre ( $TRE < 2.5\text{mm}$ ). Elles bénéficient toutes d'un apprentissage multi-tâche qui semble favoriser les performances par rapport à un apprentissage mono-tâche, à la fois pour le recalage (*SpT* et *Aux*) et pour la segmentation (*PreSeg*). La comparaison entre les deux architectures *Joint-shared* et *Joint-loss* démontre un meilleur apprentissage lorsqu'une partie de l'architecture est partagée.

Il est également important de noter que la majorité des tests statistiques réalisés ont révélé des résultats significatifs ( $p < 0.05$ ) mais pas très significatifs ( $p > 0.01$ ). Cela témoigne d'un niveau d'apprentissage et de performance assez similaires.

Concernant les résultats sur  $I_{i,t}$  à  $V_j$ , on observe également de meilleures performances pour les expériences *sJSR*, avec une correction de recalage pouvant atteindre 92% des erreurs sous la limite clinique. Cette expérience multi-tâche présente également de meilleures performances comparativement aux expériences mono-tâche de recalage (*SpT* ou *Aux*) ou mono-tâche de segmentation (*PreSeg*). En revanche, relativement à  $I_{i,t}$  à  $V_i$ , seule l'expérience combinant les deux contributions de segmentation *sJSR* présente des améliorations tandis que les expériences *Aux* ou *Joint* (contribution unique de segmentation) ne produisent plus d'amélioration des performances.

Pour finir, la diminution d'erreur (après recalage) est moins importante sur le jeu  $I_{i,t}$  à  $V_j$  que sur le jeu  $I_{i,t}$  à  $V_i$ , illustrant une tâche de recalage plus complexe à réaliser. En revanche, concernant les performances de segmentation (uniquement pour les expériences *Joint* concernées), on constate que les résultats obtenus entre les deux jeux sont similaires (voir Figure VI.6-b). Cela démontre qu'une moins bonne généralisation du réseau à des niveaux plus difficiles pour une tâche en particulier ne détériore pas l'intégralité du réseau ni son apprentissage sur la deuxième tâche malgré le partage des poids et l'optimisation commune.

TABLE VI.3 – Comparaison des performances de recalage et de segmentation pour le jeu de données  $I_{i,t}$  à  $V_i$ , à travers une évaluation séquentielle par portion, et pour différentes expériences. Les résultats statistiques entre les expériences et les scénarios  $sJSR$  ou  $SpT$  sont présentées dans les deux dernières colonnes (\* si p-valeur  $<0.05$ ). N.B : Seul le scénario  $Im + T_{suivi}$  avec  $T_{prior}$  est évalué.

Éval. séquentielle $I_{i,t}$ à $V_i$	Recalage ( $\mu \pm \sigma$ )			Segmentation ( $\mu \pm \sigma$ )		p-valeur (sJSR)	p-valeur (SpT)	Param. (M)
	TRE ( $\downarrow$ )	TRE $<2.5$ mm	SSIM ( $\uparrow$ )	HD ( $\downarrow$ )	DSC ( $\uparrow$ )			
Avant recalage								
	$9.13 \pm 6,50$	15.5%	$0.23 \pm 0.15$	-	-	*	*	
Approche précédente (sans information structurale)								
$SpT$	$0.63 \pm 2.47$	96.4%	$0.84 \pm 0.23$	$\emptyset$	$\emptyset$	*	-	71
Expériences <i>Aux</i>								
<i>Aux-Concat</i>	$0.61 \pm 2.66$	95.3%	$0.88 \pm 0.22$	$\emptyset$	$\emptyset$	*	*	71
<i>Aux-Para</i>	$0.62 \pm 3.41$	96.3%	$0.91 \pm 0.19$	$\emptyset$	$\emptyset$	*	*	139
<i>PreSeg</i>	$\emptyset$	$\emptyset$	$\emptyset$	$8.23 \pm 7.06$	$0.83 \pm 0.20$	$\emptyset$	$\emptyset$	1.9
Expériences <i>Joint</i>								
<i>Joint-shared</i>	$0.43 \pm 1.57$	97.1%	$0.86 \pm 0.21$	$6.49 \pm 4.37$	$0.86 \pm 0.16$	*	*	75.7
<i>Joint-loss</i>	$0.58 \pm 2.48$	95.6%	$0.88 \pm 0.22$	$6.58 \pm 4.60$	$0.86 \pm 0.16$	*	*	77
Approche proposée (complète)								
$sJSR$	$0.51 \pm 2.12$	96.4%	$0.86 \pm 0.21$	$4.17 \pm 3.65$	$0.89 \pm 0.19$	-	*	76

TABLE VI.4 – Comparaison des performances de recalage et de segmentation pour le jeu de données  $I_{i,t}$  à  $V_j$ , à travers une évaluation séquentielle par portion, et pour différentes expériences. Les résultats statistiques entre les expériences et les scénarios  $sJSR$  ou  $SpT$  sont présentées dans les deux dernières colonnes (\* si p-valeur  $<0.05$ ). N.B : Seul le scénario  $Im + T_{suivi}$  avec  $T_{prior}$  est évalué.

Éval. séquentielle $I_{i,t}$ à $V_j$	Recalage ( $\mu \pm \sigma$ )			Segmentation ( $\mu \pm \sigma$ )		p-valeur (sJSR)	p-valeur (SpT)	Param. (M)
	TRE ( $\downarrow$ )	TRE $<2.5$ mm	SSIM ( $\uparrow$ )	HD ( $\downarrow$ )	DSC ( $\uparrow$ )			
Avant recalage								
	$13.6 \pm 5,87$	15.5%	$0.14 \pm 0.06$			*	*	
Approche précédente (sans information structurale)								
$SpT$	$1.22 \pm 1.21$	90.2%	$0.27 \pm 0.13$	$\emptyset$	$\emptyset$	*	-	71
Expériences <i>Aux</i>								
<i>Aux-Concat</i>	$1.44 \pm 1.46$	86.5%	$0.26 \pm 0.12$	$\emptyset$	$\emptyset$	*	0.63	71
<i>Aux-Para</i>	$1.77 \pm 1.78$	79.4%	$0.26 \pm 0.13$	$\emptyset$	$\emptyset$	*	*	139
<i>PreSeg</i>	$\emptyset$	$\emptyset$	$\emptyset$	$7.74 \pm 6.74$	$0.84 \pm 0.19$	$\emptyset$	$\emptyset$	1.9
Expériences <i>Joint</i>								
<i>Joint-shared</i>	$1.43 \pm 1.47$	86.7%	$0.26 \pm 0.13$	$5.81 \pm 4.05$	$0.88 \pm 0.14$	*	*	75.7
<i>Joint-loss</i>	$1.21 \pm 1.34$	90.0%	$0.27 \pm 0.13$	$6.48 \pm 4.69$	$0.87 \pm 0.14$	*	*	77
Approche proposée (complète)								
$sJSR$	$1.08 \pm 1.06$	92.2%	$0.27 \pm 0.13$	$4.11 \pm 2.77$	$0.90 \pm 0.15$	-	*	76

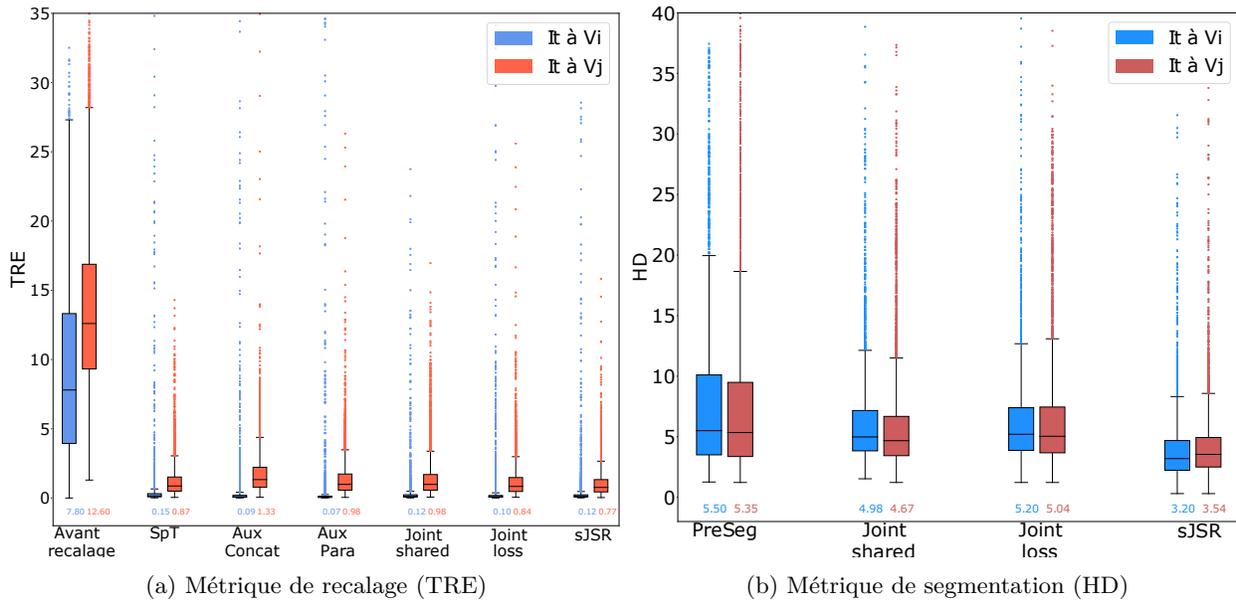


FIGURE VI.6 – Comparaison des distributions (et leur médiane) des métriques de recalage et de segmentation selon différentes expériences réalisées et sur les deux jeux de données  $I_{i,t}$  à  $V_i$  et  $I_{i,t}$  à  $V_j$ .

## VI.4.2 Analyse temporelle

L'analyse des résultats a également été étendue à une analyse temporelle. Les Figures VI.7 et VI.8 présentent l'évolution au cours d'une trajectoire des paramètres de transformation et l'évolution des métriques d'évaluation de recalage et de segmentation, respectivement sur  $I_{i,t}$  à  $V_i$  et  $I_{i,t}$  à  $V_j$ . Les quatre principales expériences *SpT* (jaune), *Aux-concat* (orange), *Joint-shared* (vert) et *sJSR* (bleu) y sont comparées.

Globalement, les atténuations d'erreurs de recalage et les suivis de trajectoire sont satisfaisants. La méthode *SpT* présente des déviations parfois plus marquées, mais suffisamment contenues grâce aux réajustements du  $T_{prior}$  par portion. Les expériences *sJSR* et *Joint-shared* montrent de meilleures performances.

L'expérience *sJSR* produit des erreurs de segmentation de plus en plus importantes (métrique de HD croissante, métrique de DSC décroissante) mais qui restent inférieures comparativement à l'expérience *Joint-shared* dont les erreurs sont plus grandes mais plus constantes. Cela reflète un phénomène d'accumulation d'erreur impliquée par l'entrée  $S_{prior}$ , fournie à partir des prédictions du réseau à l'étape précédente. Ces segmentations ne peuvent pas être corrigées et impliquent ainsi une dérive tout au long de la trajectoire pouvant impacter la totalité du réseau et ses performances de recalage.

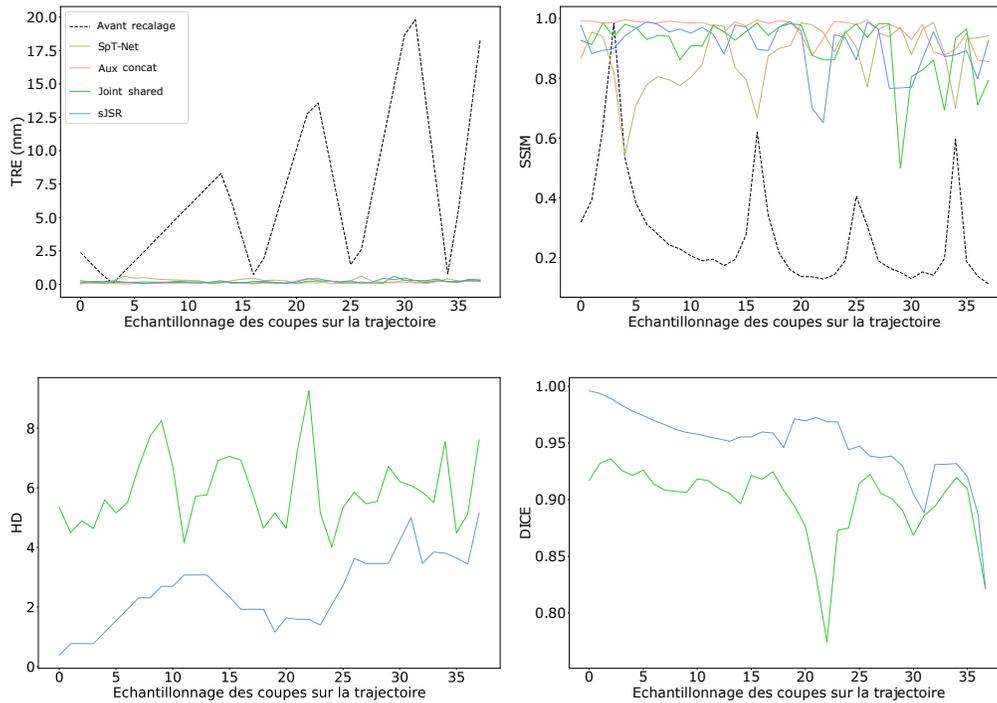
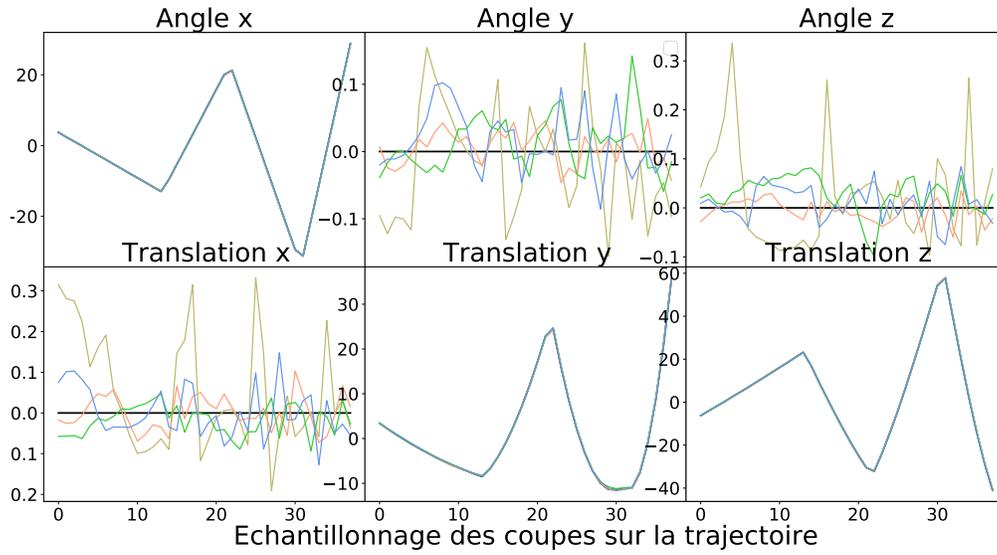


FIGURE VI.7 – Comparaison temporelle le long d’une trajectoire sur  $I_{i,t}$  à  $V_i$  entre différentes expériences : *SpT* (jaune), *Aux-concat* (orange), *Joint-shared* (vert) et *sJSR* (bleu). Évolution des paramètres de vérités terrains (courbe noire) et des erreurs avant recalage (pointillé noir).

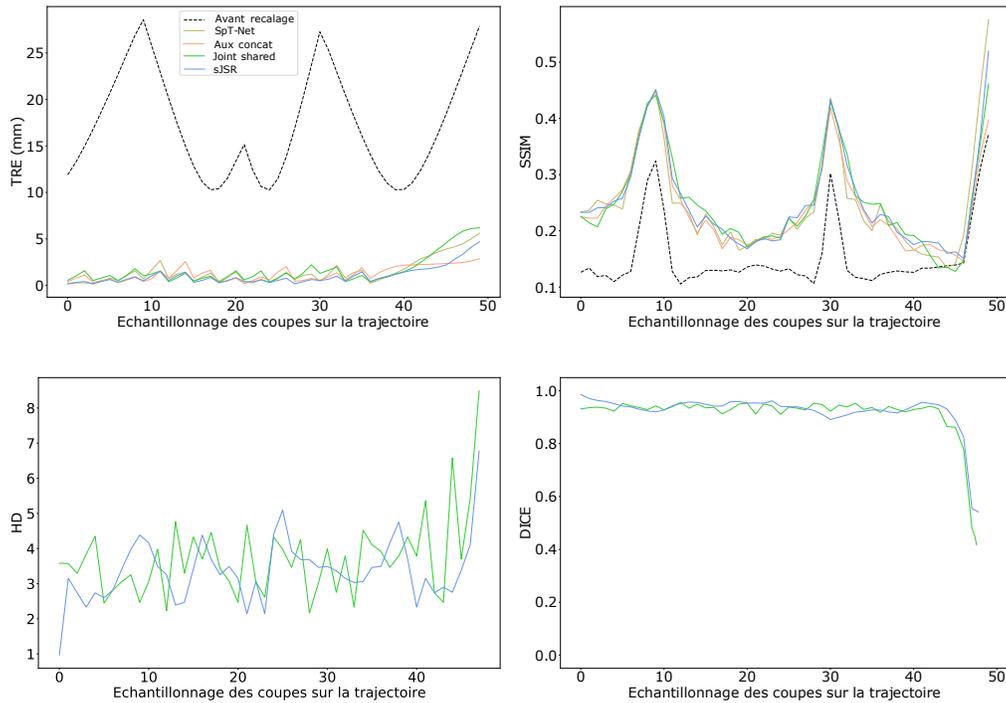
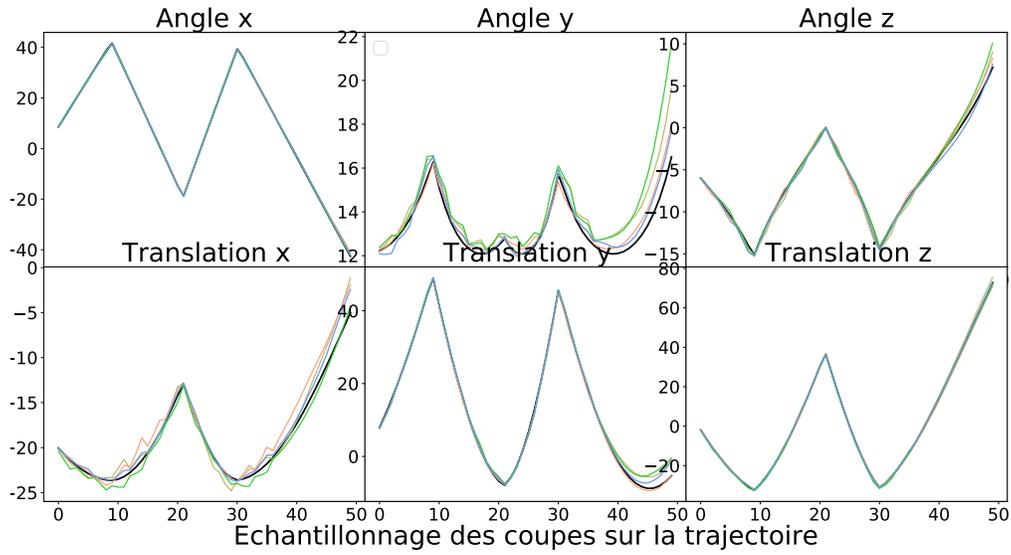


FIGURE VI.8 – Comparaison temporelle le long d’une trajectoire sur  $I_{i,t}$  à  $V_j$  entre différentes expériences : *SpT* (jaune), *Aux-concat* (orange), *Joint-shared* (vert) et *sJSR* (bleu). Évolution des paramètres de vérités terrains (courbe noire) et des erreurs avant recalage (pointillé noir).

### VI.4.3 Analyse qualitative

La Figure VI.9 présente plusieurs exemples de cartes d'activation obtenues pour les différentes expériences réalisées. Nous rappelons, que ces cartes d'activation mettent en évidence les régions de l'image ayant le plus participé à la prédiction du réseau (voir section IV.3.2). Dans ces résultats, seules les cartes calculées à partir de l'algorithme Grad-CAM en utilisant le paramètre  $\theta_x$  comme classe d'intérêt sont présentées. L'étude des autres paramètres aboutit aux mêmes types de cartes.

Sur les cartes obtenues pour l'expérience *SpT* (sans aucune information de segmentation), l'attention du réseau semble se diriger vers les régions de l'image présentant de forts contrastes et de fortes modifications iconiques. Ces régions se situent sur la totalité de l'image, à l'intérieur ou à l'extérieur de la région de la prostate, en fonction des pixels ayant fortement été impactés par les déplacements. On remarque également qu'en présence de zones noires (voir colonnes spécifiques), ce réseau semble particulièrement impacté et se concentre sur du contenu "hors-échographique".

Pour les expériences multi-tâche (*Joint-shared* et *sJSR*), l'attention se porte sur les contours de la prostate. Cela permet d'attirer l'attention sur des repères anatomiques importants et pouvant favoriser la tâche de recalage. On remarque que le *Joint-shared* combine les régions d'attention liées à la segmentation aux autres régions initialement détectées. En revanche, pour l'expérience *sJSR*, il semble que le réseau ne focalise son attention que sur ces contours, négligeant les autres informations iconiques. On peut également constater que ces deux réseaux sont moins impactés par la présence de zones noires et dirigent toujours leur attention vers les bonnes zones d'intérêt.

Pour finir, des exemples de visualisation des segmentations prédites selon les différentes expériences sont présentées dans la Figure VI.10 et comparées aux segmentations de vérité terrain (blanches). On constate de meilleures prédictions pour les deux réseaux multi-tâche (*sJSR* et *Joint-shared*) que pour le réseau mono-tâche *PreSeg*. Sur ces données de test, rappelons que seuls les volumes  $V_{pano}$  disposant de l'estimation des segmentations par des experts sont utilisés. Malgré leur expertise, on remarque que certaines segmentations restent à interpréter avec précaution.

### VI.4.4 Pondération et apprentissage

Plusieurs pondérations et associations de deux différentes tâches entraînées (segmentation et recalage) ont été testées pour les expériences concernées. Nous avons comparé une pondération avec des coefficients fixes entre les deux tâches et une pondération adaptative qui normalise les deux tâches. Seule la moitié de notre jeu de données (voir Table VI.1) a été utilisée afin de réduire les temps d'entraînement et de multiplier les expériences testées. La Table VI.5 en présente certaines. Les évolutions des profils de convergence durant l'entraînement sont illustrées dans la Figure VI.11.

Les performances obtenues sont très similaires entre toutes les expériences réalisées, avec une légère supériorité (et statistiquement significative) de la pondération ( $\alpha = 1$ ,  $\beta = 50$ ). Même si les résultats présentent de moins bonnes corrections que pour les bases de données complètes (voir Table VI.3), les performances sont toujours satisfaisantes et reflètent un bon apprentissage quelle que soit la pondération choisie. Les différents profils de convergence illustrent notamment ce bon apprentissage, avec une diminution de la fonction de coût progressive, stable et sans phénomène de sur-apprentissage. Il semble donc qu'un apprentissage adapté et robuste a pu être établi pour chaque expérience testée, indépendamment de la pondération choisie.

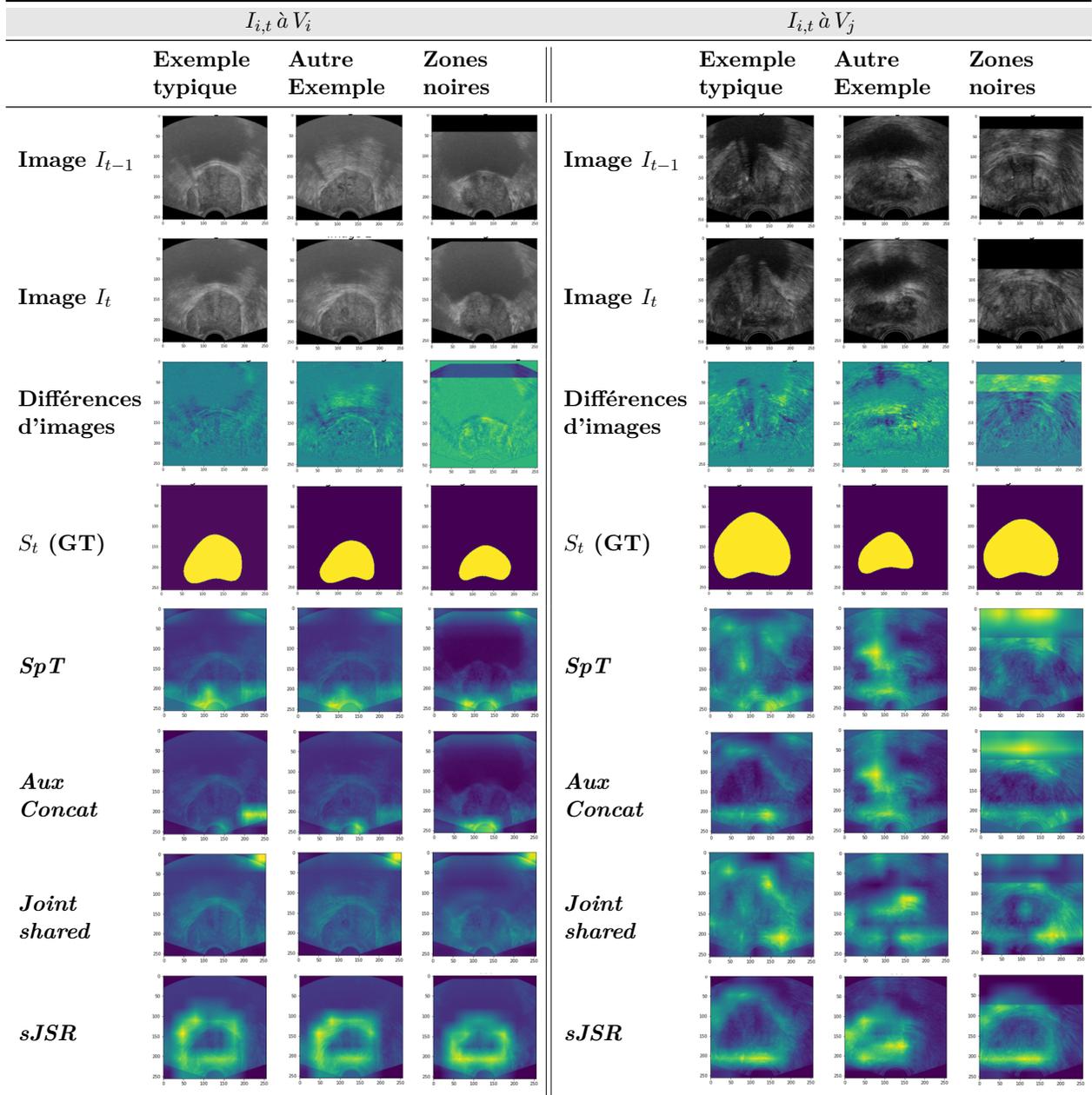


FIGURE VI.9 – Exemples de visualisation des cartes d'activation moyenne obtenues par le réseau selon les différentes expériences ( $SpT$ , *Aux-Concat*, *Joint-shared*, *sJSR*). Résultats sur  $I_{i,t}$  à  $V_i$  (à gauche) et  $I_{i,t}$  à  $V_j$  (à droite).

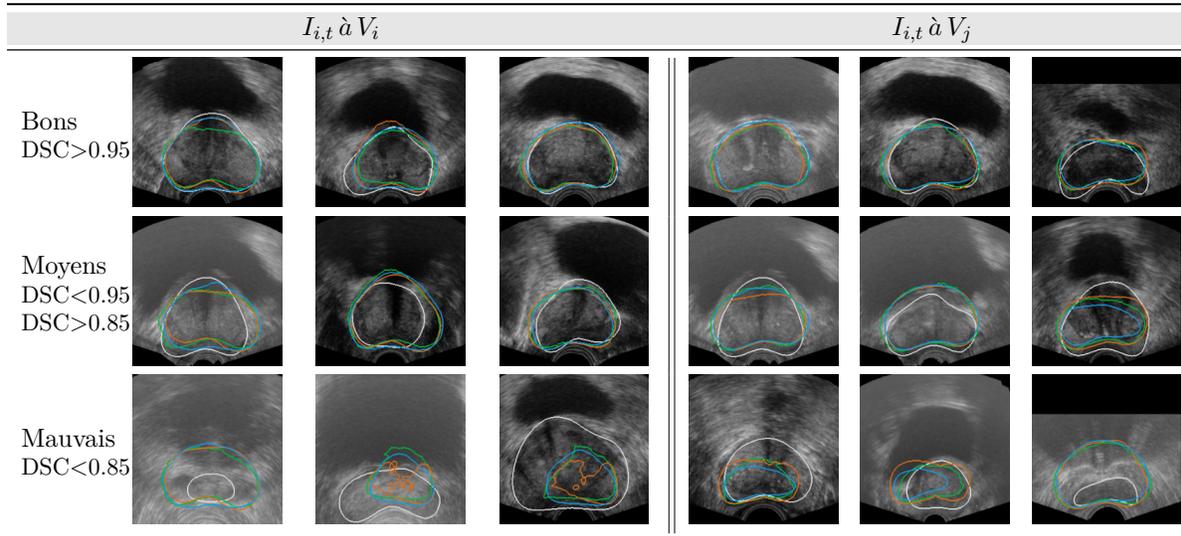


FIGURE VI.10 – Exemples de visualisation et comparaison des segmentations vérités terrain (blanches) et prédites par le réseau selon les différentes expériences : *PreSeg* (orange), *Joint-shared* (vert) et *sJSR* (bleu).

TABLE VI.5 – Comparaison des performances de recalage et de segmentation pour le jeu de données  $I_{i,t}$  à  $V_i$ , selon différentes pondérations pour les fonctions de coût.

Éval. séquentielle <i>sJSR</i>	TRE (↓)	SSIM (↑)	HD (↓)	DSC (↑)
Pondération : $L_{joint} = \alpha * L_{MSE} + \beta * L_{BCE}$				
( $\alpha = 1, \beta = 50$ )	$0.75 \pm 2.45$	$0.79 \pm 0.26$	$4.48 \pm 3.29$	$0.89 \pm 0.15$
( $\alpha = 50, \beta = 1$ )	$0.81 \pm 2.94$	$0.80 \pm 0.27$	$8.38 \pm 6.33$	$0.82 \pm 0.18$
Adaptative $\frac{L_{MSE}}{L_{BCE}}$	$0.89 \pm 2.73$	$0.73 \pm 0.24$	$6.73 \pm 4.55$	$0.85 \pm 0.16$

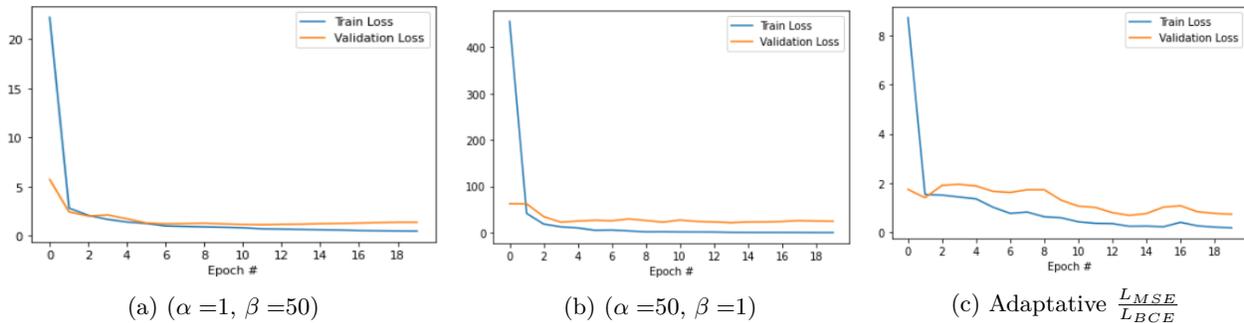


FIGURE VI.11 – Comparaison des profils de convergence selon différentes pondérations des fonctions de coût. Résultat du modèle *sJSR* et sur le jeu de données  $I_{i,t}$  à  $V_i$ .

## VI.5 Discussion

### VI.5.1 Discussion sur les contributions méthodologiques

#### VI.5.1.1 Entrée auxiliaire de segmentation

L'ajout d'information de segmentation sous forme d'entrée auxiliaire (voir bloc C, Figure VI.1) ne semble pas particulièrement bénéfique pour la tâche de recalage à réaliser.

De manière similaire aux conclusions pour les entrées spatiales additionnelles, il semble que l'ajout d'informations supplémentaires en entrée du réseau n'est pas toujours utile pour l'amélioration de l'apprentissage et de la réalisation d'une tâche. Cette observation est d'autant plus vérifiée si un tel ajout implique une augmentation de la complexité du réseau. Il semble que les réseaux plus complexes, avec notamment plus de paramètres à entraîner, sont plus difficiles à optimiser.

Ces entrées, concaténées à la branche principale ou traitées sur une branche parallèle, sont pourtant bien intégrées par le réseau, pour lequel les cartes d'activation démontrent une attention concentrée au niveau des contours de la prostate. Cependant, nous avons pu remarquer des différences de pixels importantes réparties dans toute l'image, à l'intérieur comme à l'extérieur de la prostate. Fournir les segmentations en entrée peut donc induire une attention focalisée sur les contours et défavoriser l'apprentissage du réseau, qui ne détecte plus toutes les zones d'intérêt présentes dans toute l'image.

#### VI.5.1.2 Tâche conjointe de segmentation

L'ajout d'information de segmentation sous forme d'une sortie conjointe (voir bloc B, Figure VI.1) pour un apprentissage multi-tâche semble particulièrement favorable, d'autant plus lorsque cet apprentissage est combiné à une architecture partagée et une fonction de coût communément établie.

Une architecture commune est critique pour un apprentissage multi-tâche. Après avoir comparé plusieurs degrés de partage, nous avons montré que l'utilisation d'une partie commune de l'architecture (*Joint-shared*) est plus pertinente que le simple partage dans la fonction de coût (*Joint-loss*). Nous pensons qu'avec un encodage partagé, les caractéristiques encodées correspondent alors à des zones d'intérêt pertinentes qui ont été définies de manière à optimiser les deux tâches combinées. D'autres types de partage plus superficiel existent, avec des interactions partielles et modulables des poids grâce à des unités de "cross-stitch" [39]. Une étude approfondie des différentes architectures possibles pourrait être davantage explorée dans le futur.

Une optimisation commune et pertinente implique une pondération adaptée des termes dans la fonction de coût. Cette pondération est cruciale pour la bonne association et considération des deux tâches, qui peuvent présenter différents rythmes d'apprentissage. Les quelques comparaisons que nous avons réalisées nous ont conduit à choisir une pondération fixe, pour laquelle nous avons obtenu une légère amélioration des performances. Notre protocole pour choisir cette pondération reste cependant exploratoire et nécessite une investigation plus poussée et plus méthodique. Ces problématiques sont des sujets de plus en plus présents dans la littérature [109] dont nous pourrions nous inspirer pour comparer d'autres types de pondération (alternée à chaque epoch, apprise durant l'apprentissage, etc.).

Comparativement aux expériences avec des segmentations d’entrées auxiliaires, il semble que l’apprentissage et la réalisation de telles segmentations sont bien plus bénéfiques que leur simple ajout. Nous pensons qu’apprendre à segmenter une image nécessite une analyse complète et globale de l’image avec la détection de nouveaux points d’intérêt. À l’inverse, fournir cette segmentation directement sous forme d’entrée, empêche ce travail d’analyse globale et concentre l’attention uniquement sur les contours, au détriment du reste de l’image.

Dans notre étude, nous avons démontré la supériorité d’un réseau multi-tâche par rapport au réseau mono-tâche, associé à chacune des tâches. Cela démontre que les deux tâches de recalage et segmentation sont complémentaires, est que l’apprentissage de l’une peut favoriser l’apprentissage de l’autre.

### VI.5.1.3 Combinaison de deux formes de segmentation

L’association des informations de segmentation sous forme d’a priori d’entrée et sous forme de tâche conjointe, semble être la configuration la plus optimale permettant d’améliorer à la fois les performances de recalage et de segmentation.

Alors que l’ajout de segmentation en entrée seule n’a pas été retenu comme étant une contribution significative, il semble qu’associé à un cadre multi-tâche, cet ajout améliore la tâche de segmentation et par conséquent améliore la tâche de recalage. Un a priori de segmentation de l’image précédente en entrée du réseau ( $S_{prior}$ ) est une information très bénéfique pour réaliser la tâche de segmentation de l’image courante. Cette idée est à rapprocher du concept d’initialisation et rappelle le même procédé proposé pour le recalage avec l’ajout d’un  $T_{prior}$ . Favoriser la tâche de segmentation peut finalement conduire à favoriser la tâche de recalage, notamment dans le cadre multi-tâche de *sJSR*. En effet, l’entrée de segmentation impacte la totalité du réseau, concernant à la fois la branche liée aux tâches de segmentation et la branche liée au recalage.

Fournir cet a priori de segmentation en entrée du réseau nécessite l’utilisation d’un réseau dédié à cette tâche. Nous avons comparé l’utilisation d’un réseau indépendant et dédié à cette tâche (*PreSeg*) et l’utilisation des prédictions du même réseau à l’étape précédente (*sJSR*). La première option permet d’exploiter un réseau mono-tâche pré-entraîné, dont les prédictions sont constantes et robustes au cours du temps. Pour la deuxième option, bien que les prédictions de segmentation obtenues soient de meilleure qualité, on observe un phénomène d’accumulation d’erreur qui n’est pas corrigé durant nos évaluations. Cela peut compromettre les performances du réseau, qui utilise alors des segmentations d’entrée de moins bonne qualité que celles ayant été utilisées durant l’apprentissage. Un réseau initialement meilleur qu’un autre, peut finalement voir ses performances réduites en présence de dérive. Un compromis doit donc être fait entre obtenir les meilleures performances et obtenir des performances robustes au cours de la trajectoire.

## VI.5.2 Conclusions sur ce troisième modèle

### VI.5.2.1 Conclusions sur les performances obtenues

La meilleure configuration que nous avons obtenue consiste en la combinaison des informations positionnelles ( $T_{prior}, T_{suivi}$ ) et structurelles ( $S_{prior}$  en entrée auxiliaire) à travers un apprentissage multi-tâche de recalage  $\widehat{T}_t$  et de segmentation  $\widehat{S}_t$ . Nous avons proposé une architecture avec un encodage partagé entre les deux tâches et une pondération adaptée à notre problématique, qui permet d’obtenir des performances satisfaisantes à la fois sur la tâche de segmentation et sur la tâche de recalage. Après recalage, 97% (respectivement 92%) des erreurs obtenues sont inférieures au seuil clinique à atteindre ( $TRE < 2.5\text{mm}$ ) sur les jeux de données  $I_{i,t}$  à  $V_i$  (respectivement  $I_{i,t}$  à  $V_j$ ).

### VI.5.2.2 Comparaison avec la littérature

Parmi les études référencées dans notre état-de-l’art [38, 39, 43, 48, 84], nous insistons sur le fait qu’aucune étude (à ce jour et à notre connaissance) n’a pu être trouvée exploitant à la fois des informations de segmentation tout en partageant le même objectif et les mêmes applications que les nôtres (recalage 2D/3D monomodal US). Même si ces études ont pu être des sources d’inspiration concernant le type d’intégration utilisée (concaténation, branche parallèle, encodage partagé, cross-stitch), ces architectures ont dû être adaptées à notre méthode initiale  $SpT$ . Une comparaison rigoureuse avec celles-ci est donc difficile.

Les autres méthodes de la littérature [56, 57] (sans information de segmentation) ont, quant à elles, déjà été comparées à nos méthodes précédentes qui avaient notamment démontré de meilleures performances. Comparer la nouvelle méthode proposée  $sJSR$  avec l’ancienne  $SpT$  suffit alors pour démontrer la supériorité de la nouvelle méthode sur l’état-de-l’art général.

### VI.5.2.3 Conclusions sur les objectifs

Les conclusions en réponse aux trois objectifs principaux de cette troisième méthode sont :

- Les expériences avec diverses formes de segmentation considérées (entrée auxiliaire *Aux* ou sortie conjointe *Joint*) ont permis d’obtenir de meilleurs résultats que la méthode précédente  $SpT-Net$  (sans aucune information de segmentation). Cela démontre la pertinence de l’apport d’informations structurelles pour l’amélioration de la tâche de recalage 2D/3D. L’utilisation d’un apprentissage multi-tâche est une nouvelle façon d’exploiter des informations additionnelles, qui semble plus pertinente que toutes nos autres propositions d’intégrations étudiées jusqu’à présent.
- Parmi les différentes configurations testées, nous avons proposé une architecture avec un encodage partagé entre les deux tâches et une pondération adaptée à notre problématique, qui nous ont permis d’obtenir des performances satisfaisantes à la fois sur la tâche de segmentation et sur la tâche de recalage.
- Nous avons vérifié la robustesse de notre méthode sur différents niveaux de difficulté et à travers une évaluation séquentielle sur des trajectoires complètes. En particulier, la méthode proposée dans ce chapitre semble particulièrement bénéficier aux tâches de recalage plus complexes.

**Points clé**

- ❖ La combinaison des informations de segmentation comme entrée du réseau (a priori de forme) et comme tâche conjointe (apprentissage multi-tâche) permet d'améliorer la tâche du recalage, même monomodal et multidimensionnel (2D/3D US/US).
- ❖ Le modèle proposé combine des informations positionnelles ( $T_{prior}$ ,  $T_{suivi}$ ) et structurales ( $S_{prior}$ ) à travers un apprentissage multi-tâche et un encodeur partagé. Notre modèle est autonome lors de procédures cliniques, grâce à des connexions séquentielles proposées pour fournir les entrées d'a priori directement à partir des prédictions précédentes. Le phénomène de dérive impliqué est évalué et contenu tout au long d'une trajectoire.
- ❖ Le modèle proposé obtient de meilleures performances que notre méthode précédente et que les méthodes bénéficiant de contribution unique de segmentation inspirées de la littérature. Des pistes d'améliorations concernant la pondération des tâches conjointes et des architectures multi-tâches sont encore en cours d'exploration.
- ❖ Ces travaux sont en cours de soumission au journal "Medical Physics".

---

## Conclusion

---

*Cet* ultime chapitre conclut ce manuscrit et ces années de doctorat. Nous rappelons dans une première partie l'évolution de nos différentes méthodes avec leurs contributions majeures. Une deuxième partie de cette conclusion met en avant les perspectives d'amélioration futures, pour aller plus loin dans la recherche d'information contextuelle liée à la navigation et pour étendre nos conditions expérimentales en vue d'une application clinique.

### Sommaire

---

<b>VII.1 Rappel des contributions</b> . . . . .	<b>162</b>
VII.1.1 Contributions méthodologiques . . . . .	163
VII.1.2 Contributions expérimentales . . . . .	164
<b>VII.2 Perspectives</b> . . . . .	<b>166</b>
VII.2.1 Vers une nouvelle formulation du problème . . . . .	166
VII.2.1.1 Informations séquentielles et structures dynamiques . . . . .	166
VII.2.1.2 Optimisation multi-tâche . . . . .	167
VII.2.2 Vers une application clinique . . . . .	168
VII.2.2.1 Nouvelles acquisitions de données cliniques : comment valider le recalage ? . . . . .	168
VII.2.2.2 Quelle assistance pour les praticiens ? . . . . .	168
<b>VII.3 Publication et communication</b> . . . . .	<b>169</b>

---

## VII.1 Rappel des contributions

Durant un examen de biopsies prostatiques, la navigation et la précision du ciblage des biopsies sont rendues complexes à la fois par la modalité échographique (2D US) couramment utilisée, et par de potentiels bougés ou déformations de la prostate. Ces éléments rendent la localisation des biopsies difficile et imprécise, et peuvent alors compromettre le diagnostic du cancer de la prostate ainsi que les décisions thérapeutiques.

Afin de faciliter la réalisation des biopsies, le principal objectif de cette thèse était de concevoir et développer un dispositif d'assistance à la navigation permettant à l'urologue de mieux se repérer pour ainsi mieux visualiser la position de la biopsie qu'il cible. Ce dispositif doit permettre une localisation précise et continue, par rapport à une référence anatomique fixe et connue, le tout en tenant compte des bougés et déformations de la cible.

Les outils actuels utilisés en clinique offrent soit une localisation de la sonde en temps-réel mais ne prenant pas en compte les mouvements et/ou déformations précises de l'anatomie ; soit une localisation précise de l'organe mais dont les temps d'acquisitions ne permettent pas une localisation en temps-réel. De plus en plus de méthodes visent à combiner les deux types d'approches pour trouver un meilleur compromis entre rapidité et précision. Les plus prometteuses d'entre elles présentent des fréquences et des précisions de recalage satisfaisantes, mais les conditions expérimentales mises en place pour évaluer ces méthodes sont encore très limitées en termes de réalisme et de validité clinique.

Nous avons proposé des méthodes de recalage rigide 2D/3D en temps-réel, permettant de localiser la coupe 2D US courante (guidage) par rapport à un volume de référence US (pris en début de navigation) en superposant au mieux le contenu anatomique de la prostate. Afin de répondre aux limites de l'existant, les principales originalités de nos travaux sont de considérer des informations liées à la trajectoire réalisée et de développer des approches fondées sur l'intelligence artificielle (par apprentissage profond), en respectant davantage les conditions de réalisation clinique.

### VII.1.1 Contributions méthodologiques

Plusieurs façons d’exploiter des informations liées à la trajectoire réalisée durant une procédure de biopsie prostatique ont été étudiées avec différentes propositions d’intégration ou de combinaison dans une architecture profonde. Parmi les différentes formes de données d’entrée possibles, nous nous sommes particulièrement intéressés aux informations : *positionnelles* (résultats du recalage précédent, suivi de sonde, flot optique), *spatiales* (coupe voisine, volume ou sous-volume de référence), ou *structurelles* (image de segmentation). Elles présentent toutes l’avantage de contenir des informations contextuelles permettant de mieux considérer la trajectoire réalisée. La bonne intégration de telles entrées dans une architecture d’apprentissage profond nécessite de porter une attention particulière aux structures de réseaux utilisées (convolution, récurrent, attention, etc.) et aux types d’optimisations associées (simple, auto-pénalisation, multi-tâches).

Trois principales évolutions de notre méthode ont donc été explorées selon la configuration et la combinaison des trois types d’entrées précédemment citées.

Dans un premier temps, nous nous sommes concentrés sur l’apport d’a priori *positionnels*, relatifs ou absolus. Nous avons développé un réseau convolutif 2D permettant d’estimer les paramètres de transformation, pour une tâche de recalage 2D US/ 3D US en temps-réel. Parmi différentes configurations testées, nous avons privilégié un encodage de la transformation recherchée sous la forme d’angles d’Euler et un apprentissage supervisé basé sur la MSE. En incorporant par une concaténation directe des informations liées à la trajectoire réalisée, à travers notamment les résultats du recalage précédent ( $T_{prior}$ ) et du suivi de sonde ( $T_{suivi}$ ), nous avons démontré une amélioration significative de la qualité du recalage estimé.

Puis, nous avons décidé d’exploiter davantage les informations contextuelles *spatiales* concernant l’espace de référence. Les expériences menées suggèrent que ni l’ajout de nouveaux types d’entrées *spatiales* (empilement de coupes voisines/ volume/sous-volume) ni l’apport de pénalisation spatiale (SSIM) ne participent à l’amélioration de la tâche de recalage, et ce, malgré l’utilisation d’une nouvelle architecture plus adaptée et équilibrée. Ces ajouts impliquent des entrées moins robustes face aux potentiels phénomènes d’accumulation d’erreurs et des réseaux plus gourmands et plus complexes en calculs qui peuvent ralentir la convergence et l’apprentissage.

Enfin, une dernière méthode se concentre sur l’ajout d’informations *structurelles* qui permettent de considérer les segmentations des contours de la prostate au sein des images (entrée auxiliaire  $S_{prior}$  ou sortie conjointe  $\widehat{S}_t$ ). Nous avons proposé une architecture avec un encodage partagé entre les deux tâches et une pondération adaptée à notre problématique, qui permet d’améliorer la tâche du recalage, même monomodal et multidimensionnel. Notre modèle est autonome lors de procédures cliniques, grâce à des connexions séquentielles proposées pour fournir les entrées d’a priori directement à partir des prédictions précédentes.

Finalement, en ajoutant des informations liées à la trajectoire réalisée, en considérant notamment les résultats du recalage précédent  $T_{prior}$ , le déplacement relatif de la sonde  $T_{suivi}$  et un apprentissage multi-tâche avec une tâche de segmentation  $\widehat{S}_t$ , nous avons observé une amélioration significative de la qualité du recalage estimé  $\widehat{T}_t$ . L’identification de nouveaux types d’informations de trajectoire ainsi que de nouvelles manières de les intégrer à une architecture profonde restent des perspectives de cette thèse.

### VII.1.2 Contributions expérimentales

Afin de répondre aux limites des méthodes de validation existantes, nous avons proposé, comme second type de contribution de cette thèse, de nous concentrer sur le respect des conditions cliniques de réalisme et de faisabilité de notre application. Cela s'illustre notamment par les données utilisées (simulation d'entrées et de vérités terrain) et des méthodes de validation choisies (évaluation sans intervention extérieure, déroulement réaliste, métriques pertinentes).

À défaut d'avoir des images (2D US) et leur localisation directement issues des procédures cliniques, nous avons opté pour un processus de simulation permettant d'extraire artificiellement des coupes en les rééchantillonnant à partir d'informations cliniques disponibles. Nous avons proposé plusieurs simulations (trajectoire de biopsie via un simulateur, reconstruction de trajectoire par interpolation, ou trajectoire par balayage base-apex), chacune présentant ses avantages et limites.

Dans un premier temps, nous avons utilisé un simulateur de biopsie, manipulé par des experts sur des données cliniques. Les trajectoires simulées sont fidèles à la complexité et aux caractéristiques de mouvements (vitesse, amplitude, etc.) et génèrent des erreurs initiales (avant recalage) réalistes. Cependant, ce simulateur ne dispose que d'un seul volume pour chaque patient, ce qui contraint les rééchantillonnages et les localisations au sein de ce même volume, générant ainsi un contenu iconique identique et des zones noires qui perturbent le réalisme de ces données. Par ailleurs, seuls les 3 paramètres de rotation sont étudiés.

Un deuxième processus de simulation se concentre sur une complexification de la tâche avec un recalage vers un autre volume. L'utilisation des différents volumes successifs, enregistrés au cours d'une procédure, ainsi que la reconstruction de la trajectoire réalisée par interpolation, ont permis à la fois de générer des images présentant des déformations différentes et de simuler les translations de la sonde. Les trajectoires, encore trop complexes, mènent aux mêmes conséquences d'apparition de zones noires et d'erreurs initiales trop importantes.

Enfin, le dernier processus de simulation vise à la fois à supprimer la présence de zones noires et à générer plusieurs niveaux de difficultés. Les images rééchantillonnées dans un volume peuvent être localisées soit dans ce même volume ( $I_{i,t}$  à  $V_i$ ), soit dans un volume voisin ( $I_{i,t}$  à  $V_j$ ). En utilisant des trajectoires simplifiées de balayage base-apex, les images sont générées sans zone noire arbitraire, et les erreurs avant recalage sont plus comparables à celles de la littérature. Cependant, ce processus n'intègre pas toute la complexité de trajectoires réelles (trajectoire incomplète, unidirectionnelle, vitesse du geste). Cette dernière simulation reste le meilleur compromis que nous ayons pu trouver, entre réalisme et biais occasionné, parmi nos différentes simulations et celles de la littérature.

Pour réaliser une évaluation réaliste et adaptée à la finalité clinique, nous avons proposé de mesurer les erreurs à travers une série de recalages successifs, réalisés au cours d'une trajectoire complète. Cette évaluation séquentielle est cruciale pour quantifier le risque d'accumulation d'erreur, d'autant plus avec l'utilisation de différentes entrées liées à la trajectoire réalisée. La combinaison de métriques géométriques (TRE), iconiques (SSIM/NCC) et statistiques ( $R^2$ /MAE) permet d'estimer les erreurs accumulées sur une trajectoire complète.

En accord avec les informations disponibles dans un cadre clinique, nous avons envisagé une procédure de biopsie avec une assistance à la navigation *par portion*, entre chaque site de biopsie. En particulier, les recalages 3D/3D US estimés à chaque nouveau site de biopsie peuvent être utilisés

comme des réajustements clés et ainsi contenir le potentiel phénomène de dérive.

Finalement, la meilleure configuration obtenue nous permet d'atteindre des erreurs de recalage entre 0.5 à 2 mm selon les bases de données testées, ce qui représente entre 90 et 97% des erreurs corrigées sous le seuil clinique à atteindre (TRE <2.5 mm). Ces résultats sont obtenus avec des temps de prédictions très rapides (<50 ms) sans une implémentation particulièrement optimisée du code. Ces performances sont cependant sujettes à des phénomènes de dérive, observés au cours de trajectoires complètes, avec l'utilisation des entrées d'a priori comme  $T_{prior}$  et  $S_{prior}$ .

Même si un intérêt particulier a été porté sur le réalisme de nos expériences, nos protocoles de génération de données et de validation présentent encore de nombreuses limites ou simplifications, résumées précédemment. Cela peut notamment impacter les performances obtenues au moment d'une utilisation dans un cadre clinique. Il convient alors d'explorer d'autres pistes méthodologiques pour affiner la précision et la robustesse du recalage estimé et d'améliorer la validité de nos évaluations en vue d'une intégration pour des dispositifs cliniques.

## VII.2 Perspectives

### VII.2.1 Vers une nouvelle formulation du problème

#### VII.2.1.1 Informations séquentielles et structures dynamiques

Étant donné que la navigation durant un examen prostatique est un processus temporel très stéréotypé et que nous venons de démontrer l'avantage d'utiliser des informations liées à la trajectoire déjà réalisée, l'étape suivante consiste à étudier davantage les informations *séquentielles* au travers de réseaux dits *dynamiques*.

Ces informations *séquentielles* peuvent correspondre à une séquence d'a priori  $(\widehat{T}_{t-1}, \dots, \widehat{T}_{t-S-1})$  seule ou combinée à une séquence d'images  $(I_t, \dots, I_{t-S})$  de taille  $S$ . Ces informations présentent l'avantage d'accumuler de l'information spatiale et temporelle combinée, à partir de plusieurs images issues du guidage peropératoire. Ce contexte spatio-temporel est de plus en plus présent dans la littérature, notamment pour des problématiques de reconstruction de volume (voir section II.4.1.3).

La bonne intégration des entrées *séquentielles* est souvent traitée grâce à des structures *dynamiques* (voir section II.4.3.2) de type réseaux récurrents ou encore réseaux "Transformers". Ces différentes architectures de réseaux sont reconnues pour favoriser l'interaction logique entre chaque élément d'une séquence d'entrée. Par exemple, les réseaux récurrents permettent d'initialiser chaque cellule du réseau en fonction des instants précédents de la séquence, en retenant plus ou moins d'information sur une fenêtre de temps plus ou moins longue. Les Transformers permettent, quant à eux, de combiner et de pondérer certaines prédictions entre elles en fonction de leur similarité (module d'attention) et de leur position au sein de la séquence ("embedding"). Cet encodage peut également permettre de fournir des informations additionnelles, adaptées à la problématique. De plus, des découpages locaux des images sous formes de "patch" ou sous forme de "tubelet" sont répandues pour des problématiques de traitements de séquence d'images avec les Transformers [110, 111].

En combinant des informations d'entrées *séquentielles* à travers des architectures de réseau *dynamiques*, l'idée serait ainsi de mieux analyser et traiter l'évolution de la trajectoire passée, sur une fenêtre de temps plus grande et à travers plusieurs images. Ces modules pourraient être pré-entraînés sur des tâches adaptées (prédire la suite d'une trajectoire stéréotypée ou corriger une dérive) permettant de se concentrer sur le profil global de la trajectoire. Cette approche dynamique pourrait aider à contenir le phénomène de dérive observé dans ce travail de thèse.

Quelques unes de ces pistes de réflexion ont fait l'objet de tests préliminaires mais qui n'ont pas donné les résultats escomptés. Ces architectures, très coûteuses et difficiles à faire converger, nécessitent davantage de recherche pour favoriser leur apprentissage. L'utilisation de parties d'architecture pré-entraînées (transfert d'apprentissage) reste une possibilité à considérer.

### VII.2.1.2 Optimisation multi-tâche

L'utilisation d'un apprentissage *multi-tâche* est une autre façon d'exploiter des informations additionnelles, dont nous avons déjà pu démontrer la pertinence avec l'apprentissage de la tâche de segmentation. À travers l'apprentissage d'une tâche complémentaire de la tâche de recalage, le réseau apprend à mieux concentrer son attention sur des caractéristiques pertinentes que si celles-ci lui avaient été fournies directement en données d'entrée. Cela permet non seulement d'obtenir de meilleures performances du réseau sur la tâche principale, mais également de bénéficier de l'estimation d'une deuxième tâche et utiliser ces nouvelles informations pour renforcer des contraintes d'apprentissage (voir section II.4.4.2).

Parmi les différentes tâches complémentaires possibles du recalage, nous avons déjà pu démontrer le bénéfice d'apprendre à segmenter la prostate, de manière simultanée et conjointe. Pour aller plus loin dans cette démarche, on peut imaginer exploiter la segmentation sous la forme de nuage de points, avec l'utilisation de réseaux spécifiques à cette forme d'entrée (PointNet [112]). Cette analyse permettrait de considérer davantage les connexions ordonnées et internes de ce maillage et d'envisager un recalage non-rigide avec l'estimation du déplacement pour chaque point.

Une autre piste à étudier concerne la reconstruction d'un volume à partir d'une séquence de coupes. Cette tâche de reconstruction est très complémentaire de la localisation de ces coupes, et permettrait notamment de concentrer l'apprentissage sur la cohérence des formes et des structures anatomiques du volume. De plus, cette reconstruction de volume pourrait également permettre d'ajouter des termes d'auto-pénalisation selon les similarités entre les images d'entrée et des images extraites du volume reconstruit ou selon la similarité avec le volume de référence [71, 73]. Cette estimation de volume reconstruit pourrait également être utilisé pour mettre à jour le modèle du volume de référence pré-opératoire.

Pour finir, nous pensons que la réalisation de procédures de biopsies, notamment pour les biopsies standardisées, repose sur une navigation assez stéréotypée et contrainte par l'anatomie du patient. Trois transitions se répètent notamment durant toute la durée d'un examen : médiane-latérale, base-apex, rotation autour de la sonde. On peut alors imaginer entraîner le réseau à détecter le « type » de transition réalisée. Cela permettrait de concentrer l'attention du réseau sur le profil stéréotypé des portions de trajectoires. De plus, cette classification du type de trajectoire pourrait également permettre d'initialiser une localisation moyenne des coupes [63] ou de pondérer les prédictions selon des mesures de confiance pour chaque classe [62, 73].

Ces tâches sont toutes complémentaires et pourraient être combinées à plusieurs niveaux en combinant plusieurs d'entre elles. De plus, toutes ces pistes d'apprentissage multi-tâches sont particulièrement focalisées sur des propriétés intrinsèques des trajectoires. Nous pensons alors que la combinaison d'un apprentissage multi-tâche et d'une analyse dynamique sur des entrées *séquentielles* serait particulièrement adaptée et pertinente.

## VII.2.2 Vers une application clinique

### VII.2.2.1 Nouvelles acquisitions de données cliniques : comment valider le recalage ?

Pour avoir des simulations et des validations plus réalistes, il faudrait envisager de réaliser de nouvelles acquisitions lors de biopsies prostatiques sur patients (sur fantômes de prostate ou sur cadavres), permettant de collecter le flux d'images échographiques 2D directement durant l'examen.

Se pose alors la question de l'alignement optimal des images 2D US à chaque instant compte tenu des bougés et déformations de la prostate par rapport à un espace de référence : comment établir cette localisation de manière rigoureuse comme vérité terrain ? Si localiser la sonde est possible avec un système de suivi optique (de type Polaris avec les calibration nécessaire), cela ne permet pas pour autant de localiser de façon précise les bougés et déformations élastiques subies par la prostate au cours de l'examen. S'appuyer sur des méthodes 2D/3D de l'état-de-l'art (par exemple [51, 52]) semble aussi problématique car toutes ces méthodes présentent généralement les mêmes limites que les nôtres quant à la bonne validité de la localisation continue estimée. Faire des estimations de localisations ponctuelles, seulement à chaque site de biopsie, grâce à une méthode de recalage 3D/3D (par exemple [17]), semble également trop simpliste et limite le nombre de données que nous pouvons utiliser pour quantifier la précision et robustesse de notre méthode. Une détection automatique de fiduciels pourrait, quant à elle, être utile pour établir une vérité terrain continue, mais nécessite au préalable l'implémentation d'une telle méthode et la garantie de sa validité et de ses performances.

Toutes ces questions de validité du recalage estimé restent au coeur des défis actuels. Le temps nécessaire au développement et à l'obtention d'autorisations réglementaires pour la mise en place de telles expériences pourraient être conséquent.

### VII.2.2.2 Quelle assistance pour les praticiens ?

Pour concrétiser l'aboutissement de notre travail et pour valider l'intérêt d'utiliser notre assistance à la navigation, il convient de réfléchir aux types d'informations que l'on souhaiterait restituer au chirurgien et à la meilleure façon de les présenter dans un dispositif d'interface visuelle. Ainsi, différentes questions se posent : comment rendre cette visualisation aussi efficace et intuitive que possible pour une interprétation et un suivi en temps réel par les urologues ? Comment les guider vers leur prochaine cible, s'ils le jugent nécessaire ? Comment définir un indicateur de "confiance" concernant la localisation de biopsie en cours ?

En effet, des indications d'incertitude de notre modèle peuvent être utiles pour notamment désactiver l'assistance au-delà d'un certain seuil d'incertitude et ne pas induire de mauvaises décisions du praticien. Ce seuil d'incertitude de l'algorithme de localisation proposé ainsi que la manière dont elle pourrait être estimée par le réseau lui-même reste cependant à définir.

Cette future interface d'assistance à la navigation a déjà fait l'objet de premières discussions avec des praticiens. Une fois cette interface réalisée, il sera nécessaire d'évaluer l'intérêt de cette assistance : à travers des sessions de biopsies avec ou sans cette assistance, sur le simulateur et par plusieurs praticiens. Le but est de quantifier dans quelle mesure celle-ci permet (i) d'améliorer l'échantillonnage des biopsies, (ii) de réduire la durée de l'examen, et (iii) de favoriser l'apprentissage du geste pour de nouveaux praticiens.

## VII.3 Publication et communication

Ces travaux ont fait l'objet de communications scientifiques dans différents contextes et interventions orales. Une liste des publications et des interventions est décrite ci-dessous :

- Listes des publications :
  - T.Dupuy, C.Beitone, J.Troccaz, and S. Voros, (2023). "2D/3D Deep Registration Along Trajectories With Spatiotemporal Context : Application To Prostate Biopsy Navigation". IEEE Transactions on Biomedical Engineering.
  - T.Dupuy, C.Beitone, J.Troccaz, and S. Voros, (2021). "2D/3D deep registration for real-time prostate biopsy navigation". In Medical Imaging 2021 : Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 11598, pp. 463-471). SPIE (Cum Laude Poster Winner)
  - J.B.Lefrancq, J.L.Descotes, S.Selmi, T.Dupuy, ... and G.Fiard, (2021). "External validation of a prostate biopsy simulator". Progrès en Urologie, 31(16), 1115-1122.
  
- Communication scientifique :
  - Participation à la finale 2021 de la présélection locale "Universités Grenoble Alpes & Savoie Mont Blanc" du concours international Ma thèse en 180 secondes. Vidéo disponible à l'adresse<sup>1</sup>.
  - Capsule vidéo de vulgarisation scientifique réalisée dans le cadre de la Journée des doctorants du laboratoire TIMC. Vidéo disponible à l'adresse<sup>2</sup>.

---

1. <https://www.youtube.com/watch?v=T2a7-n1SET8>

2. <https://www.youtube.com/watch?v=4Yf3fcTYgT0&t=21s>



---

## Annexe A : Architecture détaillée des réseaux proposés

---

*Cette* annexe présente les architectures des différents réseaux impliqués dans ce manuscrit en détaillant la nature, les dimensions et les paramètres couche par couche. Un exemple associé aux caractéristiques de dimensions et/ou aux paramètres spécifiques d'une base de données est également fourni en dernière colonne.

Les trois architectures associées aux trois principales évolutions de méthodes proposées durant ces travaux (Chapitres IV, V, VI) sont fournies respectivement dans les Tables 1, 3, 5. D'autres configurations et expériences complémentaires sont également fournies, en se référant aux schémas et aux sections correspondantes.

TABLE 1 – Résumés et détails par couche de l’architecture du réseau convolutif proposé en section IV.2.1 et illustré en Figure IV.1.

Numéro couche	Type de couche	Caractéristiques	Dimension de sortie	Nombre paramètres	Exemple sur Biopsym
*	Entrée	$(I_t, I_{t-1})$	$(B,H,W,D)$	0	$(B,130,250,2)$
*	Entrée	$I_{flot}$	$(B,H,W,D)$	0	$(B,130,250,2)$
<b>Convolution</b>					
L=1	Conv2D	$(k=5, f=64, s=2, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,65,125,64)$
L=2	Conv2D	$(k=5, f=64, s=2)$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,33,63,64)$
-	MaxPool2D	$(k=2, s=1)$	Idem	0	Idem
L=3	Conv2D	$(k=3, f=64, s=2, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,17,32,64)$
L=4	Conv2D	$(k=3, f=64, s=2, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,64)$
-	MaxPool2D	$(k=2, s=1)$	Idem	0	Idem
L=5	Conv2D	$(k=5, f=128, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,128)$
L=6	Conv2D	$(k=5, f=128, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,128)$
-	MaxPool2D	$(k=2, s=1)$	Idem	0	Idem
L=7	Conv2D	$(k=3, f=128, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,128)$
L=8	Conv2D	$(k=3, f=128, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,128)$
-	MaxPool2D	$(k=2, s=1)$	Idem	0	Idem
L=9	Conv2D	$(k=5, f=256, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,256)$
L=10	Conv2D	$(k=5, f=256, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,256)$
-	MaxPool2D	$(k=2, s=1)$	Idem	0	Idem
L=11	Conv2D	$(k=3, f=256, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,256)$
L=12	Conv2D	$(k=3, f=256, s=1, \text{ReLu})$	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	$(B,9,16,256)$
-	MaxPool2D	$(k=2, s=1)$	Idem	0	Idem
-	AvgPool2D	$(k=2, s=2)$	$(B, \frac{H}{2^5}, \frac{W}{2^5}, D)$	0	$(B,4,8,256)$
-	Flatten			0	$(B,8192,1)$
<b>Couches FC</b>					
L=13	Dense	$f=2048, \text{ReLu}$	$(B,f)$	$(D_{L-1}+1)*f$	$(B,2048,1)$
L=14	Dense	$f=1024, \text{ReLu}$	$(B,f)$	$(D_{L-1}+1)*f$	$(B,1024,1)$
L=15	Dense	$f=512, \text{ReLu}$	$(B,f)$	$(D_{L-1}+1)*f$	$(B,512,1)$
*	Entrée	$T_{prior}$	$(B, \#, 1)$	0	$(B,3,1)$
*	Entrée	$T_{suivi}$	$(B, \#, 1)$	0	$(B,3,1)$
L=16	Dense	$f=256, \text{ReLu}$	$(B,f)$	$(D_{L-1}+1)*f$	$(B,256,1)$
L=17	Dense	$f=64, \text{ReLu}$	$(B,f)$	$(D_{L-1}+1)*f$	$(B,64,1)$
L=18	Dense	$f=16, \text{ReLu}$	$(B,f)$	$(D_{L-1}+1)*f$	$(B,16,1)$
Sortie	Dense	$f=\#, \text{linear}$	$(B,f)$	$(D_{L-1}+1)*f$	$(B,3,1)$
Total params					24.2M

TABLE 2 – Résumés et détails par couche de l’architecture U-net, proposée pour le cas particulier d’encodage sous la forme de coordonnées en image à 3 canaux (voir Annexe-VII.3) et illustrée en Figure 1.

Numéro couche	Type de couche	Caractéristiques	Dimension de sortie	Nombre paramètres	Exemple sur Biopsym
*	Entrée	$(I_t, I_{t-1})$	(B,H,W,D)	0	(B,128,192,2)
*	Entrée	$I_{flot}$	(B,H,W,D)	0	(B,128,192,3)
*	Entrée	$T_{prior}$	(B,H,W,D)	0	(B,128,192,3)
*	Concatenate	-	-	0	(B,128,192,8)
Level l=1					
L=1	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,128,192,64)
L=2	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
-	MaxPool2D	(k=2, s=2)	$(B, \frac{H}{2}, \frac{W}{2}, f)$	0	(B,64,96,64)
Level l=2					
L=1	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,64,96,64)
L=2	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
-	MaxPool2D	(k=2, s=2)	$(B, \frac{H}{2}, \frac{W}{2}, f)$	0	(B,32,48,64)
Level l=3					
L=1	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,32,48,128)
L=2	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
-	MaxPool2D	(k=2, s=2)	$(B, \frac{H}{2}, \frac{W}{2}, f)$	0	(B,16,24,128)
Level l=4					
L=1	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,16,24,128)
L=2	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
-	MaxPool2D	(k=2, s=2)	$(B, \frac{H}{2}, \frac{W}{2}, f)$	0	(B,8,12,128)
Level l=5					
L=1	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,8,12,256)
L=2	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
-	MaxPool2D	(k=2, s=2)	$(B, \frac{H}{2}, \frac{W}{2}, f)$	0	(B,4,6,256)
Level l=6					
L=1	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,4,6,256)
L=2	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
-	MaxPool2D	(k=2, s=2)	$(B, \frac{H}{2}, \frac{W}{2}, f)$	0	(B,2,3,256)
Level l=7					
L=1	Conv2D	(k=3, f=512, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,2,3,512)
L=2	Conv2D	(k=3, f=512, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
Level l=6					
L=0	UpSampling2D	(k=2, s=2)	$(B, \frac{H}{2}, \frac{W}{2}, f)$	0	(B,4,6,512)
L=1	Concatenate	-	-	0	(B,4,6,768)
L=1	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,4,6,256)
L=2	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem

Level l=5					
L=0	UpSampling2D	(k=2, s=2)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, f)$	0	(B,8,12,256)
L=1	Concatenate	-	-	0	(B,8,12,512)
L=1	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,8,12,256)
L=2	Conv2D	(k=3, f=256, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
Level l=4					
L=0	UpSampling2D	(k=2, s=2)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, f)$	0	(B,16,24,256)
L=1	Concatenate	-	-	0	(B,16,24,384)
L=1	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,16,24,128)
L=2	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
Level l=3					
L=0	UpSampling2D	(k=2, s=2)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, f)$	0	(B,32,48,128)
L=1	Concatenate	-	-	0	(B,32,48,256)
L=1	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,32,48,128)
L=2	Conv2D	(k=3, f=128, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
Level l=2					
L=0	UpSampling2D	(k=2, s=2)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, f)$	0	(B,64,96,128)
L=1	Concatenate	-	-	0	(B,64,96,192)
L=1	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,64,96,64)
L=2	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
Level l=1					
L=0	UpSampling2D	(k=2, s=2)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, f)$	0	(B,128,192,64)
L=1	Concatenate	-	-	0	(B,128,192,128)
L=1	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,128,192,64)
L=2	Conv2D	(k=3, f=64, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	Idem
-	Activation	ReLU	Idem	0	Idem
Sortie	Conv2D	(k=3, f=3, s=1)	Idem	$(k*k*D_{L-1}+1)*f$	(B,128,192,3)
Total params					11.6 M

TABLE 3 – Résumés et détails par couche de la nouvelle architecture proposée *Main-3D* (voir section V.2.1) et illustré en Figure V.1.

Numéro couche	Type de couche	Caractéristiques	Dimension de sortie	Nombre paramètres	Exemple pour E=V=128
<b>Branche principale</b>					
*	Entrée	$I_t$	(B,H,W,1)	0	(B,256,256,1)
*	Entrée	$locale-V$	(B,H,W,V)	0	(B,256,256,V)
-	Concatenate	-	(B,H,W,V+1)	0	(B,256,256,V+1)
L=0	Conv2D	(k=5, f=E, s=1)	(B,H,W,E,1)	$(k*k*D_{L-1}+1)*E$	(B,256,256,128,1)
L=1	Conv3D	(k=5, f=64, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{E}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,128,128,64,64)
L=2	Conv3D	(k=5, f=64, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{E}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,64,64,32,64)
-	Activation	ReLU	Idem	0	Idem
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
L=3	Conv3D	(k=3, f=128, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{E}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,32,32,16,128)
L=4	Conv3D	(k=3, f=128, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{E}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,16,16,8,128)
-	Activation	ReLU	Idem	0	Idem
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
-	AvgPool3D	(k=2, s=2)	$(B, \frac{H}{2^5}, \frac{W}{2^5}, \frac{E}{2^5}, \frac{f}{2})$	0	(B,8,8,4,128)
-	Flatten			0	(B,32768,1)
<b>Branche parallèle</b>					
*	Entrée	$glob-V$ ou $part-V$	(B,H,W,V,1)	0	(B,256,256,128,1)
L=1	Conv3D	(k=5, f=64, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{V}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,128,128,64,64)
L=2	Conv3D	(k=5, f=64, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{V}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,64,64,32,64)
-	Activation	ReLU	Idem	0	Idem
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
L=3	Conv3D	(k=3, f=128, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{V}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,32,32,16,128)
L=4	Conv3D	(k=3, f=128, s=2)	$(B, \frac{H}{2L}, \frac{W}{2L}, \frac{V}{2L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,16,16,8,128)
-	Activation	ReLU	Idem	0	Idem
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
-	AvgPool3D	(k=2, s=2)	$(B, \frac{H}{2^5}, \frac{W}{2^5}, \frac{V}{2^5}, \frac{f}{2})$	0	(B,8,8,4,128)
-	Flatten			0	(B,32768,1)
<b>Concatenation et fin du réseau</b>					
-	Concatenate	-	-	0	(B,65536,1)
L=13	Dense	f=2048, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,2048,1)
L=14	Dense	f=1024, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,1024,1)
L=15	Dense	f=512, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,512,1)
*	Entrée	$T_{prior}$	(B,#,1)	0	(B,6,1)
*	Entrée	$T_{suivi}$	(B,#,1)	0	(B,6,1)
-	Concatenate	-	(B,f+#,1)	0	(B,524,1)
L=16	Dense	f=256, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,256,1)
L=17	Dense	f=64, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,64,1)
L=18	Dense	f=16, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,16,1)
Sortie	Dense	f=#, linear	(B,f)	$(D_{L-1}+1)*f$	(B,6,1)
Total params	(local)				72.2M
Total params	(part)				39M
Total params	(glob)				139M

TABLE 4 – Résumés et détails par couche de l’architecture du réseau convolutif *Main-2D* utilisé pour des expériences de comparaison des architecture (décrit en section V.3.2) et illustré en Figure V.3.

Numéro couche	Type de couche	Caractéristiques	Dimension de sortie	Nombre paramètres	Exemple pour E=V=64
<b>Branche principale</b>					
*	Entrée	$I_t$	(B,H,W,1)	0	(B,256,256,1)
*	Entrée	$locale-V$	(B,H,W,V)	0	(B,256,256,V)
-	Concatenate	-	(B,H,W,V+1)	0	(B,256,256,V+1)
L=1	Conv2D	(k=5, f=64, s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,128,128,64)
L=2	Conv2D	(k=5, f=64, s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,64,64,64)
-	MaxPool2D	(k=2, s=1)	Idem	0	Idem
L=3	Conv2D	(k=3, f=64, s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,32,32,64)
L=4	Conv2D	(k=3, f=64, s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,64)
-	MaxPool2D	(k=2, s=1)	Idem	0	Idem
L=5	Conv2D	(k=5, f=128,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,128)
L=6	Conv2D	(k=5, f=128,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,128)
-	MaxPool2D	(k=2, s=1)	Idem	0	Idem
L=7	Conv2D	(k=3, f=128,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,128)
L=8	Conv2D	(k=3, f=128,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,128)
-	MaxPool2D	(k=2, s=1)	Idem	0	Idem
L=9	Conv2D	(k=5, f=256,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,256)
L=10	Conv2D	(k=5, f=256,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,256)
-	MaxPool2D	(k=2, s=1)	Idem	0	Idem
L=11	Conv2D	(k=3, f=256,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,256)
L=12	Conv2D	(k=3, f=256,s=1, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, f)$	$(k*k*D_{L-1}+1)*f$	(B,16,16,256)
-	MaxPool2D	(k=2, s=1)	Idem	0	Idem
-	AvgPool2D	(k=2, s=2)	$(B, \frac{H}{2^5}, \frac{W}{2^5}, f)$	0	(B,8,8,256)
-	Flatten	-	-	0	(B,16384,1)
<b>Branche parallèle</b>					
*	Entrée	$glob-V$ ou $part-V$	(B,H,W,V,1)	0	(B,256,256,64,1)
L=1	Conv3D	(k=5, f=64, s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, \frac{V}{2^L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,128,128,32,64)
L=2	Conv3D	(k=5, f=64, s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, \frac{V}{2^L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,64,64,16,64)
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
L=3	Conv3D	(k=3, f=128,s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, \frac{V}{2^L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,32,32,8,128)
L=4	Conv3D	(k=3, f=128,s=2, ReLu)	$(B, \frac{H}{2^L}, \frac{W}{2^L}, \frac{V}{2^L}, f)$	$(k*k*k*D_{L-1}+1)*f$	(B,16,16,4,128)
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
-	AvgPool3D	(k=2, s=2)	$(B, \frac{H}{2^5}, \frac{W}{2^5}, \frac{V}{2^5}, \frac{f}{2})$	0	(B,8,8,2,128)
-	Flatten	-	-	0	(B,16384,1)
<b>Concatenation et fin du réseau</b>					
-	Concatenate	-	-	0	(B,32768,1)
L=13	Dense	f=2048, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,2048,1)
L=14	Dense	f=1024, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,1024,1)
L=15	Dense	f=512, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,512,1)
Entrée*		$T_{prior}$	(B,#,1)	0	(B,6,1)
Entrée*		$T_{suivi}$	(B,#,1)	0	(B,6,1)
-	Concatenate	-	(B,f+#,1)	0	(B,524,1)
L=16	Dense	f=256, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,256,1)
L=17	Dense	f=64, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,64,1)

---

L=18	Dense	f=16, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,16,1)
Sortie	Dense	f=#,linear	(B,f)	$(D_{L-1}+1)*f$	(B,6,1)
Total params					75.8M

TABLE 5 – Résumés et détails par couche de l’architecture proposée *sJSR* (voir section VI.2.1) et illustré en Figure VI.1.

Numéro couche	Type de couche	Caractéristiques	Dimension de sortie	Nombre paramètres	Exemple pour E=128
<b>Branche principale</b>					
*	Entrée	$(I_t, I_{t-1})$	(B,H,W,2)	0	(B,256,256,3)
*	Entrée	$S_{prior}$	(B,H,W,1)	0	(B,256,256,1)
-	Concatenate	-	(B,H,W,3)	0	(B,256,256,3)
Level l=0	Conv2D	(k=5, f=E, s=1, ReLu)	(B,H,W,E,1)	$(k*k*D+1)*E$	(B,256,256,128,1)
level l=1	Conv3D	(k=5, f=64, s=2, ReLu)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{E}{2^l}, f)$	$(k*k*k*D_{l-1}+1)*f$	(B,128,128,64,64)
level l=2	Conv3D	(k=5, f=64, s=2, ReLu)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{E}{2^l}, f)$	$(k*k*k*D_{l-1}+1)*f$	(B,64,64,32,64)
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
level l=3	Conv3D	(k=3, f=128, s=2, ReLu)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{E}{2^l}, f)$	$(k*k*k*D_{l-1}+1)*f$	(B,32,32,16,128)
level l=4	Conv3D	(k=3, f=128, s=2, ReLu)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{E}{2^l}, f)$	$(k*k*k*D_{l-1}+1)*f$	(B,16,16,8,128)
-	MaxPool3D	(k=2, s=1)	Idem	0	Idem
level l=5	AvgPool3D	(k=2, s=2)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{E}{2^l}, \frac{f}{2})$	0	(B,8,8,4,128)
<b>Concatenation et fin du réseau</b>					
-	Flatten	-	-	0	(B,16384,1)
L=6	Dense	f=2048, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,2048,1)
L=7	Dense	f=1024, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,1024,1)
L=8	Dense	f=512, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,512,1)
*	Entrée	$T_{prior}$	(B,#,1)	0	(B,6,1)
*	Entrée	$T_{suivi}$	(B,#,1)	0	(B,6,1)
-	Concatenate	-	(B,f+#,1)	0	(B,524,1)
L=9	Dense	f=256, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,256,1)
L=10	Dense	f=64, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,64,1)
L=11	Dense	f=16, ReLu	(B,f)	$(D_{L-1}+1)*f$	(B,16,1)
Sortie	Dense	f=#, linear	(B,f)	$(D_{L-1}+1)*f$	(B,6,1)
<b>Branche multi-tâche</b>					
level l=4	ConvTransp.	(k=2, s=2, f=128)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{D}{2^l}, f)$	0	(B,16,16,8,128)
	Concatenate	-	-	0	(B,16,16,8,256)
	Conv3D	(k=3, f=128, s=1, ReLu)	Idem	$(k*k*k*D_{l+1}+1)*f$	(B,16,16,8,128)
level l=3	ConvTransp.	(k=2, s=2, f=128)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{D}{2^l}, f)$	0	(B,32,32,16,128)
	Concatenate	-	-	0	(B,32,32,16,256)
	Conv3D	(k=3, f=128, s=1, ReLu)	Idem	$(k*k*k*D_{l+1}+1)*f$	(B,32,32,16,128)
level l=2	ConvTransp.	(k=2, s=2, f=64)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{D}{2^l}, f)$	0	(B,64,64,32,64)
	Concatenate	-	-	0	(B,64,64,32,128)
	Conv3D	(k=5, f=64, s=1, ReLu)	Idem	$(k*k*k*D_{l+1}+1)*f$	(B,64,64,32,64)
level l=1	ConvTransp.	(k=2, s=2, f=64)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{D}{2^l}, f)$	0	(B,128,128,64,64)
	Concatenate	-	-	0	(B,128,128,64,128)
	Conv3D	(k=5, f=64, s=1, ReLu)	Idem	$(k*k*k*D_{l+1}+1)*f$	(B,128,128,64,64)
level l=0	ConvTransp.	(k=2, s=2, f=1)	$(B, \frac{H}{2^l}, \frac{W}{2^l}, \frac{D}{2^l}, f)$	0	(B,256,256,128,1)
	Concatenate	-	-	0	(B,256,256,128,3)
	Conv3D	(k=5, f=1, s=1, ReLu)	Idem	$(k*k*k*D_{l+1}+1)*f$	(B,256,256,128,1)
Sortie	Conv2D	(k=5, f=1, s=1, Sig.)	Idem	$(k*k*D+1)*f$	(B,256,256,1)
Total params					75.7M

---

## Annexe B : Comparaison des encodages des transformations

---

*E*n complément des expériences présentées dans le Chapitre IV, nous avons également comparé 3 types d'encodage différents pour représenter les transformations de sortie  $T_t$  : les angles d'Euler, le couple sin/cos, les coordonnées de l'image. Chacun des ses encodages a été testé en adaptant le type d'architecture du réseau et en adaptant la fonction d'optimisation associée. Les différentes expériences et résultats sont présentés dans cette annexe.

## Définitions des encodages

Les données de sortie du réseau sont les prédictions de localisation optimale  $\hat{T}_t$  de chaque image courante  $I_t$  par rapport au volume de référence  $V_{ref}$ . Nous avons vu précédemment (voir section II.4.2), qu'il existe différentes façons de représenter une transformation rigide et notamment le paramétrage de la rotation. Parmi les différentes représentations envisagées dans la littérature, nous avons choisi de comparer les angles d'Euler, le couple sin/cos et les coordonnées de l'image.

- Angles d'Euler : tels que  $\hat{T}_t = (\theta x, \theta y, \theta z, tx, ty, tz)$ , avec  $\# = 6$  pour une représentation de tous les paramètres rigides. Les angles sont exprimés en degrés tandis que les translations sont exprimées en millimètres (mm). Les encodages des vérités terrains ( $T_t$ ) et des données d'entrée comme  $T_{prior}$  et  $T_{suivi}$  suivent ce même type d'encodage. Les expériences associées à cet encodage sont intitulées « *Euler* ».
- Couple sin/cos : tel que  $\hat{T}_t = (\sin(\theta x), \sin(\theta y), \sin(\theta z), \cos(\theta x), \cos(\theta y), \cos(\theta z), tx, ty, tz)$ , avec  $\# = 9$  pour une représentation de tous les paramètres rigides. Les valeurs des rotations sont bornées entre -1 et 1 pour chaque angle exprimé en radians tandis que les translations sont toujours exprimées en millimètres. L'encodage des vérités terrains ( $T_t$ ) et des données d'entrée comme  $T_{prior}$  et  $T_{suivi}$  utilise ce même encodage. Les expériences associées à cet encodage sont intitulées « *Sin/Cos* ».
- Coordonnées : correspondant aux coordonnées physiques (position) de chaque pixel de l'image à localiser. La prédiction  $\hat{T}_t$  peut-être convertie sous deux formes : (i) en liste des coordonnées de chaque point de l'image ou (ii) en image à 3 canaux. Le premier cas permet de conserver une forme vectorielle 1D de l'encodage (de dimension  $\# = H * W * 3$ ) et de garder l'architecture principale, tandis que le deuxième cas permet de traiter le problème de recalage sous la forme image-vers-image (de dimension  $\# = H \times W \times 3$ ) et nécessitera une architecture spécifique (décrite dans la section suivante). Les expériences associées à cet encodage sont respectivement intitulées « *Coord-1D* » et « *Coord-2D* » selon la forme des coordonnées (vectorielle 1D ou image 2D). Dans tous les cas, l'entrée  $T_{suivi}$ , donnant des informations sur les déplacements relatifs, ne peut être encodée par des coordonnées de position. Cette entrée sera exceptionnellement combinée à celle du  $T_{prior}$ , en appliquant la transformée  $T_{suivi}$  à chaque point de l'image associé à l'entrée  $T_{prior}$ .

## Architectures

- Architecture principale :

Nous rappelons que l'architecture principale proposée dans le Chapitre 4 a été précédemment décrite en section IV.2.1.2 et illustrée dans la Figure IV.1. Cette architecture sera utilisée pour la majorité des expériences d'encodage (« *Euler* »; « *Sin/Cos* », « *Coord-1D* »).

- Autre architecture :

Nous avons adapté notre architecture principale pour convenir au cas spécifique d'encodage « *Coord-2D* » décrit précédemment. Pour cela, nous avons utilisé un réseau de type U-net [105], présenté sur la Figure 1. Celui-ci présente l'avantage d'agir comme un encodeur-décodeur, prédisant les coordonnées sous le même format que les images d'entrée. L'entrée  $T_{prior}$ , également sous la forme d'image

( $\# = H \times W \times 3$ ) est alors concaténée à la paire d'images en entrée du réseau. Nous rappelons que la transformée d'entrée  $T_{suivi}$ , ne pouvant être encodée sous la forme de plan, sera directement appliquée pour transformer chaque point du plan de l'entrée  $T_{prior}$ . Ce réseau est également décrit en Annexe (voir Table 2) et totalise environ 12 M de paramètres sur l'ensemble des couches.

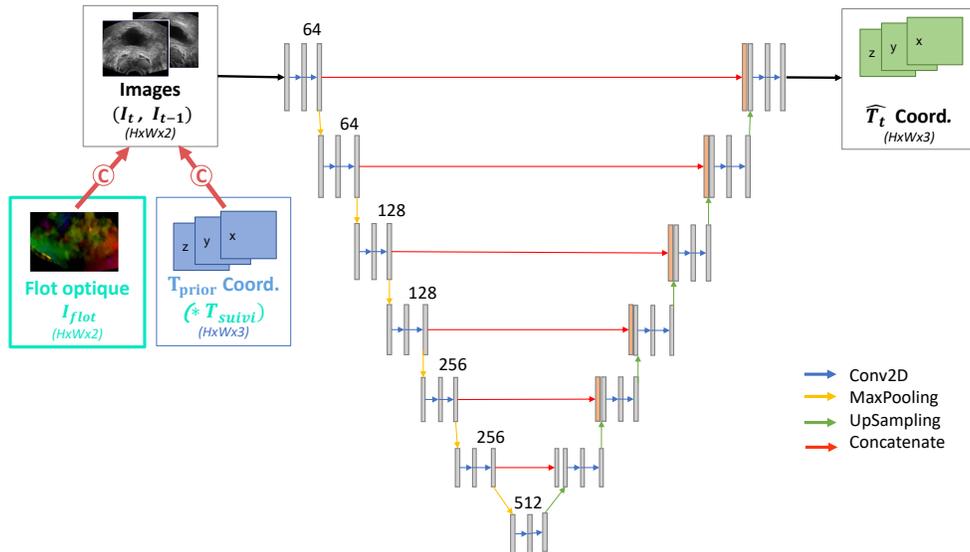


FIGURE 1 – Architecture du réseau de U-net proposé pour le cas particulier d'encodage de coordonnées « *Coord-2D* », sous la forme d'une image à 3 canaux de dimension ( $H \times W \times 3$ ).

## Fonctions de coût

En complément de la fonction principale  $L_{MSE}$  (voir section IV.2.2.1), nous avons évalué d'autres termes plus spécifiques et plus adaptés aux différents types d'encodage de sortie testés. Ceux-ci seront comparés à la fonction principale par le biais de différentes expériences. Par défaut, si aucune mention n'est spécifiée, la fonction utilisée est la fonction principale  $L_{MSE}$ .

- Terme de pénalisation mathématique :

Bien que certaines études recommandent le couple sin/cos pour représenter les rotations de manière plus robuste, plus homogène et plus continue (voir section II.4.2.1), aucune à notre connaissance n'impose de contrainte supplémentaire pour préserver la dépendance entre ces termes. Ainsi, nous avons ajouté à notre fonction de coût, un terme de régularisation visant à maintenir l'égalité mathématique et géométrique  $\cos^2(\theta) + \sin^2(\theta) = 1$  pour chaque angle  $\theta$  considéré. Pour cela, chaque couple  $\hat{T}_{k,sin}(\theta) / \hat{T}_{k,cos}(\theta)$  associé respectivement à l'encodage sin/cos de la transformée  $\hat{T}_k$  d'un élément  $k$ , est utilisé pour la pénalisation. La fonction de coût  $L_{Sin/Cos}$  est présentée dans l'équation 1a et

sera dédiée à l'encodage des sorties sous la forme « *Sin/Cos* ».

$$L_{Sin/Cos} = \frac{1}{K} \sum_{k=1}^K MSE_k + GEO_k \quad (1a)$$

$$\text{avec } GEO_k = \left\| 1 - (\hat{T}_{k,cos}(\theta)^2 + \hat{T}_{k,sin}(\theta)^2) \right\| \quad (1b)$$

- Terme de distance géométrique d'alignement :

Avec l'encodage sous la forme des coordonnées cartésiennes des pixels de l'image à localiser, l'idée est de réussir à créer une fonction de coût qui soit révélatrice de la qualité de notre recalage d'un point de vue géométrique. Pour cela, nous avons choisi d'appliquer le calcul de TRE. La mesure est calculée entre les coordonnées des points à la position prédite  $\hat{p}_{k,i}$  et des points à la position de vérité terrain  $p_{k,i}$  avec  $i = (1, \dots, N)$ . Le calcul est défini dans l'équation 2a pour tous les éléments  $k$  de l'entraînement. Cette fonction  $L_{Coord}$  sera dédiée aux expériences liées à l'encodage sous la forme de coordonnées (« *Coord-1D* » ou « *Coord-2D* »)

$$L_{Coord} = \frac{1}{K} \sum_{k=1}^K TRE_k \quad (2a)$$

$$\text{avec } TRE_k = \sqrt{\frac{\sum_{i=1}^N \|p_{k,i} - \hat{p}_{k,i}\|^2}{N}} \quad (2b)$$

## Expériences

Les conditions expérimentales sont les mêmes que celles décrites dans le Chapitre IV (voir section IV.3.4). Pour le jeu de données utilisé (Table IV.1), seuls les 3 paramètres de rotation ont été estimés ( $\# = 3$ ). Sans paramètre de translation, les dimensions des sorties selon l'encodage choisi sont :

- $T_t = (\theta x, \theta y, \theta z)$  pour l'encodage « *Euler* »
- $T_t = (\sin(\theta x), \sin(\theta y), \sin(\theta z), \cos(\theta x), \cos(\theta y), \cos(\theta z))$  pour l'encodage « *Sin/Cos* »
- $T_t = (130 \times 250 \times 3)$  pour l'encodage des coordonnées « *Coord-2D* »
- $T_t = (130 * 250 * 3, 1)$  pour l'encodage des coordonnées « *Coord-1D* ».

## Résultats et Discussion

Dans les résultats qui vont suivre, seules les performances sur le scénario 3 avec  $T_{prior}$  (*par portion*) sont présentées. Les autres expériences suivent les mêmes tendances et conclusions. La Table 6 résume l'ensemble des valeurs moyennes de TRE obtenus pour les 4 différents encodages testés.

TABLE 6 – Comparaison des encodages.

Évaluation séquentielle <i>par portion</i>	TRE ( $\downarrow$ ) : $\mu \pm \sigma$ (mm)			
	TRE avant recalage : 27.47 ( $\pm$ 16,68)			
	<i>Euler</i>	<i>Sin/Cos</i>	<i>Coord-1D</i>	<i>Coord-2D</i>
Sc.3 Avec $T_{prior}$				
$L_{MSE}$	1.91 ( $\pm$ 1.26)	8.15 ( $\pm$ 5.00)	24.57 ( $\pm$ 3.66)	28.06 ( $\pm$ 5.24)
$L_{Sin/Cos}$	N/A	9.64 ( $\pm$ 4.18)	N/A	N/A
$L_{Coord}$	N/A	N/A	27.8 ( $\pm$ 3.72)	28.8 ( $\pm$ 6.94)

Par rapport à l'encodage par défaut, sous la forme d'angles d'Euler, on constate que l'utilisation d'autres types d'encodage résulte en de moins bonnes performances.

La Figure 2 montre l'évolution des paramètres de l'encodage sin/cos le long d'une trajectoire. On remarque que les prédictions du sinus suivent assez bien la trajectoire tandis que les prédictions du cosinus semblent au contraire constantes. L'ajout du terme de pénalité  $L_{Sin/Cos}$  au sein de la fonction de coût permet de mieux conserver la relation géométrique (voir les distributions Figure 2), au détriment de l'évolution des prédictions au cours de la trajectoire.

Des exemples de visualisation des coordonnées des pixels de l'image *Coord-1D* et *Coord-2D* sont présentés sur la Figure 3 . On remarque que les prédictions des coordonnées sont plus ou moins réalistes : l'espacement entre les points n'est pas toujours respecté et la surface de l'image n'est pas toujours plane. Nous ne notons pas de différence significative entre les performances de l'encodage vectoriel (1D) ou de l'encodage image (2D), ni avec l'utilisation de la fonction de coût spécifique  $L_{Coord}$ .

Finalement, parmi les différents encodages testés et les fonctions de coût respectivement associées, il semble que l'utilisation d'angles d'Euler et d'une fonction de coût basée sur la MSE permettent d'obtenir la meilleure configuration possible. Malgré les problématiques connues liées à la non-continuité et à la non-périodicité de l'encodage sous la forme d'angles d'Euler, les prédictions obtenues respectent l'intervalle et les amplitudes des angles à établir, avec une simple fonction d'activation linéaire "ReLU", et démontrent un bon apprentissage du réseau.

Cependant l'encodage sous forme de coordonnées n'a pas été davantage exploré, et avant d'éliminer complètement cette option, il faudrait peut-être envisager d'y ajouter des termes de contraintes structurelles pour imposer à la fois un espacement régulier entre les points du voisinage ou encore imposer une surface de prédiction plane (dans le cadre de recalage rigides). Ces aspects seront discutés dans le Chapitre VII.

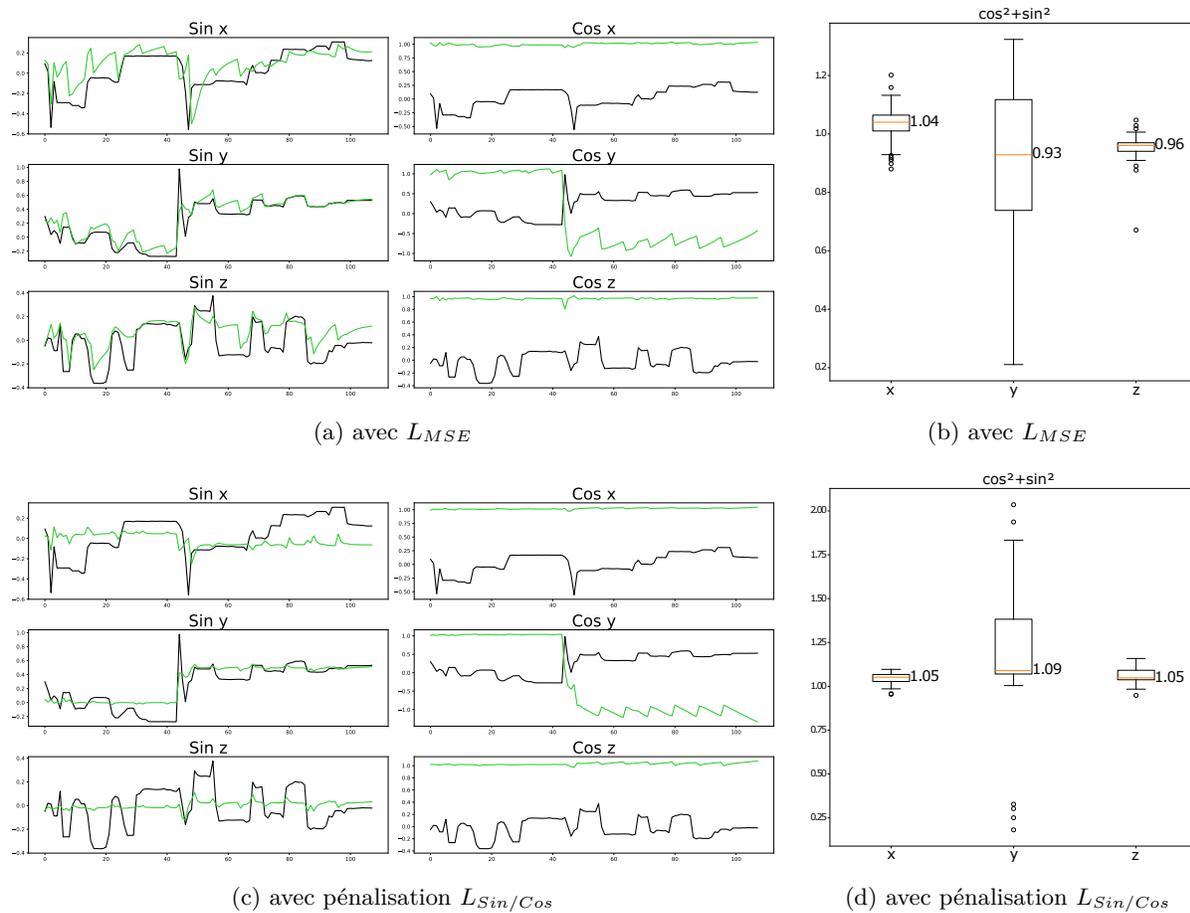


FIGURE 2 – Analyse temporelle de l'évolution des paramètres de l'encodage Sin/Cos le long d'une trajectoire (à gauche). Distribution de la conservation mathématique de l'égalité remarquable  $\cos^2 + \sin^2 = 1$  (à droite). Comparaison des expériences selon la fonction de coût utilisée : principale  $L_{MSE}$  ou celle spécifique  $L_{Sin/Cos}$

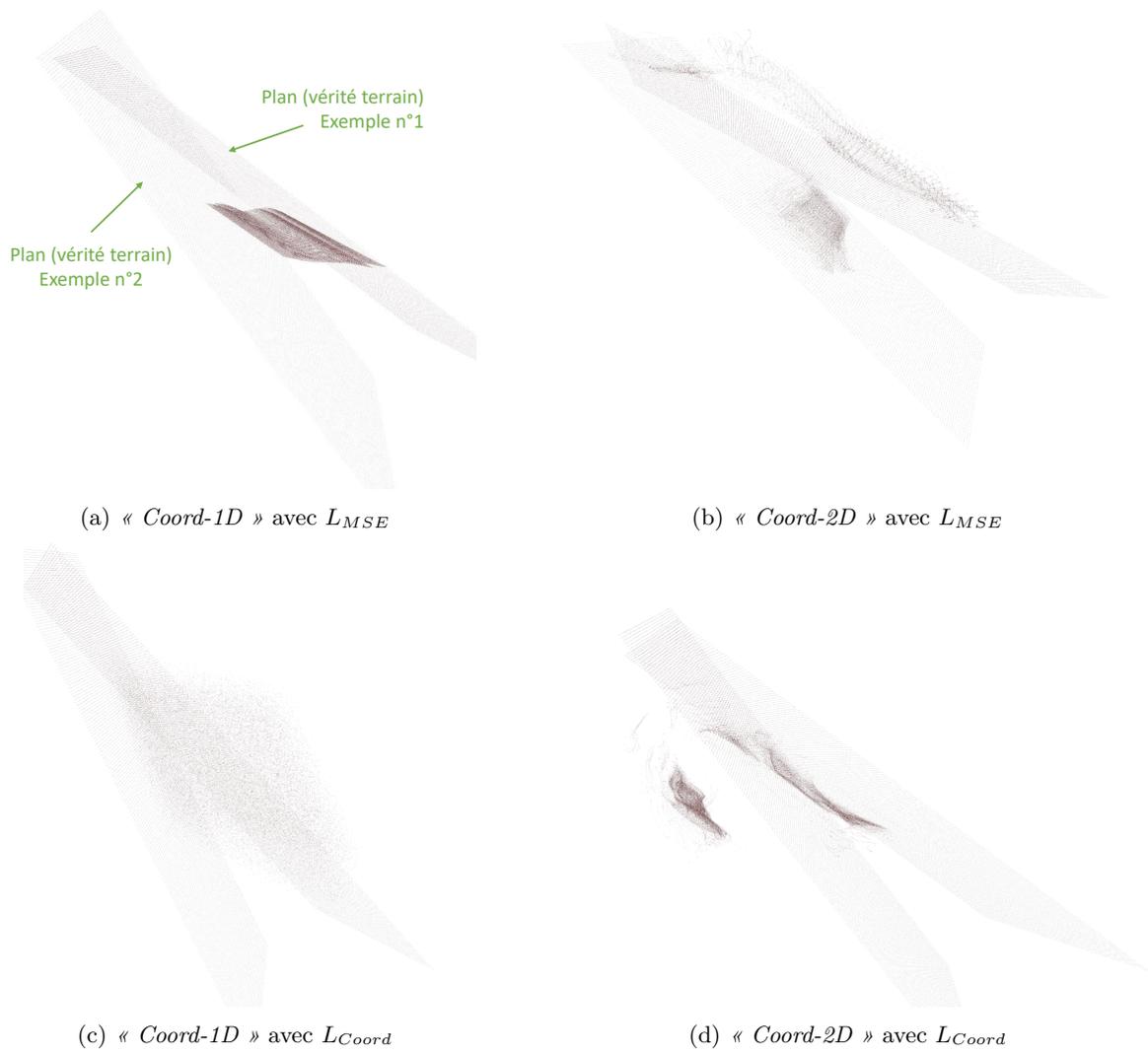


FIGURE 3 – Visualisation de deux exemples de prédictions des coordonnées de l'image selon le format de l'encodage testé (Coord-1D ou Coord-2D) et selon les fonctions de coût utilisées.



---

## Bibliographie

---

- [1] J. E. McNeal, “The zonal anatomy of the prostate,” *The Prostate*, vol. 2, no. 1, pp. 35–49, 1981.
- [2] C. J. o. Das, A. Razik, S. Sharma, and S. Verma, “Prostate biopsy : When and how to perform,” *Clinical Radiology*, vol. 74, no. 11, pp. 853–864, 2019.
- [3] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global cancer statistics 2020 : Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA : A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [4] P. Rawla, “Epidemiology of prostate cancer,” *World Journal of Oncology*, vol. 10, no. 2, pp. 63–89, 2019.
- [5] N. Mottet, R. C. N. van den Bergh, E. Briers, T. Van den Broeck, M. G. Cumberbatch, M. De Santis, S. Fanti, N. Fossati, G. Gandaglia, S. Gillessen, N. Grivas, J. Grummet, A. M. Henry, T. H. van der Kwast, T. B. Lam, M. Lardas, M. Liew, M. D. Mason, L. Moris, D. E. Oprea-Lager, H. G. van der Poel, O. Rouvière, I. G. Schoots, D. Tilki, T. Wiegel, P.-P. M. Willemse, and P. Cornford, “Eau-eanm-estro-esur-siog guidelines on prostate cancer—2020 update. part 1 : Screening, diagnosis, and local treatment with curative intent,” *European Urology*, vol. 79, no. 2, pp. 243–262, 2021.
- [6] M. S. Litwin and H.-J. Tan, “The diagnosis and treatment of prostate cancer : A review,” *JAMA*, vol. 317, no. 24, pp. 2532–2542, 2017.
- [7] M. D. Michaelson, S. E. Cotter, P. C. Gargollo, A. L. Zietman, D. M. Dahl, and M. R. Smith, “Management of complications of prostate cancer treatment,” *CA : A Cancer Journal for Clinicians*, vol. 58, no. 4, pp. 196–213, 2008.
- [8] M. Ahdoot, A. H. Lebastchi, B. Turkbey, B. Wood, and P. A. Pinto, “Contemporary treatments in prostate cancer focal therapy,” *Current opinion in oncology*, vol. 31, no. 3, p. 200, 2019.

- 
- [9] E. Even-Sapir, U. Metser, E. Mishani, G. Lievshitz, H. Lerman, and I. Leibovitch, “The detection of bone metastases in patients with high-risk prostate cancer : 99mTc-mdp planar bone scintigraphy, single- and multi-field-of-view spect, 18F-fluoride pet, and 18F-fluoride pet/ct,” *Journal of Nuclear Medicine*, vol. 47, no. 2, pp. 287–297, 2006.
- [10] M. Mitterberger, W. Horninger, F. Aigner, G. Pinggera, I. Steppan, P. Rehder, and F. Frauscher, “Ultrasound of the prostate,” *Cancer imaging : the official publication of the International Cancer Imaging Society*, vol. 10, pp. 40–8, 03 2010.
- [11] “PET Scanning For Prostate Cancer,” Jan. 2023. [Online]. Available : <https://www.petimagingflorida.com/pet-scanning-for-prostate-cancer/>
- [12] K. K. Hodge, J. E. McNeal, M. K. Terris, and T. A. Stamey, “Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate,” *Journal of Urology*, vol. 142, no. 1, pp. 71–74, 1989.
- [13] K. K. Hodge, J. E. McNeal, and T. A. Stamey, “Ultrasound guided transrectal core biopsies of the palpably abnormal prostate,” *Journal of Urology*, vol. 142, no. 1, pp. 66–70, 1989.
- [14] A. Jager, J. C. Vilanova, M. Michi, H. Wijkstra, and J. R. Oddens, “The challenge of prostate biopsy guidance in the era of mpMRI detected lesion : Ultrasound-guided versus in-bore biopsy,” *The British Journal of Radiology*, vol. 95, no. 1131, p. 20210363, 2022.
- [15] V. M. Skouteris, E. D. Crawford, V. Mouraviev, P. Arangua, M. P. Metsinis, M. Skouteris, G. Zacharopoulos, and N. N. Stone, “Transrectal ultrasound-guided versus transperineal mapping prostate biopsy : Complication comparison,” *Reviews in Urology*, vol. 20, no. 1, pp. 19–25, 2018.
- [16] J. C. Presti, “Prostate biopsy : Current status and limitations,” *Reviews in Urology*, vol. 9, no. 3, pp. 93–98, 2007.
- [17] M. Baumann, P. Mozer, V. Daanen, and J. Troccaz, “Prostate biopsy tracking with deformation estimation,” *Medical Image Analysis*, vol. 16, no. 3, pp. 562–576, apr 2012.
- [18] T. De Silva, A. Fenster, J. Bax, C. Romagnoli, J. Izawa, J. Samarabandu, and A. D. Ward, “Quantification of prostate deformation due to needle insertion during trus-guided biopsy : Comparison of hand-held and mechanically stabilized systems,” *Medical Physics*, vol. 38, no. 3, pp. 1718–1731, 2011.
- [19] G. Fiard, “Apprentissage des biopsies prostatiques par la simulation : Vers la validation du simulateur biopsym,” Ph.D. dissertation, Université Grenoble Alpes, 2018.
- [20] D. J. o. Gillies, L. Gardi, T. De Silva, S.-r. Zhao, and A. Fenster, “Real-time registration of 3d to 2d ultrasound images for image-guided prostate biopsy,” *Medical Physics*, vol. 44, no. 9, pp. 4708–4723, 2017.
- [21] V. Karnik, A. Fenster, J. Bax, D. Cool, L. Gardi, I. Gyacskov, C. Romagnoli, and A. Ward, “Assessment of image registration accuracy in three-dimensional transrectal ultrasound guided prostate biopsy,” *Medical Physics*, vol. 37, pp. 802–13, feb 2010.
- [22] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, “Medical image registration,” *Physics in medicine & biology*, vol. 46, no. 3, p. R1, 2001.
- [23] F. P. Oliveira and J. M. R. Tavares, “Medical image registration : A review,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, no. 2, pp. 73–93, 2014.

- 
- [24] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, “Deep learning in medical image registration : A review,” *Physics in Medicine & Biology*, vol. 65, no. 20, p. 20TR01, 2020.
- [25] J. B. West, J. o. M. Fitzpatrick, S. A. Toms, C. R. Maurer, and R. J. o. Maciunas, “Fiducial point placement and the accuracy of point-based, rigid body registration,” *NeuroSurgery*, vol. 48, no. 4, pp. 810–817, apr 2001.
- [26] H. R. Boveiri, R. Khayami, R. Javidan, and A. Mehdizadeh, “Medical image registration using deep neural networks : a comprehensive review,” *Computing Elec. Engineering*, vol. 87, p. 106767, 2020.
- [27] G. Christensen and H. o. Johnson, “Consistent image registration,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 568–582, 2001.
- [28] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [29] G. Haskins, U. Kruger, and P. Yan, “Deep learning in medical image registration : A survey,” *Machine Vision and Applications*, vol. 31, no. 1-2, p. 8, 2020.
- [30] X. Chen, A. Diaz-Pinto, N. Ravikumar, and A. Frangi, “Deep learning in medical image registration,” *Progress in Biomedical Engineering*, 2020.
- [31] F. Cornud, L. Brolis, N. B. Delongchamps, D. Portalez, B. Malavaud, R. Renard-Penna, and P. Mozer, “Trus–mri image registration : A paradigm shift in the diagnosis of significant prostate cancer,” *Abdominal Imaging*, vol. 38, no. 6, pp. 1447–1463, dec 2013.
- [32] A. M. Brown, O. Elbuluk, F. Mertan, S. Sankineni, D. J. o. Margolis, B. J. o. Wood, P. A. Pinto, P. L. Choyke, and B. Turkbey, “Recent advances in image-guided targeted prostate biopsy,” *Abdominal Imaging*, vol. 40, no. 6, pp. 1788–1799, aug 2015.
- [33] P. A. Pinto, P. H. Chung, A. R. Rastinehad, A. A. Baccala, J. Kruecker, C. J. o. Benjamin, S. Xu, P. Yan, S. Kadoury, C. Chua, J. K. Locklin, B. Turkbey, J. H. Shih, S. P. Gates, C. Buckner, G. Bratslavsky, W. M. Linehan, N. D. Glossop, P. L. Choyke, and B. J. o. Wood, “Magnetic resonance imaging/ultrasound fusion guided prostate biopsy improves cancer detection following transrectal ultrasound biopsy and correlates with multiparametric magnetic resonance imaging,” *Journal of Urology*, vol. 186, no. 4, pp. 1281–1285, 2011.
- [34] S. Xu, J. Kruecker, B. Turkbey, N. Glossop, A. K. Singh, P. Choyke, P. Pinto, and B. J. o. Wood, “Real-time mri-trus fusion for guidance of targeted prostate biopsies,” *Computer Aided Surgery*, vol. 13, no. 5, pp. 255–264, 2008.
- [35] J. Bax, D. Cool, L. Gardi, K. Knight, D. Smith, J. Montreuil, S. Sherebrin, C. Romagnoli, and A. Fenster, “Mechanically assisted 3d ultrasound guided prostate biopsy system : Mechanically assisted 3d ultrasound guided prostate biopsy system,” *Medical Physics*, vol. 35, no. 12, pp. 5397–5410, 2008.
- [36] N. Zhu, M. Najafi, B. Han, S. Hancock, and D. Hristov, “Feasibility of image registration for ultrasound-guided prostate radiotherapy based on similarity measurement by a convolutional neural network,” *Technology in Cancer Research & Treatment*, vol. 18, p. 153303381882196, 2019.

- 
- [37] H. Guo, M. Kruger, S. Xu, B. J. o. Wood, and P. Yan, “Deep adaptive registration of multimodal prostate images,” *Computerized Medical Imaging and Graphics*, vol. 84, p. 101769, 2020.
- [38] Y. Fu, Y. Lei, T. Wang, P. Patel, A. B. Jani, H. Mao, W. J. o. Curran, T. Liu, and X. Yang, “Biomechanically constrained non-rigid mr-trus prostate registration using deep learning based 3d point cloud matching,” *Medical Image Analysis*, vol. 67, p. 101845, 2021.
- [39] M. S. Elmahdy, L. Beljaards, S. Yousefi, H. Sokooti, F. Verbeek, U. A. Van Der Heide, and M. Staring, “Joint registration and segmentation via multi-task learning for adaptive radiotherapy of prostate cancer,” *IEEE Access*, vol. 9, pp. 95 551–95 568, 2021.
- [40] Q. Zeng, Y. Fu, Z. Tian, Y. Lei, Y. Zhang, T. Wang, H. Mao, T. Liu, W. J. o. Curran, A. B. Jani, P. Patel, and X. Yang, “Label-driven magnetic resonance imaging (mri)-transrectal ultrasound (trus) registration using weakly supervised learning for mri-guided prostate radiotherapy,” *Physics Medical Ampmathsemicolon Biology*, vol. 65, no. 13, p. 135002, jun 2020.
- [41] Y. Hu, M. Modat, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, J. o. A. Noble, D. C. Barratt, and T. Vercauteren, “Label-driven weakly-supervised learning for multimodal deformable image registration,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 1070–1074.
- [42] G. Haskins, J. Kruecker, U. Kruger, S. Xu, P. A. Pinto, B. J. o. Wood, and P. Yan, “Learning deep similarity metric for 3d mr–trus image registration,” *International Journal of Computing Assisted Radiology Surgery*, vol. 14, no. 3, pp. 417–425, oct 2018.
- [43] Y. Chen, L. Xing, L. Yu, W. Liu, B. Pooya Fahimian, T. Niedermayr, H. P. Bagshaw, M. Buyyounouski, and B. Han, “Mr to ultrasound image registration with segmentation-based learning for hdr prostate brachytherapy,” *Medical Physics*, vol. 48, no. 6, pp. 3074–3083, 2021.
- [44] Q. Zeng, Y. Fu, J. Jeong, Y. Lei, T. Wang, H. Mao, A. B. Jani, P. Patel, W. J. o. Curran, T. Liu, and X. Yang, “Weakly non-rigid mr-trus prostate registration using fully convolutional and recurrent neural networks,” in *Medical Imaging 2020 : Image Processing*, I. Išgum and B. A. Landman, Eds., vol. 11313, 2020, pp. 754–760.
- [45] S. Zhang, S. Jiang, Z. Yang, and R. Liu, “2d ultrasound and 3d mr image registration of the prostate for brachytherapy surgical navigation,” *Medicine (Baltimore)*, vol. 94, no. 40, 2015.
- [46] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, “Robust non-rigid registration through agent-based action learning,” in *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2017*, ser. Lecture Notes in Computer Science, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds., 2017, vol. 10433, pp. 344–352.
- [47] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, S. Ourselin, J. o. A. Noble, D. C. Barratt, and T. Vercauteren, “Weakly-supervised convolutional neural networks for multimodal image registration,” *Medical Image Analysis*, vol. 49, pp. 1–13, 2018.
- [48] Z. M. Baum, Y. Hu, and D. C. Barratt, “Real-time multimodal image registration with partial intraoperative point-set data,” *Medical image Analysis*, vol. 74, p. 102231, 2021.

- 
- [49] D. Tessier, D. J. o. Gillies, L. Gardi, A. Mercado, and A. Fenster, “Ring navigation : An ultrasound-guided technique using real-time motion compensation for prostate biopsies,” in *Medical Imaging 2018 : Image-Guided Procedures, Robotic Interventions, and Modeling*, R. J. o. Webster and B. Fei, Eds., 2018, p. 52.
- [50] T. De Silva, A. Fenster, D. W. Cool, L. Gardi, C. Romagnoli, J. Samarabandu, and A. D. Ward, “2d-3d rigid registration to compensate for prostate motion during 3d trus-guided biopsy,” *Medical Physics*, vol. 40, no. 2, p. 022904, feb 2013.
- [51] S.-Y. Selmi, E. Promayon, and J. Troccaz, “Hybrid 2d-3d ultrasound registration for navigated prostate biopsy,” *International J Computing Assisted Radiology Surgery*, vol. 13, no. 7, pp. 987–995, jul 2018.
- [52] C. Beitone, G. Fiard, and J. Troccaz, “Towards real-time free-hand biopsy navigation,” *Medical Physics*, vol. 48, no. 7, pp. 3904–3915, jul 2021.
- [53] S. Y. Selmi, “Navigation en temps-réel pour la biopsie de prostate,” Ph.D. dissertation, Université Grenoble Alpes (ComUE), 2017.
- [54] A. Bhardwaj, J.-S. Park, S. Mukhopadhyay, S. Sharda, Y. Son, B. Ajani, and S. R. Kudavelly, “Rigid and deformable corrections in real-time using deep learning for prostate fusion biopsy,” in *Medical Imaging 2020 : Image-Guided Procedures, Robotic Interventions, and Modeling*, B. Fei and C. A. Linte, Eds., vol. 11315, 2020, pp. 486–499.
- [55] A. Bhardwaj, P. Mathur, T. Singh, V. Suryanarayana, Y. Son, S. R. Kudavelly, S. Song, and H. Kang, “An approach for live motion correction for trus-mr prostate fusion biopsy using deep learning,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 2993–2996.
- [56] H. Guo, X. Xu, S. Xu, B. J. o. Wood, and P. Yan, “End-to-end ultrasound frame to volume registration,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021*, 2021, pp. 56–65.
- [57] H. Guo, X. Xu, X. Song, S. Xu, H. Chao, J. Myers, B. Turkbey, P. A. Pinto, B. J. o. Wood, and P. Yan, “Ultrasound frame-to-volume registration via deep learning for interventional guidance,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2022.
- [58] B. Hou, A. Alansary, S. McDonagh, A. Davidson, M. Rutherford, J. V. Hajnal, D. Rueckert, B. Glocker, and B. Kainz, “Predicting slice-to-volume transformation in presence of arbitrary subject motion,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, 2017, pp. 296–304.
- [59] T.-F. Yu, P. Liu, Y. Peng, J. Liu, H. Yin, and D. C. Liu, “Slice localization for three-dimensional breast ultrasound volume using deep learning,” *DEStech Transactions Engineering Technology Res.*, 2019.
- [60] S. S. Mohseni Salehi, S. Khan, D. Erdogmus, and A. Gholipour, “Real-time deep pose estimation with geodesic loss for image-to-template rigid registration,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 470–481, 2019.
- [61] B. Almogadwy, K. McLeod, and A. Burger, “A deep learning approach for slice to volume biomedical image integration,” in *Proceedings of the 2019 11th International Conference Bioinf. Biomedical Technology*, ser. ICBBT’19, Stockholm, Sweden, 2019, pp. 62–68.

- 
- [62] Y. Li, B. Khanal, B. Hou, A. Alansary, J. J. o. Cerrolaza, M. Sinclair, J. Matthew, C. Gupta, C. Knight, B. Kainz, and D. Rueckert, “Standard plane detection in 3d fetal ultrasound using an iterative transformation network,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018*, 2018, pp. 392–400.
- [63] Z. Chen, Z. Xu, W. Yi, X. Yang, W. Hou, M. Ding, and O. Granichin, “Real-time and multimodal brain slice-to-volume registration using cnn,” *Expert Systems with Applications*, vol. 133, pp. 86–96, 2019.
- [64] C. Di Vece, B. Dromey, F. Vasconcelos, A. L. David, D. Peebles, and D. Stoyanov, “Deep learning-based plane pose regression in obstetric ultrasound,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 5, pp. 833–839, 2022.
- [65] S. Miao, Z. J. Wang, and R. Liao, “A cnn regression approach for real-time 2d/3d registration,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.
- [66] W. Wei, M. Rak, J. Alpers, and C. Hansen, “Towards fully automatic 2d us to 3d ct/mr registration : A novel segmentation-based strategy,” in *2020 IEEE 17th International S. Biomedical Imaging (ISBI)*, 2020, pp. 433–437.
- [67] Y. Fu, Y. Lei, T. Wang, M. Axente, J. Roper, J. D. Bradley, T. Liu, and X. Yang, “Deep learning based volume-to-slice mri registration via intentional overfitting,” in *Medical Imaging 2022 : Image-Guided Procedures, Robotic Interventions, and Modeling*, C. A. Linte and J. H. Siewerdsen, Eds., 2022, p. 50.
- [68] Y. Zhao, K. Zeng, Y. Zhao, P. Bhatia, M. Ranganath, M. L. Kozhikkavil, C. Li, and G. Hermosillo, “Deep learning solution for medical image localization and orientation detection,” *Medical Image Analysis*, vol. 81, p. 102529, 2022.
- [69] V. Markova, M. Ronchetti, W. Wein, O. Zettinig, and R. Prevost, “Global multi-modal 2d/3d registration via local descriptors learning,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022 : 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, 2022, pp. 269–279.
- [70] M. A. Fischler and R. C. Bolles, “Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [71] M. Luo, X. Yang, X. Huang, Y. Huang, Y. Zou, X. Hu, N. Ravikumar, A. F. Frangi, and D. Ni, “Self context and shape prior for sensorless freehand 3d ultrasound reconstruction,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021*, 2021, pp. 201–210.
- [72] H. Guo, S. Xu, B. Wood, and P. Yan, “Sensorless freehand 3d ultrasound reconstruction via deep contextual learning,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2020*, 2020, pp. 463–472.
- [73] J. Xu, D. Moyer, P. E. Grant, P. Golland, J. E. Iglesias, and E. Adalsteinsson, “Svort : Iterative transformer for slice-to-volume registration in fetal brain mri,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2022 : 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, 2022, pp. 3–13.

- 
- [74] M. Luo, X. Yang, H. Wang, L. Du, and D. Ni, “Deep motion network for freehand 3d ultrasound reconstruction,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022 : 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*, 2022, pp. 290–299.
- [75] G. Ning, H. Liang, L. Zhou, X. Zhang, and H. Liao, “Spatial position estimation method for 3d ultrasound reconstruction based on hybrid transformers,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 2022, pp. 1–5.
- [76] R. Prevost, M. Salehi, S. Jagoda, N. Kumar, J. Sprung, A. Ladikos, R. Bauer, O. Zettinig, and W. Wein, “3d freehand ultrasound without external tracking using deep learning,” *Medical Image Analysis*, vol. 48, pp. 187–202, 2018.
- [77] T. Leblanc, F. Lalys, Q. Tollenaere, A. Kaladji, A. Lucas, and A. Simon, “Stretched reconstruction based on 2d freehand ultrasound for peripheral artery imaging,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 7, pp. 1281–1288, 2022.
- [78] K. Miura, K. Ito, T. Aoki, J. Ohmiya, and S. Kondo, “Localizing 2d ultrasound probe from ultrasound image sequences using deep learning for volume reconstruction,” in *Medical Ultrasound, Preterm, Perinatal Paediatric Image Analysis*, 2020, pp. 97–105.
- [79] A. Singh, S. S. M. Salehi, and A. Gholipour, “Deep predictive motion tracking in magnetic resonance imaging : Application to fetal imaging,” *IEEE Transactions Medical Imaging*, vol. 39, no. 11, pp. 3523–3534, 2020.
- [80] P.-H. Yeung, M. Aliasi, A. T. Papageorghiou, M. Haak, W. Xie, and A. I. L. Namburete, “Learning to map 2d ultrasound images into 3d space with minimal human annotation,” *Medical Image Analysis*, vol. 70, p. 101998, 2021.
- [81] M. Mikaeili and H. Şakir Bilge, “Trajectory estimation of ultrasound images based on convolutional neural network,” *Biomedical Signal Processing and Control*, vol. 78, p. 103965, 2022.
- [82] L. Qiu and H. Ren, “Rsegnet : A joint learning framework for deformable registration and segmentation,” *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 2499–2513, 2022.
- [83] L. Canalini, J. Klein, D. Miller, and R. Kikinis, “Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 10, pp. 1697–1713, 2019.
- [84] T. Estienne, M. Vakalopoulou, S. Christodoulidis, E. Battistella, M. Lerousseau, A. Carre, G. Klausner, R. Sun, C. Robert, S. Mougiakakou, N. Paragios, and E. Deutsch, “U-resnet : Ultimate coupling of registration and segmentation with deep nets,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019*, ser. Lecture Notes in Computer Science, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds., 2019, vol. 11766, pp. 310–319.
- [85] K. Hara, R. Vemulapalli, and R. Chellappa, “Designing deep convolutional neural networks for continuous object orientation estimation,” *arXiv preprint arXiv :1702.01499*, 2017.
- [86] T. Horstmann, O. Zettinig, W. Wein, and R. Prevost, “Orientation estimation of abdominal ultrasound images with multi-hypotheses networks,” in *Medical Imaging with Deep Learning*, 2022.

- 
- [87] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, “On the continuity of rotation representations in neural networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5745–5753.
- [88] S. Mahendran, H. Ali, and R. Vidal, “3d pose regression using convolutional neural networks,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2174–2182.
- [89] D. Q. Huynh, “Metrics for 3d rotations : Comparison and analysis,” *Journal of Mathematical Imaging and Vision*, vol. 35, no. 2, pp. 155–164, 2009.
- [90] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in *International conference on machine learning*, 2013, pp. 1310–1318.
- [91] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [92] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in *Proceedings of the 28th International Conference Neural Information Processing Systems - Volume 2*, ser. NIPS’15, 2015, p. 2017–2025.
- [93] G. Fiard, S.-Y. Selmi, E. Promayon, J.-L. Descotes, and J. Troccaz, “Simulation-based training for prostate biopsies : Towards the validation of the biopsym simulator.” *Minim Invasive Ther Allied Technol*, vol. 29, no. 6, pp. 359–365, dec 2020.
- [94] J.-B. Lefrancq, J.-L. Descotes, S.-Y. Selmi, T. Dupuy, E. Promayon, M. Chabenes, J.-A. Long, J. Troccaz, and G. Fiard, “External validation of a prostate biopsy simulator,” *Progrès en Urologie*, 2021.
- [95] G. Fiard, S.-Y. Selmi, M. Maignon, A. Bellier, E. Promayon, J.-L. Descotes, and J. Troccaz, “Validating the transfer of skills acquired on a prostate biopsy simulator : A prospective, randomized, controlled study,” *Journal of Surgical Education*, vol. 77, no. 4, pp. 953–960, 2020.
- [96] M. Kok, J. D. Hol, and T. B. Schön, “Using inertial sensors for position and orientation estimation,” *Found. Trends® Signal Process.*, vol. 11, no. 1-2, pp. 1–153, 2017.
- [97] G. Farneäck, “Two-frame motion estimation based on polynomial expansion,” in *Image Analysis*, J. Bigun and T. Gustavsson, Eds., 2003, vol. 2749, pp. 363–370.
- [98] D. Chicco, M. J. o. Warrens, and G. Jurman, “The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation,” *PeerJ Computing Sci.*, vol. 7, p. e623, jul 2021.
- [99] A. Tato and R. Nkambou, “Improving adam optimizer,” 2018.
- [100] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “Tensorflow : A system for large-scale machine learning,” vol. 16, pp. 265–283, 2016.
- [101] D. T. Huff, A. J. Weisman, and R. Jeraj, “Interpretation and visualization techniques for deep learning models in medical imaging,” *Physics in medicine and Biology*, vol. 66, pp. 04TR01 – 04TR01, 2020.

- 
- [102] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam : Visual explanations from deep networks via gradient-based localization.”
- [103] T. Dupuy, C. Beitone, J. Troccaz, and S. Voros, “2d/3d deep registration for real-time prostate biopsy navigation,” in *SPIE Medical Imaging 2021*, ser. Medical Imaging 2021 : Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 11598, feb 2021, p. 115981P.
- [104] T. Dupuy, C. Beitone, J. Troccaz, and S. Voros, “2d/3d deep registration along trajectories with spatiotemporal context : Application to prostate biopsy navigation,” *IEEE Transactions on Biomedical Engineering*, pp. 1–12, 2023.
- [105] O. Ronneberger, P. Fischer, and T. Brox, “U-net : Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., 2015, pp. 234–241.
- [106] S. Jadon, “A survey of loss functions for semantic segmentation,” in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, 2020, pp. 1–7.
- [107] V. Yeghiazaryan and I. Voiculescu, “Family of boundary overlap metrics for the evaluation of medical image segmentation,” *Journal of Medical Imaging*, vol. 5, no. 1, pp. 015 006–015 006, 2018.
- [108] A. A. Taha and A. Hanbury, “Metrics for evaluating 3d medical image segmentation : analysis, selection, and tool,” *BMC Medical imaging*, vol. 15, no. 1, pp. 1–28, 2015.
- [109] T. Gong, T. Lee, C. Stephenson, V. Renduchintala, S. Padhy, A. Ndirango, G. Keskin, and O. H. Elibol, “A comparison of loss weighting strategies for multi task learning in deep neural networks,” *IEEE Access*, vol. 7, pp. 141 627–141 632, 2019.
- [110] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić, and C. Schmid, “Vivit : A video vision transformer,” in *Proceedings of the IEEE/CVF International Conference on computer vision*, 2021, pp. 6836–6846.
- [111] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words : Transformers for image recognition at scale,” in *International Conference on Learning Representations*, 2021.
- [112] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet : Deep learning on point sets for 3d classification and segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.



## Navigation en temps-réel de biopsie prostatique assistée par intelligence artificielle

**Mot clés :** Recalage d'images 2D/3D, apprentissage profond, biopsie prostatique, guidage ultrasonore

Le cancer de la prostate est un problème de santé publique majeur. Seul un examen de biopsie prostatique permet d'établir définitivement le diagnostic de ce cancer. Durant cet examen, la navigation ainsi que la précision du ciblage des biopsies sont rendues complexes par la modalité échographique (2D US) couramment utilisée pour le guidage, et par de potentiels bougés ou déformations de la prostate. Ces éléments rendent le repérage et le ciblage des biopsies difficiles et imprécis, pouvant compromettre le diagnostic et les décisions thérapeutiques. Dans ce contexte, le principal objectif de cette thèse est de concevoir et développer un dispositif d'assistance à la navigation permettant à l'urologue de se repérer, en continu et avec précision, pour ainsi mieux visualiser la position de la biopsie qu'il cible. Pour cela, nous avons proposé des méthodes de recalage rigide 2D/3D en temps-réel, permettant d'aligner au mieux la coupe 2D US courante (guidage) par rapport à un volume de ré-

férence US (pris en début de navigation). Les principales originalités de nos travaux sont de considérer des informations liées à la trajectoire réalisée et de développer des approches fondées sur l'intelligence artificielle (par apprentissage profond), tout en respectant les conditions de réalisation clinique. En ajoutant des informations liées à la trajectoire réalisée, en considérant notamment les résultats du recalage précédent, le déplacement relatif de la sonde et un apprentissage multiple avec une tâche de segmentation, nous avons observé une amélioration significative de la qualité du recalage. Ces résultats ont été obtenus sur une base de données cliniques conséquente et évalués grâce à un protocole séquentiel pertinent et fidèle au déroulement clinique de notre application. L'identification de nouveaux types d'informations de trajectoire ainsi que de nouvelles manières de les intégrer à une architecture profonde restent des perspectives de cette thèse.

---

### AI Assistance for Real-Time Prostate Biopsy Navigation

**Keywords :** 2D/3D Medical image registration, Deep learning, Prostate biopsy, Ultrasound-guided procedure

Prostate cancer is a major public health problem. Only a prostate biopsy examination can confirm its diagnosis. During this examination, the accuracy of biopsy targeting is complicated by the ultrasound (2D US) modality commonly used for guidance, and by potential movements or deformations of the prostate. These elements make biopsy location and targeting difficult and imprecise, which can compromise both diagnosis and therapeutic decisions. In this context, the main objective of this thesis is to design and develop a navigation assistance device allowing the urologist to have a continuous and precise tracking of the biopsy location he/she is targeting, considering the prostate deformation. For this purpose, we proposed 2D/3D real-time rigid registration methods, allowing to localize the current intraoperative 2D

US image (guidance) with respect to a US reference volume (acquired at the beginning of navigation). The main contributions of our work are to consider prior trajectory information using artificial intelligence approaches (by deep learning), and by respecting the clinical conditions and realism. By adding prior trajectory information, considering in particular the previous registration results, the relative probe tracking and a segmentation-based multitask learning, we observed a significant improvement of the registration quality. These results were obtained on a large clinical database and evaluated through a cumulative protocol, relevant and adapted to the clinical application. The identification of new types of trajectory information as well as new ways to integrate them into a deep architecture remain perspectives.