



**HAL**  
open science

# Fusion de segmentations complémentaires d'images médicales par Intelligence Artificielle et autres méthodes de gestion de conflits

Lisa Corbat

► **To cite this version:**

Lisa Corbat. Fusion de segmentations complémentaires d'images médicales par Intelligence Artificielle et autres méthodes de gestion de conflits. Imagerie médicale. Université Bourgogne Franche-Comté, 2020. Français. NNT : 2020UBFCD029 . tel-03022231

**HAL Id: tel-03022231**

**<https://theses.hal.science/tel-03022231>**

Submitted on 24 Nov 2020

**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 DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ**

**PRÉPARÉE À L'UNIVERSITÉ DE FRANCHE-COMTÉ**

École doctorale n°37

SPIM : Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Informatique

par

**LISA CORBAT**

**Fusion de segmentations complémentaires d'images médicales par  
Intelligence Artificielle et autres méthodes de gestion de conflits**

Unité de Recherche :  
Femto-st/DISC

Thèse présentée et soutenue à Besançon, le 13 octobre 2020

Composition du Jury :

FRÉDÉRIC AUBER	Prof., UBFC Centre Hospitalier Régional Universitaire de Besançon (CHRUB)	Président
JEAN-PHILIPPE THIRAN	Prof., École Polytechnique Fédérale de Lausanne (EPFL)	Rapporteur
DANIEL RACOCÉANU	Prof., Univ. Sorbonne de Paris	Rapporteur
HENRI SOUCHAY	Dr., Clinical Research Manager for France, Global Research Organization, GE Healthcare	Examineur
JEAN-CHRISTOPHE LAPAYRE	Prof., Univ. de Bourgogne/Franche-Comté	Directeur
JULIEN HENRIET	MCF, Univ. de Bourgogne/Franche-Comté	Co-Directeur



**Résumé :**

Le néphroblastome est la tumeur du rein la plus fréquente chez l'enfant et son diagnostic est exclusivement basé sur l'imagerie. Ce travail qui fait l'objet de nos recherches s'inscrit dans le cadre d'un projet de plus grande envergure : le projet européen SAIAD (Segmentation Automatique de reins tumoraux chez l'enfant par Intelligence Artificielle Distribuée). L'objectif du projet est de parvenir à concevoir une plate-forme capable de réaliser différentes segmentations automatiques sur les images sources à partir de méthodes d'Intelligence Artificielle (IA), et ainsi obtenir une reconstruction fidèle en trois dimensions. Dans ce sens, des travaux réalisés dans une précédente thèse de l'équipe de recherche ont mené à la création d'une plate-forme de segmentation. Elle permet la segmentation de plusieurs structures individuellement, par des méthodes de type Deep Learning, et plus particulièrement les réseaux de neurones convolutifs (CNNs), ainsi que le Raisonnement à Partir de Cas (RàPC). Cependant,

il est ensuite nécessaire de fusionner de manière automatique les segmentations de ces différentes structures afin d'obtenir une segmentation complète pertinente. Lors de l'agrégation de ces structures, des pixels contradictoires peuvent apparaître. Ces conflits peuvent être résolus par diverses méthodes basées ou non sur l'IA et font l'objet de nos recherches. Nous proposons tout d'abord une première approche de fusion non focalisée sur l'IA en utilisant la combinaison de six méthodes différentes, basées sur différents critères présents sur l'imagerie et les segmentations. En parallèle, deux autres méthodes de fusion sont proposées en utilisant, un CNN couplé au RàPC pour l'une, et un CNN utilisant une méthode d'apprentissage spécifique existante en segmentation pour l'autre. Ces différentes approches ont été testées sur un ensemble de 14 patients atteints de néphroblastome et démontrent leur efficacité dans la résolution des pixels conflictuels et leurs capacités à améliorer les segmentations résultantes.

**Mots-clés :** Néphroblastome, Imagerie médicale, Fusion de segmentations, Gestion de conflits, Deep Learning

**Abstract:**

Nephroblastoma is the most common kidney tumour in children and its diagnosis is based exclusively on imaging. This work, representing the focus of our research, is part of a larger project: the European project SAIAD (Automated Segmentation of Medical Images Using Distributed Artificial Intelligence). The aim of the project is to design a platform capable of performing different automatic segmentations from source images using Artificial Intelligence (AI) methods, and thus obtain a faithful three-dimensional reconstruction. In this sense, work carried out in a previous thesis of the research team led to the creation of a segmentation platform. It allows the segmentation of several structures individually, by methods such as Deep Learning, and more particularly Convolutional Neural Networks (CNNs), as well as Case Based Reasoning (CBR).

However, it is then necessary to automatically fuse the segmentations of these different structures in order to obtain a complete relevant segmentation. When aggregating these structures, contradictory pixels may appear. These conflicts can be resolved by various methods based or not on AI and are the subject of our research. First, we propose a fusion approach not focused on AI using the combination of six different methods, based on different imaging and segmentation criteria. In parallel, two other fusion methods are proposed using, a CNN coupled to the CBR for one, and a CNN using a specific existing segmentation learning method for the other. These different approaches were tested on a set of 14 nephroblastoma patients and demonstrated their effectiveness in resolving conflicting pixels and their ability to improve the resulting segmentations.

**Keywords:** Nephroblastoma, Medical imaging, Segmentations fusion, Conflict management, Deep Learning



# REMERCIEMENTS

Je tiens à remercier l'ensemble des membres du jury pour avoir accepté d'y faire partie et d'évaluer mon travail. Je remercie également Jean-Philippe THIRAN et Daniel RACOCÉANU d'avoir accepté d'être les rapporteurs de cette thèse.

Mes remerciements vont ensuite à Julien HENRIET et Jean-Christophe LAPAYRE, mes co-directeur et directeur de thèse, pour m'avoir aidée et guidée tout le long de ces trois années. Ce fut un réel plaisir de travailler avec vous, et mes félicitations pour avoir su supporter mon caractère parfois compliqué à vivre.

Je remercie le programme européen Interreg ainsi que le projet SAIAD pour avoir financé mes travaux. Je tiens également à remercier tous les membres du projet, pour leur collaboration et leur motivation, ce qui nous a permis de mener à bien ce projet. Je tiens en particulier à remercier toute l'équipe du CHU de Besançon pour leur expertise médicale, sans quoi ce travail n'aurait pu aboutir. En particulier, merci à Yann CHAUSSY, Lorédane VIEILLE et Élise LACROIX d'avoir toujours été présents et répondu à mes nombreuses interrogations.

Ensuite, je tiens à remercier l'ensemble de l'équipe du DISC pour leur sympathie et leur accueil au sein du laboratoire. Merci à mes amis et collègues doctorants pour la bonne ambiance, la découverte de multiples restaurants, les pauses croissants du mercredi, et les soirées jeux de société. Enfin, merci à toi Florent, mon ami et collègue de bureau, pour tous ces bons moments. Nous nous sommes aidés et soutenus mutuellement durant nos thèses, et j'aurai toujours en mémoire nos débuts où nous apprenions ensemble les bases des réseaux de neurones à coup de conversations très animées.

Merci à Guillaume, mon compagnon, d'avoir toujours été à mes côtés et de m'avoir épaulée dans toutes les situations. Nos trois petits chats, Niva, Nelly et Plume, ont également été d'un soutien et d'un réconfort sans faille (tant que de la nourriture était en jeu). Je remercie également Isabelle, sa maman au grand cœur, sur qui je sais que je peux toujours compter.

Enfin, mes remerciements ne seraient pas complets sans parler de mes parents. Maman, Papa, vous avez toujours été présents pour moi, et m'avez toujours soutenue et poussée à aller encore et toujours plus loin. Pour tout ça, MERCI.



# SOMMAIRE

<b>Introduction</b>	<b>7</b>
Introduction générale . . . . .	7
Plan du mémoire . . . . .	8
Quelques indications pour la lecture de ce rapport . . . . .	10
<b>I Contexte et état de l'art</b>	<b>11</b>
<b>1 Contexte de ces travaux de recherche</b>	<b>15</b>
1.1 Le contexte du projet SAIAD . . . . .	15
1.1.1 Le néphroblastome . . . . .	15
1.1.2 Les reins . . . . .	16
1.1.3 Les objectifs du projet . . . . .	17
1.1.4 Le financement et les partenaires . . . . .	18
1.2 L'acquisition des données . . . . .	18
1.2.1 Les images numériques . . . . .	18
1.2.2 Les images scanner . . . . .	19
1.2.3 Deux temps d'acquisition : les temps vasculaire et tardif . . . . .	20
1.3 La segmentation et la reconstruction 3D . . . . .	21
1.3.1 La définition d'une segmentation d'image . . . . .	21
1.3.2 Les méthodes de segmentation existantes . . . . .	22
1.3.3 La construction et la représentation 3D . . . . .	24
1.4 La plate-forme existante COLISEUM-3D . . . . .	25
1.4.1 La couche de données . . . . .	25
1.4.2 La segmentation des différentes structures . . . . .	26
1.4.3 Le recalage temps vasculaire / temps tardif . . . . .	27
1.4.4 La fusion des différentes segmentations . . . . .	28
1.5 Les métriques d'évaluation de segmentations . . . . .	28
1.5.1 Les métriques utilisant la matrice de confusion . . . . .	28
1.5.2 La métrique Vol . . . . .	32

1.5.3	La métrique GCE . . . . .	33
	Synthèse . . . . .	35
<b>2</b>	<b>Les méthodes classiques de fusion de segmentations</b>	<b>37</b>
2.1	Les méthodes dites intuitives . . . . .	37
2.1.1	Le vote . . . . .	38
2.1.2	L'intersection et l'union . . . . .	39
2.2	Les méthodes de fusion dans le cadre de la segmentation Multi-Atlas . . .	40
2.2.1	La méthode SBA . . . . .	40
2.2.2	La méthode STAPLE . . . . .	41
2.3	Les méthodes basées sur une ou plusieurs métriques . . . . .	42
2.3.1	La méthode basée sur F-Mesure . . . . .	43
2.3.2	La méthode basée sur Vol . . . . .	46
2.3.3	La méthode basée sur PRI . . . . .	47
2.3.4	La méthode basée sur GCE . . . . .	48
2.4	Les autres méthodes basées sur les distances et les similarités . . . . .	49
2.4.1	La méthode <i>Neighbours and Intensity</i> : Nandl . . . . .	49
2.4.2	La méthode basée sur K-moyenne . . . . .	51
2.5	Les méthodes multicritères . . . . .	51
2.5.1	La méthode EFA-BMFM . . . . .	52
2.5.2	La méthode GCE+F-Mesure . . . . .	53
	Synthèse . . . . .	54
<b>3</b>	<b>La fusion de segmentations par réseaux de neurones</b>	<b>55</b>
3.1	Le principe des réseaux de neurones . . . . .	56
3.1.1	Les notions générales . . . . .	56
3.1.2	Les réseaux de neurones convolutifs . . . . .	60
3.1.3	Les CNNs pour la segmentation . . . . .	62
3.2	Différentes applications des réseaux de neurones à la fusion de segmen- tations . . . . .	64
3.2.1	La fusion par une couche de convolution . . . . .	65
3.2.2	La fusion par trois couches de convolution . . . . .	66
3.2.3	La fusion par FCM . . . . .	66
3.2.4	La fusion par un réseau plus profond . . . . .	67
	Synthèse . . . . .	70

<b>II Contribution</b>	<b>71</b>
<b>4 Agrégation de différentes méthodes classiques pour la fusion de segmentations et la gestion de conflits</b>	<b>75</b>
4.1 Les différentes méthodes de gestion de conflits utilisées et adaptées . . . .	75
4.1.1 Notre nouvelle méthode de gestion de conflits nommée ADS2 . . . .	76
4.1.2 L'adaptation de la méthode basée sur Vol et la nouvelle méthode basée sur Dice . . . . .	77
4.1.3 Nouvelle méthode 3D-Nandl issue de Nandl . . . . .	79
4.1.4 L'adaptation des champs aléatoires conditionnels pour la gestion de conflits . . . . .	80
4.2 Les modèles de gestion de conflits implémentés . . . . .	82
4.2.1 Le fonctionnement des modèles . . . . .	82
4.2.2 La fusion des résultats de la gestion de conflits . . . . .	84
4.3 Les résultats sur nos différentes méthodes de gestion de conflits . . . . .	84
4.3.1 L'acquisition des données . . . . .	84
4.3.2 Le calcul des paramètres optimaux pour la gestion des conflits . . . .	89
4.3.3 Les résultats sur les deux modèles . . . . .	89
4.4 Discussions . . . . .	96
4.4.1 L'efficacité et la comparaison des différentes méthodes utilisées . .	96
4.4.2 Le meilleur modèle de gestion des conflits à utiliser . . . . .	97
4.4.3 Les différents degrés de difficulté de résolution de conflits . . . . .	97
Synthèse . . . . .	99
<b>5 Méthodes de fusion de segmentations basées sur l'Intelligence Artificielle</b>	<b>101</b>
5.1 L'architecture du réseau de neurones créée pour la fusion . . . . .	101
5.2 Le système Deep Learning couplé au RàPC pour la fusion de segmentations	102
5.2.1 La modélisation d'un cas . . . . .	104
5.2.2 La phase de remémoration . . . . .	105
5.2.3 Le pré-traitement et la phase d'entraînement du réseau . . . . .	106
5.3 La méthode OVASSION pour la fusion de segmentations . . . . .	107
5.3.1 Le principe de la nouvelle méthode d'entraînement . . . . .	107
5.3.2 Formalisation . . . . .	108
5.4 Les résultats . . . . .	110
5.4.1 Les paramètres d'entraînement . . . . .	110
5.4.2 L'évaluation des différentes méthodes . . . . .	111
5.4.3 La représentation visuelle des résultats . . . . .	115

5.5	Discussions . . . . .	118
5.5.1	Les performances des différents réseaux de neurones . . . . .	118
5.5.2	La différence de performances entre les deux structures anatomiques	118
5.5.3	Discussion sur les performances de la méthode DL-RàPC . . . . .	119
5.5.4	Le temps de calcul . . . . .	120
5.5.5	Comparaison des performances des méthodes utilisant ou non l'Intelligence Artificielle . . . . .	120
	Synthèse . . . . .	122
	<b>Conclusion et perspectives</b>	<b>123</b>
	Conclusion . . . . .	123
	Perspectives . . . . .	124
	<b>Ma bibliographie personnelle</b>	<b>129</b>
	<b>Bibliographie</b>	<b>138</b>
	<b>Nomenclature</b>	<b>139</b>

# INTRODUCTION

## INTRODUCTION GÉNÉRALE

### L'ÈRE DU TRAITEMENT D'IMAGES

Les données numériques de types images et vidéos sont omniprésentes et de plus en plus massives, dans bien des domaines : télédétection, imagerie médicale ou encore robotique. Le traitement d'images est une discipline large, permettant de traiter des images et ayant différents intérêts, comme l'amélioration de leur qualité ou la possibilité d'extraire de multiples informations. Le traitement d'images est étroitement lié à la vision par ordinateur, qui consiste en la compréhension et l'interprétation d'une scène à partir d'une ou plusieurs images. La vision par ordinateur cherche alors, par le biais d'une machine, à imiter le système visuel humain, notamment grâce à différents outils de traitement d'images. L'intelligence artificielle joue également un grand rôle dans la vision par ordinateur et le traitement d'images complexes afin, par exemple, d'y détecter et d'y reconnaître des éléments précis. Depuis quelques années, il est possible d'interpréter l'information issue de données numériques, grâce notamment aux réseaux de neurones qui imitent le principe général de fonctionnement des neurones du cerveau humain.

### CONTEXTE GÉNÉRAL

Le néphroblastome est la tumeur rénale la plus fréquente chez l'enfant. Cette tumeur déforme et pousse le rein pathologique, et peut potentiellement envahir les voies urinaires. La réalisation de segmentations de l'ensemble de ces structures permet aux personnels médicaux d'obtenir une vision plus précise de la situation, et ainsi de pouvoir planifier au mieux la suite de la prise en charge du jeune patient. En effet, si la tumeur n'est pas trop envahissante et qu'elle ne recouvre pas les voies urinaires, il est alors possible de réaliser une néphrectomie partielle du rein touché, également appelée chirurgie conservatrice. Dans le cas contraire, si l'envahissement est trop important, alors une néphrectomie totale est recommandée, car le risque de récurrence est alors trop élevé. Ainsi, les segmentations permettent, en plus de déterminer le type d'intervention chirurgicale à réaliser, d'aider les chirurgiens lors de la planification de l'intervention (les outils à utiliser, l'anticipation d'éventuelles difficultés, ...). Cependant, ces segmentations doivent actuellement être réalisées manuellement ou à l'aide d'outils guidés par un personnel médical. Dans tous les cas, une forte intervention humaine est nécessaire. La réalisation de ces segmentations est alors très chronophage (entre 6 et 9 heures en moyenne) et doit être réalisée par un praticien qualifié : chirurgien ou radiologue. C'est pourquoi, ces segmentations ne sont quasiment jamais réalisées, obligeant ainsi les experts médicaux à analyser et à interpréter directement les dizaines à centaines de coupes 2D en noir et blanc réalisées par le scanner.

Les travaux de cette thèse s'inscrivent dans le cadre du projet Européen **SAIAD**, signifiant **S**egmentation **A**utomatique de reins tumoraux chez l'enfant par **I**ntelligence **A**rtificielle **D**istribuée. Ce projet a pour objectif de fournir une représentation fidèle du néphroblastome et des structures avoisinantes, à l'aide de segmentations calculées par des techniques d'Intelligence Artificielle (IA). La représentation de l'abdomen de chaque patient peut alors être fournie, sans intervention humaine ou a minima avec une intervention limitée, ce qui permet de ne pas faire perdre un temps précieux à un ou plusieurs personnels de santé, et de mieux évaluer l'opportunité de réaliser une néphrectomie partielle lorsque cela est possible.

Le néphroblastome et le rein pathologique ont déjà pu être segmentés via différentes techniques de segmentations par IA, notamment par apprentissage profond (**Deep Learning**) et par **Raisonnement à Partir de Cas** (RàPC), dans le cadre d'une précédente thèse. Ces techniques de segmentations sont regroupées au sein d'une plate-forme de segmentation appelée COLISEUM-3D (**COL**aborative platform with artificial **I**ntelligence for **SE**gmentation of **tU**moral kidney in **M**edical images in **3D**). Cette plate-forme permet de segmenter individuellement chaque structure, par des techniques conçues et optimisées pour chacune. Elle prévoit également, à la suite du système de segmentations, l'agrégation de ces structures, afin d'obtenir la reconstruction totale du patient. Cette thèse, a donc pour objectif de définir les meilleurs moyens de fusion de ces différentes segmentations complémentaires.

## OBJECTIF DE NOS TRAVAUX

Après la segmentation des différentes structures indépendamment, il est alors légitime de se demander comment les fusionner, afin d'obtenir des segmentations consensus contenant l'ensemble des structures. En théorie, comme chaque segmentation est composée d'une seule structure indépendante des autres, la fusion de ces segmentations complémentaires ne devrait pas poser de problème. Or, ces segmentations calculées par IA ne sont pas parfaites, et donc certaines structures peuvent se chevaucher. Certains pixels appartiennent alors à plusieurs structures dans les différentes segmentations. On parle dans ce cas de pixels en conflit. Mais comment résoudre ces conflits lors de la fusion ? Comment déterminer à quelle structure ces pixels doivent appartenir ?

Les travaux de recherche présentés ici s'inscrivent donc dans la suite logique des précédents travaux accomplis dans le cadre du projet SAIAD. Ils se situent également au niveau de la dernière couche de la plate-forme COLISEUM-3D, afin d'effectuer la segmentation par IA de différentes structures de patients atteints de néphroblastome. Cette fusion de segmentations des différentes structures est alors la dernière étape avant l'obtention d'une visualisation plus précise de l'abdomen du patient, et ainsi aider au mieux le personnel médical dans la planification de l'intervention chirurgicale du jeune patient.

## PLAN DU MÉMOIRE

La première partie de ce manuscrit est composée de 3 chapitres et est consacrée à la définition du contexte et à notre état de l'art. Dans le premier chapitre, nous abordons plus en détails le contexte dans lequel s'est déroulée cette thèse. Nous détaillons la problématique médicale du projet SAIAD pour les patients atteints de

néphroblastome. Nous expliquons les modes d'acquisition des différentes données numériques, ainsi que des notions succinctes d'imagerie. Nous abordons le principe de segmentation d'images et de leur représentation 3D, ainsi qu'un rapide aperçu des différentes techniques de segmentation existantes. Puis, nous introduisons la plateforme COLISEUM-3D existante, réalisée dans le cadre du projet SAIAD, composée de différentes couches, dont une permettant la segmentation de différentes structures abdominales et une autre permettant la fusion de leur représentations. Enfin, ce chapitre se termine par la présentation de différentes métriques d'évaluation de segmentations présentes dans la littérature : métriques très importantes pour s'assurer de la justesse de ces représentations numériques, mais également utiles dans l'agrégation de ces différentes segmentations.

Le second chapitre est consacré aux différentes méthodes existantes issues de la littérature, non basées sur l'IA, pour l'agrégation de multiples représentations. Nous abordons notamment les méthodes intuitives comme le vote majoritaire, les méthodes de fusion utilisées dans le cadre de techniques de segmentation Multi-Atlas, en passant par les méthodes se basant sur des métriques d'évaluation aussi bien que celles basées sur d'autres informations présentes dans les images, comme les distances entre pixels et les similarités de niveaux de gris.

Le troisième chapitre est dédié aux approches de fusion par IA et notamment par réseau de neurones. Dans un premier temps, nous présentons les notions essentielles à la bonne compréhension des réseaux de neurones convolutifs (CNN) pour la segmentation, en passant du perceptron à l'apprentissage profond (*Deep Learning*), ainsi qu'aux différentes couches utiles dans ces réseaux pour le traitement d'images. Dans un second temps, nous abordons les réseaux utilisés dans la littérature pour la fusion de segmentations, avec principalement les réseaux utilisés dans le cadre de la détection de saillances, domaine dans lesquelles l'agrégation d'images est largement utilisée.

La deuxième partie, composée de 2 chapitres, est consacrée à nos différentes contributions. Le premier chapitre de cette partie aborde notre première contribution basée sur des techniques de fusion dites *classiques*, ou non basées sur des concepts d'IA. Deux modèles de fusion et de gestion des pixels en conflit (conflits dûs à la fusion) sont proposés. Ces modèles utilisent un ensemble de méthodes issues de la littérature (et dans ce cas adaptées) ou créées, afin de résoudre ces conflits. Les résultats de ces méthodes sont ensuite agrégés dans un processus final de fusion permettant d'obtenir une segmentation consensus de ces différentes méthodes classiques. Les résultats des différentes méthodes et des différents modèles sont présentés et discutés. Ce chapitre intègre également une discussion sur la comparaison et l'efficacité de ces différentes méthodes, ainsi que du modèle le plus optimal.

Enfin, le dernier chapitre aborde nos deux contributions basées sur l'IA, avec l'utilisation de CNN ainsi que de RàPC. La première combine un CNN avec un RàPC permettant ainsi d'entraîner le réseau à l'aide de cas sélectionnés dans la base de cas du RàPC. L'entraînement est également ajusté en fonction des similarités des différents cas choisis. La seconde contribution par IA est également présentée : elle utilise un CNN avec une méthode d'entraînement particulière, nommée OV<sup>2</sup>ASSION, basée sur un sur-apprentissage par patient. Enfin, les résultats et les discussions de ces méthodes par IA sont également présentés avec une analyse de leur gestion des pixels en conflit ainsi que de l'amélioration de la qualité des segmentations.

## QUELQUES INDICATIONS POUR LA LECTURE DE CE RAPPORT

Le plan est organisé sur quatre niveaux : Parties, Chapitres, Sections et Sous-Sections. Il est utile de savoir que la Partie II, présentant les contributions et résultats, est indépendante de la première Partie, dédiée au contexte de la thèse et à l'état de l'art sur différentes approches.

Chaque partie possède une introduction présentant son sujet, tout comme chaque chapitre. Une synthèse est disponible à la fin de chaque chapitre afin de résumer son contenu.

Les résultats de nos travaux ont pour la plupart été publiés. La liste des publications personnelles est donnée avant la bibliographie placée en fin du document.



## CONTEXTE ET ÉTAT DE L'ART



Cette première partie du manuscrit de thèse présente le contexte général ainsi que l'état de l'art réalisé sur les différentes méthodes de fusion de segmentations existantes. Elle introduit également différentes notions médicales et d'Intelligence Artificielle (IA), indispensables à la bonne compréhension de mes travaux de thèse et des contributions présentées.

Le premier chapitre introduit le projet SAIAD (*Segmentation Automatique de reins tumoraux chez l'enfant par Intelligence Artificielle Distribuée*), qui a pour objectif de réaliser des représentations 3D automatiques, grâce aux segmentations, calculées par IA, de néphroblastome : qui est la tumeur rénale la plus fréquente chez l'enfant. Le principe de la segmentation et les techniques pour son évaluation sont abordés, ainsi que les différents avancements du projet dans la création d'une plate-forme et de plusieurs méthodes de segmentation.

Le second chapitre présente les différentes méthodes que nous appelons *classiques* pour la fusion de segmentations dans la mesure où elles ne sont pas basées sur des concepts d'IA. Ces méthodes sont classées en cinq catégories : les méthodes intuitives, les méthodes pour la segmentation Multi-Atlas, les méthodes par l'utilisation de métriques, les méthodes par distances et similarités, et enfin les méthodes combinées.

Pour clore cette partie, le troisième chapitre introduit les réseaux de neurones, en passant du principe général d'un neurone artificiel aux réseaux utilisés dans le cadre du traitement d'images, appelés réseaux de neurones convolutifs (CNNs). Puis, les différents CNNs utilisés particulièrement pour la fusion de segmentations dans la littérature sont présentés.



# CONTEXTE DE CES TRAVAUX DE RECHERCHE

Les travaux de cette thèse s'inscrivent dans le cadre du projet Européen **SAIAD** qui a pour objectif de proposer une plate-forme capable de segmenter automatiquement, c'est-à-dire sans intervention humaine, ou a minima avec une intervention limitée, le néphroblastome, qui est la tumeur rénale la plus fréquente chez l'enfant. Ces segmentations permettent de fournir aux chirurgiens et aux radiologues des représentations 3D de l'abdomen du patient, leur permettant ainsi d'obtenir de nombreuses informations utiles pour la planification de l'acte chirurgical qui sera la néphrectomie (l'ablation totale ou partielle du rein touché).

Ce projet ayant initialement débuté en 2017, plusieurs travaux ont déjà été menés parallèlement à cette thèse sur la recherche des meilleures techniques de segmentation par Intelligence Artificielle (IA). Ainsi, différentes structures de l'abdomen de patients atteints de néphroblastome ont pu être segmentées par **Deep Learning**, apprentissage profond, utilisant le principe des réseaux de neurones, et segmentées par **Case Based Reasoning**, Raisonnement à Partir de Cas (RàPC), couplé à des méthodes classiques de segmentation. Les segmentations de ces différentes structures doivent ensuite pouvoir être agrégées afin d'obtenir la segmentation consensus finale. C'est sur ce dernier point que s'inscrivent les travaux de cette thèse.

La section suivante introduit plus en détails le projet SAIAD, et les contours de nos travaux de thèse, et donne les connaissances de base sur la segmentation et les métriques d'évaluation, notions essentielles à la bonne compréhension du projet et des contributions de cette thèse.

## 1.1/ LE CONTEXTE DU PROJET SAIAD

### 1.1.1/ LE NÉPHROBLASTOME

Le néphroblastome ou tumeur de Wilms est une tumeur spécifique à l'enfant qui constitue 5 à 10% des tumeurs malignes chez l'enfant. Il s'agit de la tumeur rénale la plus fréquente chez l'enfant et elle survient généralement entre 1 et 5 ans [Plantaz15, IncaNephro07]. Cependant, il s'agit d'un cancer rare dont une centaine de cas est diagnostiquée chaque année en France. Aucune cause particulière n'a, pour l'heure, été prouvée dans l'apparition du néphroblastome. La tumeur est rapidement évolutive, et son traitement, vi-

sant à guérir totalement l'enfant, est pluridisciplinaire, en associant la chirurgie, la chimiothérapie et la radiothérapie, ce qui permet un taux de guérison de 90% pour les formes communes [Kaste2008]. Le principe du traitement est le suivant : une chimiothérapie préopératoire peut être effectuée afin de réduire la taille de la tumeur et faciliter l'intervention. Une ablation du rein touché est alors réalisée, que l'on appelle également néphrectomie qui peut être totale ou partielle. Enfin, une chimiothérapie ou une radiothérapie est préconisée afin de détruire d'éventuelles métastases et cellules résiduelles. De préférence, la chimiothérapie est préconisée face à la toxicité de la radiothérapie pour l'enfant.

### 1.1.2/ LES REINS

De couleur rouge brun, les reins sont des organes puissants jouant de nombreux rôles essentiels pour notre organisme. Par paire, ils sont situés de chaque côté de la colonne vertébrale sous la cage thoracique. Ils permettent la régulation de la quantité d'eau présente dans notre organisme, ainsi que la filtration du sang et l'élimination par les urines des déchets présents dans le sang. Les reins permettent également de réguler la pression du sang dans les artères grâce à la production d'hormones. La Figure 1.1 présente schématiquement le rein humain et sa vascularisation.

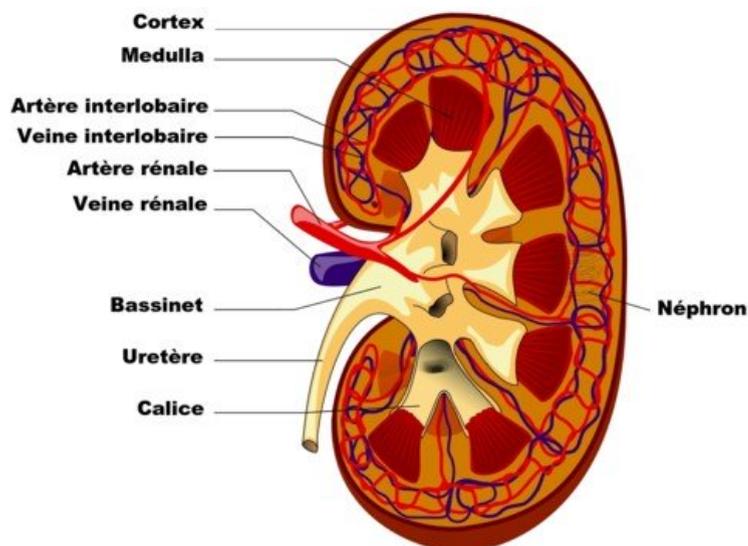


FIGURE 1.1 – Schéma du rein et de sa vascularisation (issue de *KidneyStructures* by P.M. Jaworski [SiteWebKidney2020]).

Le rein est formé de deux parties : le parenchyme rénal qui est composé du cortex et de la médulla (la partie centrale du rein), et les voies excrétrices. L'**artère rénale** permet au sang d'arriver jusqu'au rein. Les néphrons, d'environ un million, se situent dans le parenchyme rénal, et filtrent le sang qui ressort par la **veine rénale**. Les urines formées dans les **cavités excrétrices**, et plus particulièrement dans les calices puis dans le bassinet, sont évacuées par l'uretère, qui permet son transport vers la vessie. Par jour, les reins filtrent en moyenne 120 litres de sang et produisent 1 à 2 litres d'urine.

### 1.1.3/ LES OBJECTIFS DU PROJET

L'intervention chirurgicale du néphroblastome mène la plupart du temps à une néphrectomie totale du rein touché, consistant à enlever la tumeur ainsi que la totalité du rein. Cependant, compte tenu du très jeune âge des patients, l'idéal est de réaliser, lorsqu'il est possible, une néphrectomie partielle en épargnant ainsi la partie saine et fonctionnelle du rein. Ce type de chirurgie est dite conservatrice. Le choix de l'acte chirurgical se base sur les recommandations de la Société Internationale d'Oncologie Pédiatrique (la *SIO*) qui a établi un certain nombre de critères à vérifier tels que la taille de la tumeur ou encore sa localisation. L'envahissement tumoral des voies urinaires est également un critère important. En effet, si l'envahissement de la tumeur est limité au cortex rénal, alors une chirurgie conservatrice est envisageable. Au contraire, si les voies urinaires sont envahies, alors une néphrectomie totale est imposée pour éviter tous risques de récurrence. Jusqu'à très récemment, la volonté de sauvegarde du rein tumoral n'était envisagée qu'en cas de néphroblastome bilatérale. Cependant, la *SIO* recommande maintenant dès que possible la sauvegarde du rein touché, car il a été prouvé que le taux de guérison est proche de 90% en cas de chirurgie conservatrice respectant les critères décrits ci-dessus, et qu'en cas de néphrectomie totale se pose un risque prédominant de développer, un jour au cours du reste de vie du jeune patient, une insuffisance rénale.

Une représentation complète du néphroblastome et des structures abdominales voisines (comme les reins, les vaisseaux sanguins et les cavités rénales), le tout par une segmentation 3D réalisée à partir d'images scanner, permet d'offrir aux personnels médicaux toutes les informations dont ils ont besoin afin de planifier l'intervention : prendre une décision sur le type de néphrectomie à réaliser, anticiper sur ce que l'on va rencontrer pendant l'opération, sa durée, la quantité de matériel sanguin nécessaire. . . Mais cette extraction de données (segmentation, reconstruction 3D. . .) est actuellement réalisée manuellement par les radiologues ou les chirurgiens et prend en général entre 6 et 9 heures d'un temps précieux de spécialiste. Ce travail manuel utilise parfois l'aide de logiciels assez complexes comme ITK Snap [SiteWebITK2020] ou 3D-Slicer [SiteWebCFI2020] qui réalisent des segmentations à partir d'algorithmes simples comme le seuillage ou la croissance de région. Que les segmentations soient réalisées manuellement ou avec l'aide d'un logiciel, la tâche n'en reste pas moins longue et chronophage.

En général, ces segmentations ne sont donc pas effectuées car le temps d'un personnel médical (radiologue ou chirurgien) est précieux et, sans information sur l'envahissement tumoral, le chirurgien n'a principalement pas d'autres choix que de recourir à une néphrectomie totale du rein touché. L'objectif du projet est donc de permettre l'automatisation des segmentations de tumeurs rénales par des outils d'Intelligence Artificielle, le tout sans intervention humaine ou, a minima, avec très peu d'intervention humaine.

Actuellement, plusieurs verrous scientifiques se présentent. Les segmentations de tumeurs rénales peuvent être longues et difficiles à réaliser même par un expert du domaine car certaines structures sont difficilement reconnaissables (différenciables) sur les images scanner. Par exemple, les enfants ne possèdent que très peu de tissus adipeux, ce qui a pour conséquence que la tumeur peut se confondre avec les muscles, dont les intensités de gris sont similaires. Enfin, nous sommes confrontés à une grande variabilité des cas. Le projet a pour objectif de réaliser une visualisation de l'abdomen de tous les enfants atteints de néphroblastome, ce qui recouvre des enfants d'âges relativement différents. La taille des organes peut donc fortement varier d'un patient à un autre. La tumeur peut également être de n'importe quelle taille et forme et elle peut se situer à

n'importe quel endroit sur le rein (collée au rein, plus invasive ou encore recouvrante).

#### 1.1.4/ LE FINANCEMENT ET LES PARTENAIRES

**SAIAD** (Segmentation **A**utomatique de reins tumoraux chez l'enfant par Intelligence **A**rtificielle **D**istribuée) est un projet Interreg V Franco-Suisse d'une durée de 36 mois et d'un montant de 1 281 938,81 €, dont le soutien financier des Fonds Européen de Développement Régional (FEDER) est de 520 223,28 €, de fonds fédéraux Interreg Suisse de 132 900 € et du canton de Neuchâtel de 90 000 €.

Plusieurs partenaires ont activement participé au projet :

- Le CHRU de Besançon fournit les données anonymisées des patients dont les images scanner et les segmentations correspondantes réalisées manuellement,
- L'agence bisontine de l'entreprise MainCare Solution [SiteWebIDOin2020] a créé une plate-forme sécurisée de stockage et de visualisation des données patients,
- L'EPFL (École Polytechnique Fédérale de Lausanne) a été chargée de l'état de l'art des techniques de segmentations non réalisées par IA,
- L'entreprise CFI [SiteWebCFI2020] travaille sur la reconstruction en 3D des segmentations,
- Notre département informatique (DISC) de l'institut CNRS FEMTO-ST de l'Université de Franche-Comté est chargé de l'état de l'art et de l'implémentation des techniques de segmentations automatiques par IA.

Notons : Julien Henriet, codirecteur de cette thèse, est le chef de file scientifique de ce projet Interreg V et j'ai été durant cette thèse l'ingénieure en charge du projet.

## 1.2/ L'ACQUISITION DES DONNÉES

La première étape du projet a été, au fil des derniers mois, l'acquisition de données, c'est-à-dire l'acquisition des représentations abdominales de différents patients par les radiologues. Dans cette section, nous définissons les images numériques ainsi que les images numériques acquises par scanographie, appelées images scanner. Pour finir, nous définissons un peu plus en détails les types d'images obtenues dans le cadre du projet.

### 1.2.1/ LES IMAGES NUMÉRIQUES

Une image est un regroupement de pixels ayant chacun une valeur qui lui est propre représentant une couleur. Une image en noir est blanc est composée uniquement de nuances de gris. Le codage de la couleur est réalisé sur un octet et s'effectue sur une valeur comprise entre 0 et 255 : 0 étant la couleur noire et 255 la couleur blanche. Chaque pixel d'une image en couleurs est composé de 3 octets correspondant à l'espace colorimétrique RVB pour Rouge Vert et Bleu. Le mélange de ces trois valeurs donne une

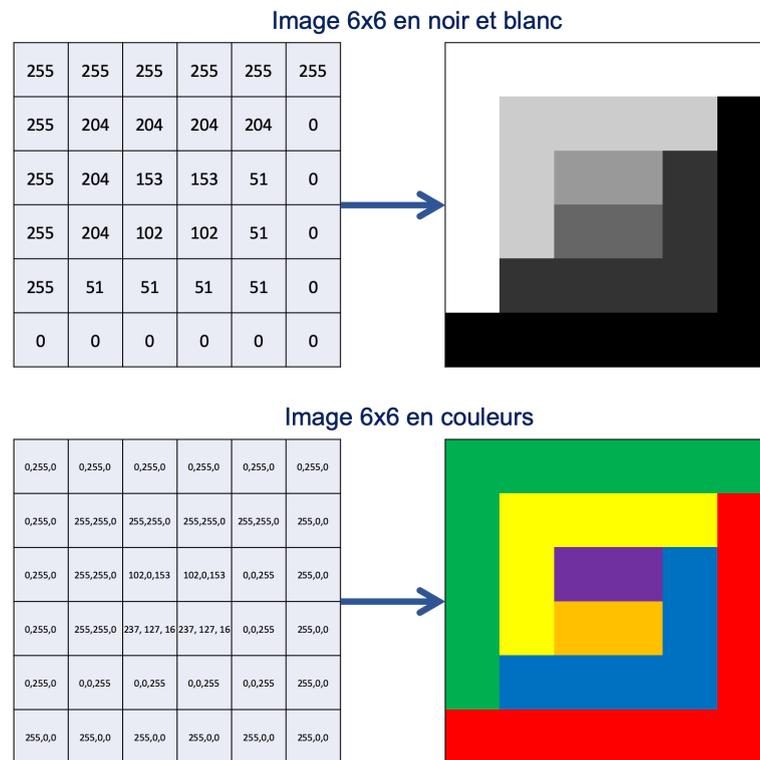


FIGURE 1.2 – Illustration de la représentation d'images de taille  $6 \times 6$  en noir et blanc et en couleurs.

couleur unique et il existe ainsi  $256 \times 256 \times 256$  couleurs possibles pour chaque pixel de l'image. La Figure 1.2 illustre la représentation d'une image en un ou plusieurs octets.

Une image en noir et blanc de taille  $H \times L$  (Le nombre de pixels en hauteur multiplié par le nombre de pixels en largeur) peut alors être représentée sous forme d'une matrice de même taille. Une image en couleurs sera quant à elle représentée par une matrice de taille  $3 \times H \times L$  afin d'ajouter pour chaque pixel les 3 valeurs RVB.

### 1.2.2/ LES IMAGES SCANNER

Une acquisition par scanner, également appelée tomographie ou CT-scan (pour *Computerized Tomography scan*) est un examen qui, grâce à un appareil d'imagerie médicale, utilise des rayons X afin de mesurer leur absorption par les tissus. Ceci permet d'obtenir la représentation des structures anatomiques, coupe par coupe, du patient. L'assemblage de ces coupes permet ainsi d'obtenir une visualisation de l'abdomen du patient en 3 dimensions. Ces représentations sont ensuite stockées sous format DICOM (pour *Digital Imaging and Communications in Medicine*). Ce format est une norme qui a pour but de standardiser les données issues des différentes modalités d'imagerie médicale (comme la radiographie (rayons-X), la tomographie (CT-Scan), les ultrasons, l'imagerie nucléaire, l'imagerie par résonance... ). Son but est notamment de faciliter la communication et le transfert des images entre les différentes modalités d'imagerie médicale et les différents hôpitaux. En plus de la représentation des structures anatomiques, un fichier DICOM comporte de nombreuses autres informations relatives au patient, à l'examen

ainsi qu'à l'acquisition des images.

### 1.2.3/ DEUX TEMPS D'ACQUISITION : LES TEMPS VASCULAIRE ET TARDIF

Les images scanner d'un patient atteint de néphroblastome doivent permettre de visualiser différentes structures abdominales : le rein sain, le rein pathologique, le néphroblastome, le système de circulation sanguine (l'artère rénale et la veine rénale) et enfin les cavités rénales. Mais cette dernière structure est difficile, voire impossible, à visualiser sur les images scanner de base. Il est nécessaire d'utiliser un produit de contraste, produit iodé hydrosoluble, qui permet d'améliorer le contraste des structures ingérant le produit : en particulier les cavités rénales.

Ainsi, l'examen doit être réalisé en deux temps, comme illustré par la Figure 1.3 : le **temps vasculaire** et le **temps tardif**. Dans un premier temps, lorsque le produit de contraste est administré au patient, les structures fortement vascularisées deviennent plus discernables avec des intensités de contrastes plus importantes. C'est le cas pour les reins, les cavités et la tumeur. Un premier jeu d'images scanner (en moyenne 96 coupes transversales successives par patient) est alors réalisé en temps vasculaire. Plusieurs minutes après l'injection, le produit est ensuite filtré par les reins et passe dans les voies urinaires. Ainsi, les cavités rénales apparaissent plus distinctement en blanc dans le deuxième jeu d'images scanner réalisé en temps tardif (à nouveau en moyenne par patient 96 coupes transversales successives).

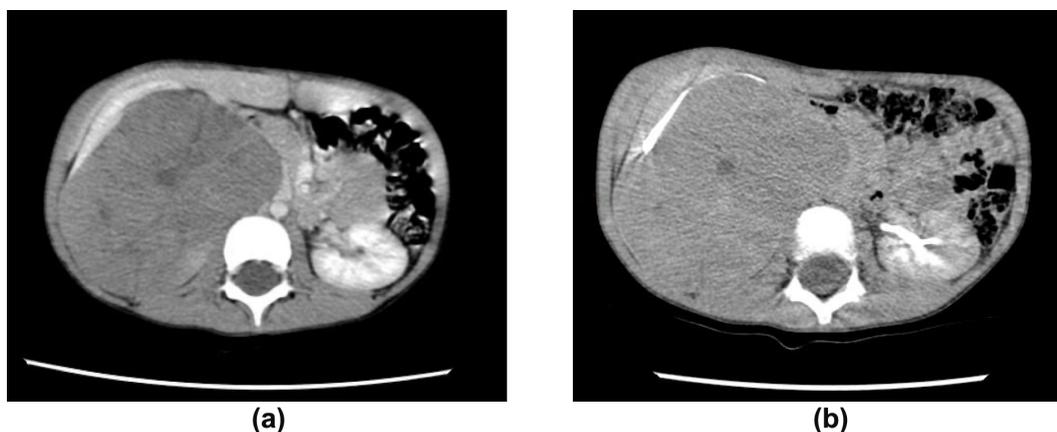


FIGURE 1.3 – Exemple d'une image scanner d'une coupe d'un patient atteint de néphroblastome (a) en temps vasculaire et (b) en temps tardif.

Comme plusieurs minutes se sont écoulées entre les deux jeux de données, le jeune patient, malgré les consignes données par le personnel soignant, peut bouger, entraînant ainsi un décalage important entre les images acquises au cours des deux temps. Afin d'obtenir une bonne correspondance entre ces deux temps d'images, il faut alors recourir à un recalage (cf. Section 1.4.3).

Il est à noter que nous avons, dans le projet SAIAD, seulement des acquisitions d'images scanner, et non des images de type IRM (Imagerie par Résonance Magnétique). En effet, l'accès à l'examen scanner est plus rapide, et la durée s'écoulant entre deux temps d'acquisition par IRM est beaucoup plus longue, rendant alors le fait de maintenir le jeune patient le plus immobile possible, plus complexe.

## 1.3/ LA SEGMENTATION ET LA RECONSTRUCTION 3D

### 1.3.1/ LA DÉFINITION D'UNE SEGMENTATION D'IMAGE

La segmentation d'une image permet de regrouper les pixels de l'image entre eux selon certains critères et caractéristiques. L'image est alors divisée en différentes **classes**, **régions**, également appelées segments, et une **étiquette** (ou en anglais *label*) sera attribuée à chaque pixel, de telle sorte que les pixels ayant les mêmes caractéristiques possèdent la même étiquette. Le processus de segmentation permet alors d'obtenir une image modifiée plus significative, ce qui permet d'extraire les informations caractéristiques contenues dans l'image.

Il existe différents types de segmentations. L'image peut être segmentée dans sa globalité, ou la segmentation peut ne cibler que certains éléments, voire seulement les contours de ces éléments. Il existe également une différence entre la segmentation sémantique et la segmentation d'instance :

- La segmentation sémantique segmente l'image en attribuant une classe à chaque type d'élément présent dans l'image. Nous avons un concept de catégorie.
- La segmentation d'instance attribue une nouvelle étiquette à chaque élément rencontré dans l'image, peu importe s'il s'agit du même type d'élément. Nous avons ici un concept d'instance. Ce type de segmentation peut être intéressant dans le but de compter le nombre total d'objets présents, sans distinction par catégorie.

Il existe également des méthodes de segmentation par **superpixels** permettant une *sur-segmentation* de l'image dont le but est de diviser l'image en un nombre assez conséquent de régions, et ainsi obtenir une grande quantité d'informations.

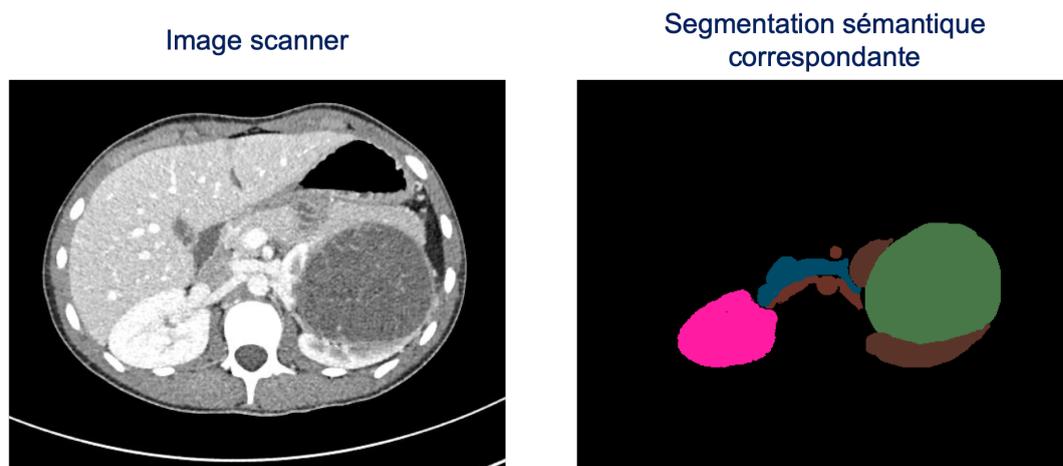


FIGURE 1.4 – Représentation de la segmentation sémantique du néphroblastome à partir d'une image scanner en temps vasculaire.

La Figure 1.4 illustre la segmentation sémantique du néphroblastome que nous voulons obtenir dans le cadre du projet SAIAD. L'étiquette rose est attribuée au rein sain, l'étiquette bleue à la veine, l'étiquette rouge à l'artère, l'étiquette marron au rein pathologique, l'étiquette verte à la tumeur et enfin l'étiquette noire à l'arrière-plan de l'image. Il en

est de même pour le temps tardif, dans lequel cette fois-ci seulement les cavités rénales et l'arrière-plan seront représentés dans la segmentation.

### 1.3.2/ LES MÉTHODES DE SEGMENTATION EXISTANTES

Il existe une grande variété de méthodes de segmentation, que ce soit par des procédés simples comme un simple seuillage ou par des techniques d'IA. Zaitoun et al., dans [Zaitoun2015], présentent en 2015 une classification de toutes les méthodes de segmentations existantes, et bien que les recherches sur l'utilisation des réseaux de neurones dans ce domaine étaient à leur début, ils sont également cités. Dans cet état de l'art, la plupart des méthodes sont classées en deux groupes :

- Les **méthodes basées sur les régions** : elles divisent l'image en plusieurs régions de sorte que les pixels d'une région possèdent au moins une caractéristique commune. C'est le cas pour l'algorithme des K-Moyenne (*K-Mean*), la méthode de Division et Fusion (*Split and Merge*), la croissance de régions (*Region Growing*) ou encore les algorithmes de seuillage (*Threshold*).
- Les **méthodes basées sur les contours** et plus particulièrement sur les similitudes : elles détectent les différents éléments dans l'image grâce à la différence de niveaux de gris obtenue au niveau de leurs contours. Sont alors mises en évidence des méthodes plus poussées comme les approches basées sur la logique floue [Soesanti2011], les algorithmes génétiques [Tosta2017] [Khan2015], et les réseaux de neurones.

Ces différentes approches sont également présentées dans les travaux de N. Senthilkumar et R. Rajesh [Senthilkumar2009].

#### 1.3.2.1/ LA SEGMENTATION PAR RÉSEAUX DE NEURONES

L'engouement pour l'utilisation des réseaux de neurones pour la segmentation, à partir de 2015, a permis à cet axe de recherche d'être l'un des plus explorés dans la littérature, avec un nombre considérable de nouveaux réseaux. Pour n'en citer que quelques uns, en 2015 a été produit le réseau FCN-8s pour la segmentation [Long2015] (réseau que nous développons dans la Section 3.1.3 et la Figure 3.8), puis DecoupledNet [Hong2015], U-Net [Ronneberger2015], DeepLab [Chen2017] et des réseaux avec le principe **encoder-decoder** comme le réseau DeconvNet [Noh2015] et SegNet [Badrinarayanan2017]. Le principe *encoder-decoder* est d'avoir :

- une première partie **encoder** du réseau permettant d'extraire les caractéristiques,
- et la seconde partie **decoder** pour suréchantillonner/agrandir les images obtenues de la première partie.

L'état de l'art réalisé par Zhao et al. [Zhao2017] présente les différents réseaux de neurones profonds de la littérature, dans le cadre d'objets difficilement différentiables, que ce soit pour la classification, la détection ou la segmentation d'images.

Les réseaux de neurones dans le contexte médical sont également nombreux, comme le présentent Litjens et al. dans [Litjens2017]. De nombreuses études ont été menées sur la segmentation de reins sains et même pathologiques ([Torres2018], [Thong2018]), mais la segmentation automatique du néphroblastome reste encore un défi. Le fonctionnement général d'un réseau de neurones et les différentes couches utilisées pour la segmentation sont développés dans la suite de ce document dans la Section 3.1.

### 1.3.2.2/ LA SEGMENTATION PAR RAISONNEMENT À PARTIR DE CAS

Le Raisonnement à Partir de Cas (RàPC ou encore *CBR* pour *Case-Based Reasoning*), est une méthode de résolution de problèmes basée sur l'analogie s'appuyant sur les expériences passées pour résoudre un nouveau problème. Il s'agit d'une méthode d'IA avec une approche orientée connaissance. En se remémorant des situations similaires au nouveau cas à résoudre, cette méthode peut construire une nouvelle solution en se basant sur les solutions des situations remémorées. C'est dans les années 1970 que le concept du RàPC est né, mais sa première implémentation est réalisée dix ans plus tard par Kolodner et al. [Kolodner1983], et Reisbeck et al. qui nomment pour la première fois ce principe comme *Case-Based Reasoning* [Reisbeck1989].

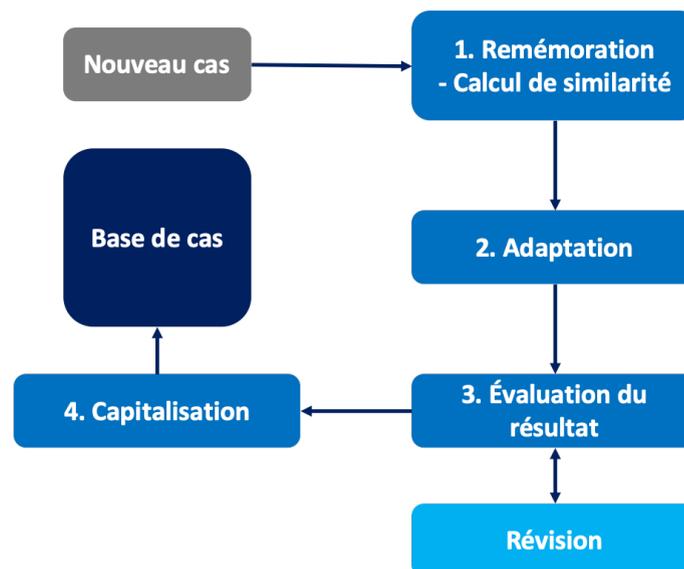


FIGURE 1.5 – Exemple d'un cycle basique d'un système RàPC.

Selon JL. Kolodner en 1992 [Kolodner1992], le cycle de travail d'un RàPC peut être décrit en quatre étapes, illustrées par la Figure 1.5. Ces étapes servent de base pour la majorité des systèmes RàPC introduits par la suite dans la littérature :

- La phase de remémoration : lors de l'entrée d'un nouveau cas dans le système, il s'agit de chercher et de récupérer le cas le plus similaire dans la base de cas. Plusieurs cas peuvent être remémorés. La sélection d'un cas utilise une mesure de similarité calculée entre le nouveau cas et les cas de la base de données.
- La phase d'adaptation : la solution du cas remémoré est modifiée pour s'adapter au nouveau cas.

- La phase d'évaluation : la solution adaptée est appliquée au nouveau cas. Si le résultat n'est pas satisfaisant, la solution doit passer par une révision, généralement manuelle, afin d'obtenir une solution correcte.
- La phase de stockage/capitalisation : c'est cette étape qui offre au RàPC la capacité d'apprentissage. Lorsque la solution est correcte, le nouveau cas et sa solution sont intégrés à la base de cas, afin de l'enrichir. Ainsi, le RàPC sera plus performant avec le temps, grâce à une base de cas plus conséquente et hétérogène. La base de cas peut également avoir une organisation hiérarchique permettant une optimisation en temps de la phase de remémoration, dans le cas d'une base de cas conséquente.

Aamodt et al. [Aamodt1994] proposent également un cycle de RàPC similaire mais plus simple contenant également 4 étapes : l'étape de remémoration, l'étape de réutilisation, dans laquelle il n'y a pas d'adaptation de la solution, l'étape de révision, et enfin l'étape de rétention. Les principes de ces phases sont globalement les mêmes que pour les phases de JL. Kolodner.

Le RàPC, combiné à une méthode de segmentation ou de fusion, peut alors être utilisé dans la segmentation d'images ou la fusion de segmentations, grâce aux paramètres de ces méthodes stockés dans la base. En 1999, P. Perner propose un système RàPC couplé à une méthode de seuillage pour la segmentation d'images médicales [Perner1999]. En 2008, M. Frucci, Perner et al. [Frucci2008] reprennent cette contribution en proposant un système RàPC pour la segmentation par *watershed* (ligne de partage des eaux). Cette méthode de segmentation considère l'image comme un relief topographique dont on simule l'inondation [Beucher1993]. Ces travaux sont eux-mêmes repris en 2012 pour être améliorés par A. Attig et P. Perner en 2012 [Attig2012].

### 1.3.3/ LA CONSTRUCTION ET LA REPRÉSENTATION 3D

La visualisation 3D à partir des segmentations calculées offre une aide intéressante pour la caractérisation du néphroblastome, et permet ainsi aux professionnels de visualiser les critères de la SIOP, planifier l'intervention et anticiper les éventuelles difficultés. Cette modélisation permet également de mieux illustrer l'envahissement tumoral. La Figure 1.6 montre la modélisation souhaitée du néphroblastome, qui a été réalisée sur le logiciel 3D-Slicer à partir des segmentations 2D, segmentations effectuées par les radiologues en semi-automatique sur le même logiciel à partir des images scanner.

La représentation holographique, comme présentée dans la Figure 1.7, permet une meilleure visualisation dans l'espace du néphroblastome ainsi qu'une meilleure communication avec le patient et la famille. À terme, elle pourrait servir, couplée à un système de reconnaissance de mouvements (de type *Kinect* [Gallo2011] ou *Leap Motion* [Weichert2013]), de représentation interactive. Les chirurgiens auraient alors la possibilité de visualiser la tumeur pendant l'opération, juste avec des gestes sans contact direct, ce qui limiterait également les risques d'infection. Il est également possible de réaliser des impressions 3D de cette reconstruction, avec des matériaux et des couleurs différents pour des raisons de communication, comme la représentation holographique.

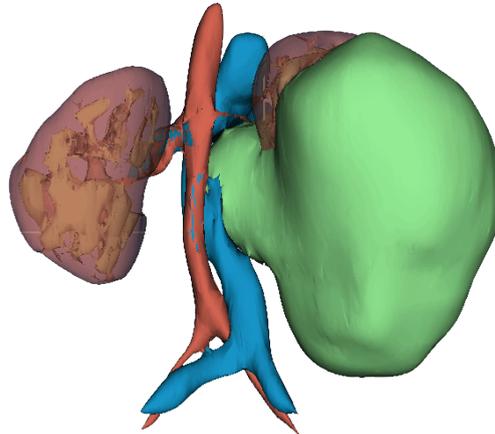


FIGURE 1.6 – Représentation 3D de la segmentation sémantique du néphroblastome, obtenue à partir des segmentations successives des coupes 2D des images scanner. La couleur marron représente les reins, le rouge l'artère, le bleu la veine, le vert la tumeur, et en jaune les cavités rénales.



FIGURE 1.7 – Représentation en pyramide holographique de la segmentation sémantique 3D du néphroblastome - réalisée par l'entreprise CFI dans le cadre du projet SAIAD.

## 1.4/ LA PLATE-FORME EXISTANTE COLISEUM-3D

Une plate-forme dédiée à la segmentation des différentes structures abdominales chez l'enfant atteint de néphroblastome, a déjà été conçue, dans le cadre de la thèse de Florent Marie [Marie2019]. Cette plate-forme porte le nom de **COLISEUM-3D** pour **COLaborative platform with artificial Intelligence for SEgmentation of tUmoral kidney in Medical images in 3D**. Présentée dans la Figure 1.8, elle est composée de trois couches : la **couche données**, la **couche segmentation**, et la **couche fusion**.

### 1.4.1/ LA COUCHE DE DONNÉES

La première couche de la plate-forme est la couche de données, regroupant l'ensemble des données utiles pour la réalisation des segmentations et des fusions de segmentations. Elle est composée de trois sous-couches :

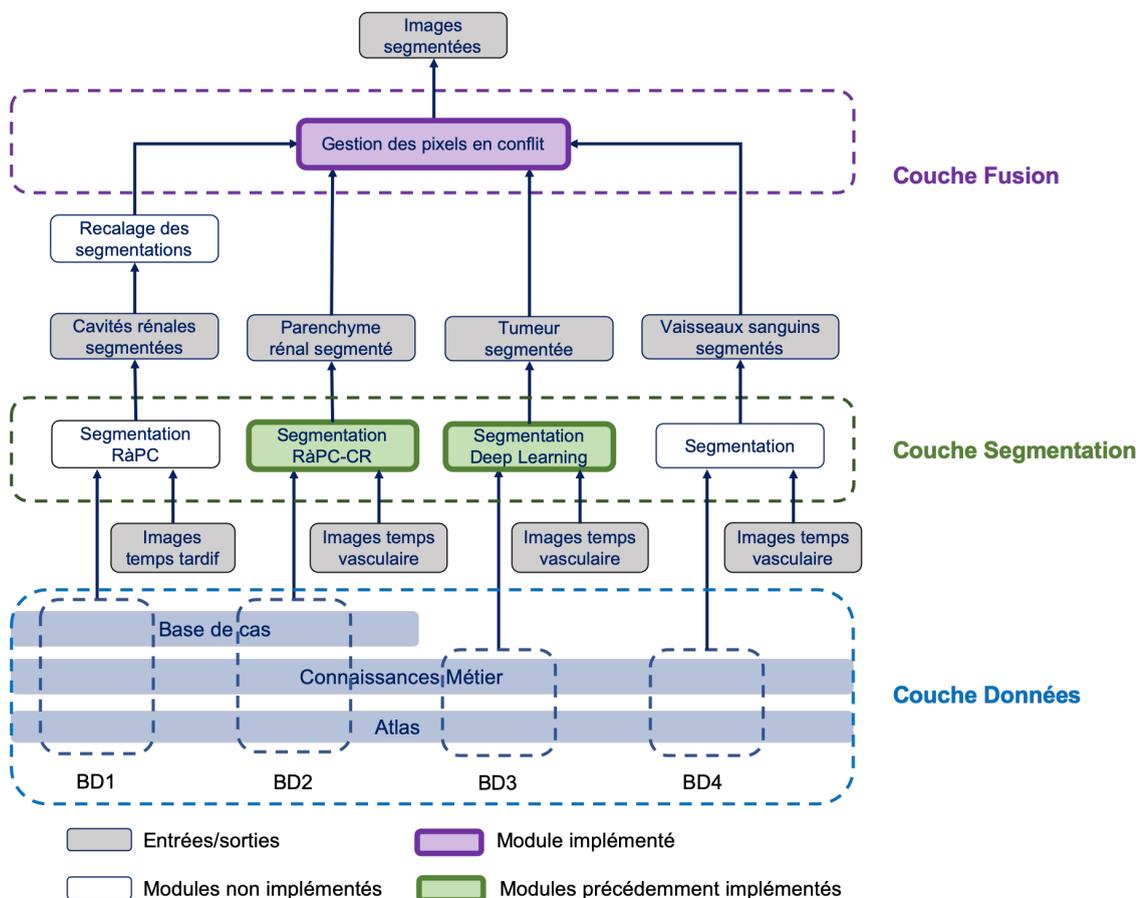


FIGURE 1.8 – Organisation de la plate-forme COLISEUM-3D.

- l'Atlas qui contient les images scanner des patients en format DICOM ou JPEG,
- les connaissances métier qui regroupent toute l'expertise médicale comme les segmentations réalisées manuellement ou semi-automatiquement par les radiologues afin de servir de référence et de résultat idéal (que l'on appelle généralement les **vérités terrain**),
- et enfin la base de cas qui s'appuie sur les deux autres sous-couches, dans le cas de techniques de segmentation comme le RàPC afin de rendre les données exploitables.

Ces 3 sous-couches permettent la réalisation d'une base de données pour chaque technique de segmentation utilisée. Par conséquent, cette couche est fortement liée à la couche de segmentation.

#### 1.4.2/ LA SEGMENTATION DES DIFFÉRENTES STRUCTURES

Le principe de la plate-forme COLISEUM-3D est de concevoir un système de segmentation propre à chaque structure. La couche de segmentation contient l'ensemble des processus de segmentation utilisés. Cette couche fait partie intégrante de la thèse de

Florent Marie [Marie2019]. Ainsi, pour la segmentation des reins pathologiques, un module de segmentation par **RàPC** couplé à un algorithme de segmentation par **croissance région** a été implémenté [Marie2018]. Dans ce système RàPC, une phase d'adaptation spécifique a été développée : elle permet de modifier les paramètres de la croissance région du cas remémoré afin de proposer une segmentation finale plus juste.

Pour la segmentation de la tumeur, un autre module a été implémenté utilisant du Deep Learning, et plus particulièrement un **réseau de neurones convolutif (CNN)** appelé FCN-8s [Long2015] utilisé avec une méthode originale d'apprentissage appelée **OVASSION (Overlearning Vector for Valid Sparse SegmentatIONS)** [Marie2019]. Cette méthode d'apprentissage consiste à réaliser un entraînement personnalisé à chaque patient et sur un faible jeu de données. Elle est présentée plus en détail dans la suite de ce document, Section 5.3. À la suite du processus de segmentation par Deep Learning, un post-traitement par **CRF (Champs Aléatoires Conditionnels, Conditional Random Field)** [Krahenbuhl2011] est finalement appliqué.

Dans cette couche de segmentation, il reste à concevoir et à implémenter un module de segmentation pour les vaisseaux sanguins à partir des images en temps vasculaire d'une part, et un module de segmentation par RàPC pour les cavités urinaires à partir des images en temps tardif d'autre part.

#### 1.4.3/ LE RECALAGE TEMPS VASCULAIRE / TEMPS TARDIF

La tumeur rénale, les reins et les vaisseaux sanguins sont visibles dans les images scanner en temps vasculaire. À l'inverse, les cavités rénales ne sont visibles que dans les images scanner en temps tardif pour des raisons présentées dans la Section 1.2.3. Cependant, plusieurs minutes s'écoulent entre les acquisitions des images en temps vasculaire puis en temps tardif. Le jeune patient, malgré les consignes de rester immobile, est susceptible de bouger, ce qui entraîne un décalage potentiel entre les deux temps d'acquisition. De plus, le simple fait de respirer peut également engendrer un décalage. Ainsi, afin d'obtenir l'ensemble de ces structures sur une seule représentation numérique, ces deux temps doivent être mis en correspondance : c'est ce que l'on entend par **recalage temps vasculaire / temps tardif**.

Dans le cas d'un recalage entre deux images, l'une d'elles est qualifiée d'*image de référence*, c'est-à-dire qu'aucune modification ne sera appliquée sur cette image, tandis que la seconde image sera alors une image dite *flottante*, transformée pour correspondre à l'image de référence. Il existe différentes méthodes de recalage qui consistent à appliquer une transformation spatiale sur l'image flottante. Des transformations simples peuvent être appliquées, comme une translation ou une rotation. Comme avec ces méthodes, l'image n'est pas déformée et les angles et distances sont conservés, on parle alors de recalage rigide. On peut également pratiquer des transformations affines ([Jenkinson2001]), ou encore élastiques ([Shen2002], [Kybic2003]) dans lesquelles toutes les déformations sont possibles. Ces différentes méthodes de recalage sont présentées dans l'état de l'art de Lisa Gottesfeld Brown [Brown1992].

L'application d'un recalage entre les images en temps vasculaire et en temps tardif ne résulte pas uniquement de simples transformations rigides, et il est nécessaire d'aller plus loin avec l'utilisation de transformations plus poussées. En effet, les reins ont tendance à avoir des mouvements plus complexes induits notamment par la respiration. Afin de déduire les transformations adéquates à appliquer, des points de repères présents

dans les deux images ont été déterminés (on parle alors d'une approche géométrique). Le pelvis est une structure osseuse intéressante et nous permet d'obtenir les points de repère requis. Pour l'heure, seules des transformations rigides ont été appliquées en déduction de ces points de repère. Bien que ces résultats préliminaires soient encourageants, les travaux futurs visent à utiliser des transformations élastiques afin de mieux prendre en compte les mouvements dus à la respiration.

#### 1.4.4/ LA FUSION DES DIFFÉRENTES SEGMENTATIONS

**Ces travaux de thèse s'intègrent dans le cadre de la couche fusion de la plateforme.** L'objectif de cette couche est de fusionner plusieurs segmentations en une seule segmentation consensus. Actuellement, le rein tumoral et la tumeur étant segmentés, la couche de fusion gère l'agrégation de ces deux structures. Comme ces structures sont complémentaires, il devrait, à priori, dans le cas de segmentations parfaitement calculées (par la couche précédente), ne pas y avoir de réelle difficulté. Chaque pixel des segmentations étant, dans ce cas idéal, associé à une seule structure, une simple union des segmentations devrait permettre d'obtenir la segmentation consensus.

Cependant, dans la pratique, les segmentations calculées ne sont pas parfaites et un pixel peut être associé à différentes structures. Il y a alors sur ce pixel un conflit qu'il faut résoudre. Le principe, en cas de conflits, est alors de rechercher la meilleure stratégie ou bien le meilleur critère de fusion afin de résoudre ces pixels en conflit et rendre une segmentation consensus la plus *juste* possible, la *justesse* d'une segmentation étant évaluée par rapport aux vérités terrain fournies par les experts du domaine (en l'occurrence ici, les segmentations guidées manuellement par les radiologues et chirurgiens du CHRU de Besançon). Un état de l'art sur les différentes méthodes de fusion de segmentations est présenté dans la suite de ce document dans les Chapitres 2 et 3.

### 1.5/ LES MÉTRIQUES D'ÉVALUATION DE SEGMENTATIONS

Lorsqu'une segmentation est réalisée, il est important de pouvoir évaluer la conformité du résultat par rapport à la vérité terrain. Pour cela, différentes métriques d'évaluation existent, permettant de comparer une image calculée avec une autre, appelée vérité terrain. Certaines de ces métriques peuvent être calculées avec la matrice de confusion, utilisant les facteurs **Vrai Positif (VP)**, **Vrai Négatif (VN)**, **Faux Positif (FP)** et **Faux Négatif (FN)**. La plupart des modèles de fusion classiques de segmentation se basent également sur certaines de ces métriques comme critères de fusion.

#### 1.5.1/ LES MÉTRIQUES UTILISANT LA MATRICE DE CONFUSION

La matrice de confusion [Provost1998], présentée par la Figure 1.9, peut être utilisée dans le cas plus large de l'évaluation d'un système de classification.

- Un pixel est dit Vrai Positif (VP) pour une classe donnée s'il a été étiqueté à cette classe dans la segmentation calculée et qu'il y est également dans la vérité terrain. Le pixel est alors correctement étiqueté.

		Prédiction	
		Vrai	Faux
Vérité terrain	Vrai	Vrai Positif (VP)	Faux Négatif (FN)
	Faux	Faux Positif (FP)	Vrai Négatif (VN)

FIGURE 1.9 – La matrice de confusion.

- Un pixel est dit Vrai Négatif (VN) pour une classe donnée s'il n'a pas été étiqueté à cette classe dans la segmentation calculée et qu'il n'y est également pas dans la vérité terrain. Le pixel est alors également correctement étiqueté.
- Un pixel est dit Faux Positif (FP) pour une classe donnée s'il a été étiqueté à cette classe dans la segmentation calculée mais qu'il n'y est pas dans la vérité terrain. Le pixel n'est alors pas correctement étiqueté.
- Un pixel est dit Faux Négatif (FN) pour une classe donnée s'il n'a pas été étiqueté à cette classe dans la segmentation calculée mais qu'il y est dans la vérité terrain. Le pixel n'est également pas correctement étiqueté.

Les métriques qui suivent utilisent cette matrice de confusion dans leur comparaison entre deux segmentations.

### 1.5.1.1/ LA SENSIBILITÉ ET LA SPÉCIFICITÉ

Les termes de sensibilité (également appelée **rappel**) et de spécificité sont des termes statistiques. La sensibilité est une mesure de la fraction Vrai Positif, tandis que la spécificité, opposée à la sensibilité, est une mesure de la fraction Vrai Négatif.

#### Définitions

La sensibilité  $SE$  et la spécificité  $SP$  sont définies par :

$$SE = \frac{N_{VP}}{N_{VP} + N_{FN}} \quad (1)$$

$$SP = \frac{N_{VN}}{N_{VN} + N_{FP}} \quad (2)$$

avec  $N_{XX}$  le nombre de pixels appartenant aux différentes catégories de la matrice de fusion.

## 1.5.1.2/ LE SCORE DICE

La métrique Dice, ou coefficient de Sørensen, est couramment utilisée par les experts en imagerie médicale. Elle donne une valeur de similarité, comprise entre 0 et 1, entre deux ensembles  $X$  et  $Y$ . Dans le contexte de l'imagerie et de la segmentation,  $X$  représente la segmentation calculée et  $Y$  la segmentation désirée (vérité terrain).

**Définition**

Soient deux ensembles  $X$  et  $Y$  à comparer, la métrique Dice (ou score Dice) est définie par :

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

Elle peut également être notée avec la matrice de confusion comme :

$$Dice = \frac{2N_{VP}}{2N_{VP} + N_{FP} + N_{FN}} \quad (4)$$

## 1.5.1.3/ LE SCORE IOU

L'indice *IoU* pour *Intersection over Union* ou appelé également indice de Jaccard est généralement utilisé dans la comparaison de segmentations calculées avec des méthodes de segmentation automatiques. Comme pour le score Dice, il compare deux ensembles  $X$  et  $Y$ , mais est plus "punitif" sur les erreurs de segmentations.

**Définition**

Soient deux ensembles  $X$  et  $Y$  à comparer, l'indice *IoU* est défini par :

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \quad (5)$$

Il peut également être noté avec la matrice de confusion comme :

$$IoU = \frac{N_{VP}}{N_{VP} + N_{FP} + N_{FN}} \quad (6)$$

## 1.5.1.4/ LA MÉTRIQUE F-MESURE

La métrique F-mesure (également nommée *F-measure* ou *harmonic mean of precision-recall measure*) [Martin2004] est une métrique utilisée pour évaluer la qualité de contours produits par une segmentation. Elle calcule un score de performance en évaluant l'accord entre les contours d'une segmentation calculée et les contours de la vérité terrain grâce à la mesure de **précision** et de **sensibilité** (Équation 1).

**Définitions**

La métrique F-Mesure est définie par :

$$F = \frac{2(P * SE)}{P + SE} \quad (7)$$

Où :

- $SE$  est la sensibilité
- $P$  est la précision définie par :

$$P = \frac{N_{VP}}{N_{VP} + N_{FP}} \quad (8)$$

## 1.5.1.5/ L'INDICE DE RAND

L'indice de Rand (RI) est introduit par [Rand1971] pour calculer la similarité entre deux ensembles de données. Dans le cas de la comparaison entre une segmentation calculée  $S^t$  et sa vérité terrain  $S^g$ , l'indice de Rand est donné par la somme du nombre de paires de pixels ayant la même étiquette dans  $S^t$  et  $S^g$  et de ceux ayant des étiquettes différentes, le tout divisé par le nombre total de paires de pixels possibles.

**Définitions**

Soit  $S^t$  la segmentation calculée et  $S^g$  la vérité terrain, l'indice de Rand est défini ainsi :

$$RI(S^t, S^g) = \frac{a + b}{a + b + c + d} = \frac{a + b}{\binom{n}{2}} \quad (9)$$

Où :

- $n$  est le nombre total de pixels d'une image,
- $a$  représente le nombre de paires de pixels assignées à la même région dans  $S^t$  et  $S^g$ ,
- $b$  représente le nombre de paires de pixels assignées dans la même région dans  $S^t$  mais dans des régions différentes dans  $S^g$ ,
- $c$  représente le nombre de paires de pixels assignées dans la même région dans  $S^g$  mais dans des régions différentes dans  $S^t$ ,
- et  $d$  représente le nombre de paires de pixels assignées à des régions différentes dans les deux segmentations.

Dans le cas où nous comparons deux segmentations ayant respectivement deux classes, alors  $a$ ,  $b$ ,  $c$  et  $d$  sont définis ainsi :

#### Définitions

$$a = \frac{1}{2}[N_{VP}(N_{VP} - 1) + N_{FP}(N_{FP} - 1) + N_{VN}(N_{VN} - 1) + N_{FN}(N_{FN} - 1)] \quad (10)$$

$$b = \frac{1}{2}[(N_{VP} + N_{FN})^2 + (N_{VN} + N_{FP})^2 - (N_{VP}^2 + N_{VN}^2 + N_{FP}^2 + N_{FN}^2)] \quad (11)$$

$$c = \frac{1}{2}[(N_{VP} + N_{FP})^2 + (N_{VN} + N_{FN})^2 - (N_{VP}^2 + N_{VN}^2 + N_{FP}^2 + N_{FN}^2)] \quad (12)$$

$$d = \frac{n(n-1)}{2} - (a + b + c) \quad (13)$$

Nous pouvons noter que le coefficient binomial  $\binom{n}{2}$  de l'équation 9 peut également être calculé comme  $\frac{n(n-1)}{2}$ .

### 1.5.2/ LA MÉTRIQUE VOI

La mesure Vol (*Variation of Information*) [Meilua2003] [Meilua2007] permet de comparer deux segmentations et de quantifier la quantité d'informations perdues ou gagnées en passant d'une étiquette à une autre. Il s'agit d'une métrique de comparaison probabiliste basée sur les régions qui utilise l'**entropie** et l'**information mutuelle** des segmentations, et qui est définie ci-dessous. Elle n'utilise donc pas la matrice de confusion. En imagerie, l'entropie peut également être calculée par l'histogramme normalisé de l'image. Plus le résultat de la métrique Vol est faible, et plus les deux segmentations sont similaires.

#### Définitions

Soit une segmentation  $S^t$ , l'entropie  $H$  de cette segmentation est définie par :

$$H(S^t) = - \sum_{i=1}^{R^t} P(i) \log P(i) = - \sum_{i=1}^{R^t} \frac{n_i^t}{n} \log \frac{n_i^t}{n} \quad (14)$$

et l'information mutuelle  $I$  entre deux segmentations  $S^t$  et  $S^g$  est définie par :

$$I(S^t, S^g) = \sum_{i=1}^{R^t} \sum_{j=1}^{R^g} P(i, j) \log \left( \frac{P(i, j)}{P(i)P(j)} \right) \quad (15)$$

Où :

- $R^t$  correspond au nombre de régions de la segmentation  $S^t$ ,

- $R^g$  correspond au nombre de régions de la segmentation  $S^g$ ,
- et  $P(i) = \frac{n_i}{n}$  est la probabilité qu'un pixel appartienne à la classe  $i$  de la segmentation  $S^t$  ( $n$  étant le nombre total de pixels de la segmentation),
- $P(i, j) = \frac{n_{ij}}{n}$ , la probabilité qu'un pixel appartienne à la classe  $i$  de la segmentation  $S^t$  et à la classe  $j$  de la segmentation  $S^g$ .

La mesure de la métrique Vol entre deux segmentations est alors la somme de leurs entropies respectives moins deux fois leur information mutuelle.

#### Définition

Soit  $S^t$  et  $S^g$  la segmentation calculée et la segmentation désirée, le Vol entre ces deux segmentations est défini par :

$$Vol(S^t, S^g) = H(S^t) + H(S^g) - 2 * I(S^t, S^g) \quad (16)$$

Où :

- $H()$  est l'entropie d'une segmentation,
- $I()$  est l'information mutuelle entre deux segmentations.

#### 1.5.3/ LA MÉTRIQUE GCE

L'erreur de la cohérence globale, ou en anglais *Global Consistency Error*, nommée *GCE* [Martin2001], est une métrique d'évaluation qui considère la nature multi-échelle des segmentations en évaluant la manière dont une segmentation peut constituer un raffinement d'une autre segmentation. Elle est évidemment utilisée pour évaluer la similarité d'une segmentation calculée avec la vérité terrain. Une segmentation est similaire à une autre si chacune de ces régions obtenues est un sous-ensemble, ou géométriquement identique, à une région dans la deuxième segmentation.

#### Définition

Soit  $S^t = \{C_1^t, C_2^t, \dots, C_{R^t}^t\}$  et  $S^g = \{C_1^g, C_2^g, \dots, C_{R^g}^g\}$  la segmentation calculée  $S^t$  et la vérité terrain  $S^g$  comprenant chacune  $R^t$  et  $R^g$  régions, ainsi que  $p_i$  un pixel de l'image. La mesure GCE utilise l'**erreur de raffinement local (LRE)** notée :

$$LRE(S^t, S^g, p_i) = \frac{|C_{<p_i>}^t \setminus C_{<p_i>}^g|}{|C_{<p_i>}^t|} \quad (17)$$

Où :

- $C_{\langle p_i \rangle}^t$  est la région de la segmentation  $S^t$  incluant le pixel  $p_i$ ,
- $C_{\langle p_i \rangle}^g$  est la région de la vérité terrain  $S^g$  incluant le pixel  $p_i$ .

Dans l'équation 17,  $|C_{\langle p_i \rangle}^t|$  est alors la cardinalité de l'ensemble des pixels de la région  $C$  dans lequel se situe le pixel  $p_i$  de la segmentation  $S^t$ , et  $\setminus$  est l'opérateur de différenciation des ensembles<sup>1</sup>.

La mesure  $LRE$  n'est pas symétrique, et permet seulement de savoir si  $S^t$  est un raffinement de  $S^g$  et non le contraire. La métrique  $GCE$  vise à rendre la mesure symétrique et à l'appliquer à chaque pixel d'une segmentation. Son résultat se situe dans l'intervalle  $[0,1]$ , 0 signifiant que la segmentation calculée correspond parfaitement à la vérité terrain et 1 représente une segmentation complètement différente de la vérité terrain.

#### Définition

Soit  $S^t$  et  $S^g$  la segmentation calculée et la vérité terrain, la formule  $GCE$  est définie par :

$$GCE(S^t, S^g) = \frac{1}{n} \min \left\{ \sum_{i=1}^n LRE(S^t, S^g, p_i), \sum_{i=1}^n LRE(S^g, S^t, p_i) \right\} \quad (18)$$

Où :

- $n$  est le nombre total de pixels de l'image des segmentations,
- $LRE$  est le calcul de l'erreur de raffinement local présenté dans l'équation 17.

1. Également appelé complément relatif. Si  $A \setminus B$  alors il s'agit de l'ensemble des éléments appartenant à  $A$  mais pas à  $B$ .

**Synthèse du chapitre****SYNTHÈSE**

Dans ce premier chapitre nous avons détaillé le projet SAIAD (**S**egmentation **A**utomatique de reins tumoraux chez l'enfant par **I**ntelligence **A**rtificielle **D**istribuée) qui a pour objectif de proposer une plate-forme capable de segmenter automatiquement, c'est-à-dire sans intervention humaine, ou a minima avec une intervention limitée, le néphroblastome, qui est la tumeur rénale la plus fréquente chez l'enfant. Différentes structures sont impactées par le néphroblastome, et l'objectif général est d'obtenir une représentation 3D fidèle de l'abdomen du patient, permettant aux professionnels (chirurgiens et radiologues) d'acquérir de nombreuses informations utiles. La représentation de ces structures est réalisée par des segmentations, à partir des images scanner du patient.

Les techniques de segmentation d'images sont nombreuses, et celles par IA le sont encore plus, en particulier par réseau de neurones : un axe de recherche présent dans la littérature de manière très importante ces dernières années.

La plate-forme COLISEUM-3D, développée au sein du projet par Florent Marie, récent Docteur de notre équipe de recherche, a pour objectif de faire collaborer différentes approches de segmentations et de fusion de segmentations afin de fournir la représentation 3D du néphroblastome. Elle est composée d'une couche de données, d'une couche de segmentation et d'une couche de fusion. C'est dans la couche de segmentation que sont réalisées les segmentations des structures, ce travail faisant partie de la thèse de Florent Marie. Ainsi, le rein pathologique a été segmenté par une technique de RàPC couplée à la croissance région, et la segmentation de la tumeur a été calculée par réseau de neurones, et plus particulièrement par un CNN (FCN-8s). Une méthode spécifique d'apprentissage des réseaux de neurones pour la segmentation, appelée OV<sup>2</sup>ASSION, a également été conçue dans le cas de jeu d'apprentissage contenant une faible quantité de données.

À la suite de la couche de segmentation, le verrou le plus important était de développer la **couche de fusion**, qui est le cœur de la présente thèse. Cette couche a pour objectif d'agrèger les différentes segmentations calculées automatiquement, contenant chacune une structure, afin d'obtenir une segmentation finale regroupant l'ensemble de ces structures. Lors de la fusion, des conflits entre les pixels peuvent apparaître et des méthodes pour gérer ces conflits sont nécessaires.

Enfin, afin de valider la conformité par rapport à la vérité terrain des segmentations calculées, il existe de nombreuses métriques d'évaluation comparant deux segmentations. En présence d'une segmentation calculée et d'une segmentation dite vérité terrain, on peut alors évaluer la *justesse* de la segmentation calculée. Certaines de ces métriques peuvent également être utilisées dans les méthodes de gestion de conflits lors de la fusion de ces segmentations, que nous verrons dans le chapitre suivant.

Dans le chapitre suivant, nous présentons l'état de l'art des différentes méthodes de fusion de segmentations que nous avons appelées classiques. Ainsi, des méthodes simples, comme le vote, des méthodes de fusion utilisées dans le cadre de la segmentation dites Multi-Atlas, ou encore des méthodes de fusion utilisant les métriques d'évaluation sont présentées.

## LES MÉTHODES CLASSIQUES DE FUSION DE SEGMENTATIONS

La fusion de multiples segmentations est étudiée principalement afin d'améliorer la segmentation d'images. Certaines approches ont pour principe de concevoir une méthodologie qui segmente plusieurs fois une image avec des algorithmes différents, voire le même algorithme avec différents paramètres (l'algorithme de segmentation utilisé étant généralement assez simple et rapide), dans le but d'obtenir différentes segmentations à fusionner à l'aide d'une procédure de fusion. L'hypothèse avancée est que la segmentation consensus générée soit plus juste. Au lieu de chercher le meilleur algorithme de segmentation avec des paramètres optimaux, ce qui peut devenir lourd et coûteux, il s'agit plutôt de s'intéresser à trouver **les meilleurs stratégies et critères de fusion** de segmentations.

Dans la littérature est également décrite une approche de segmentation appelée segmentation Multi-Atlas utilisant une étape de fusion sur plusieurs segmentations afin d'obtenir la segmentation finale. Les méthodes de fusion utilisées sont principalement les mêmes et seront décrites dans ce chapitre.

Ainsi, nous proposons un état de l'art général sur les différentes méthodes de fusion de segmentations dites *classiques*, dans le sens où elles ne font pas appel à des concepts d'IA. Ces dernières ont été classées en plusieurs catégories : les méthodes de fusion intuitives (comme le vote majoritaire ou l'intersection et l'union), les méthodes de fusion utilisées dans le cadre de la segmentation Multi-Atlas, les méthodes utilisant des métriques, les méthodes basées sur les distances et les similarités et enfin les combinaisons de différentes méthodes de fusion.

### 2.1/ LES MÉTHODES DITES INTUITIVES

Les méthodes *intuitives* sont celles qui paraissent les plus évidentes et les plus faciles pour fusionner les segmentations et gérer les éventuels conflits qui peuvent dès lors apparaître. Cependant, ces méthodes sont loin d'être les plus efficaces dans le cas de segmentations plus complexes. En effet, si des erreurs d'étiquetage se retrouvent dans la plupart voire dans une seule des segmentations à fusionner, alors ces erreurs peuvent très facilement se retrouver dans la segmentation finale, la rendant ainsi inexacte.

## 2.1.1/ LE VOTE

### 2.1.1.1/ LA FUSION PAR VOTE MAJORITAIRE

La technique de fusion par vote majoritaire est la plus simple à mettre en œuvre. Elle consiste simplement à affecter à chaque pixel de la segmentation finale l'étiquette qui apparaît le plus fréquemment pour ce pixel dans toutes les segmentations à fusionner ([Rohlfing2004a, Mitchell2010a]). La Figure 2.1 présente la fusion par vote majoritaire de 3 segmentations de taille  $5 \times 5$  et étiquetées 0 ou 1.

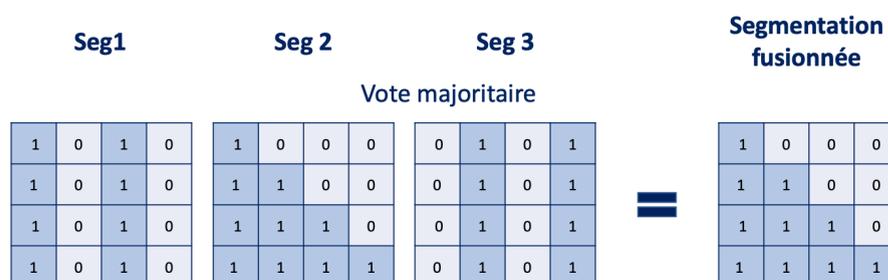


FIGURE 2.1 – Exemple de la fusion de 3 segmentations de taille  $5 \times 5$  par vote majoritaire.

Cependant, en cas d'égalité pouvant survenir dans le cas de la fusion d'un nombre pair de segmentations, il n'est pas possible de définir l'étiquette du pixel. Il est alors nécessaire de trouver une autre solution afin de pouvoir déterminer l'étiquette finale.

### 2.1.1.2/ LA FUSION PAR VOTE PONDÉRÉ

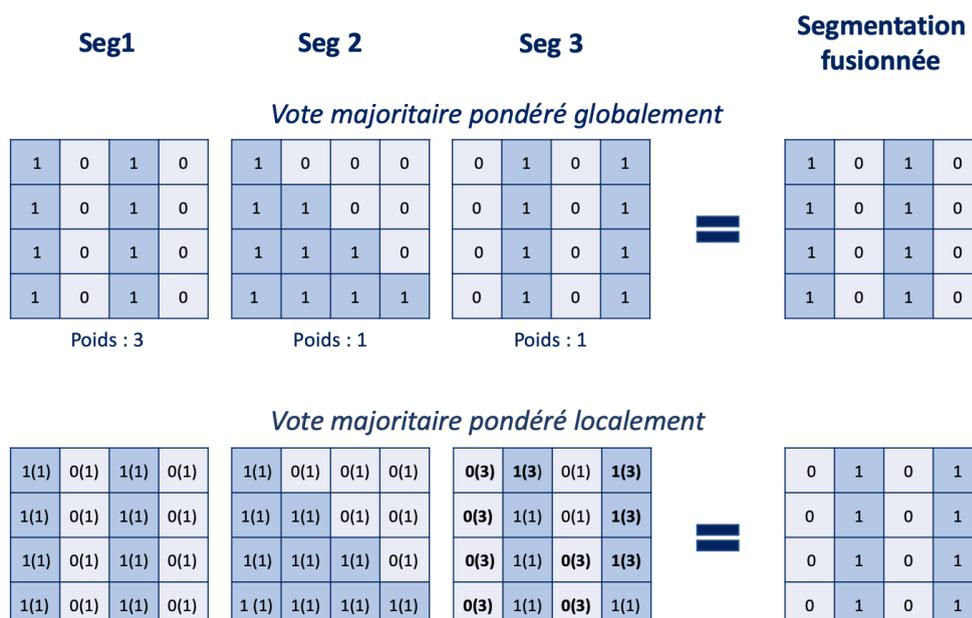


FIGURE 2.2 – Exemple de la fusion de 3 segmentations de taille  $5 \times 5$  par vote majoritaire pondéré globalement et localement.

La fusion par vote pondéré permet de donner un poids, soit globalement, soit localement aux segmentations à combiner ([Artaechevarria2009]). Pour le définir de manière simple, un vote pondéré consiste à donner un poids à chaque segmentation à fusionner, influençant donc la fusion par vote sur les segmentations ayant les poids les plus élevés. Un vote pondéré localement permet de donner un poids à chaque pixel ou région de l'image. Les segmentations sont alors composées de différents poids, et certaines parties des segmentations sont donc privilégiées. La Figure 2.2 représente ces deux types de vote avec, dans le cas du vote pondéré globalement, un poids de 3 pour la première segmentation et 1 pour les deux autres, et dans le cas du vote pondéré localement, un poids de 3 pour certains pixels de la troisième segmentation et un poids de 1 pour tous les autres pixels.

Comme pour la fusion par vote majoritaire sans pondération, des égalités peuvent se produire et il ne sera alors pas possible de fusionner ces pixels sans utiliser une autre méthode.

### 2.1.2/ L'INTERSECTION ET L'UNION

Un moyen simple et rapide de fusionner plusieurs segmentations est de réaliser leur intersection ou leur union. Pour calculer la segmentation consensus en utilisant l'intersection de multiples segmentations, les différentes étiquettes d'un pixel doivent être les mêmes sur toutes les segmentations à fusionner pour que l'étiquette du pixel de la segmentation consensus possède la même valeur. Pour utiliser la règle de l'union, les segmentations à fusionner doivent être de type *arrière-plan* et *avant-plan* : dit autrement, des segmentations ne contenant que deux étiquetages possibles (une étiquette pour l'arrière-plan et une étiquette pour l'objet de premier plan, avec une prédominance pour l'étiquette de premier plan). L'étiquette du pixel consensus sera alors étiquetée comme avant-plan si au moins une étiquette possède cette valeur. Dans le cas contraire, l'étiquette consensus aura la valeur de l'arrière-plan.

Ces deux règles de fusion nécessitent au préalable de bonnes segmentations calculées. En effet, avec la règle de l'intersection, si toutes les segmentations calculées ne sont pas de bonnes qualités et qu'elles diffèrent beaucoup les unes des autres, la segmentation consensus sera très différente (voire absence de l'objet segmenté dans certains cas dans lesquels il n'y a aucune étiquette en commun). Dans le cas de la règle de l'union, une simple erreur d'étiquetage de l'objet en avant plan dans une segmentation se retrouvera automatiquement dans la segmentation consensus. De plus, la règle de l'union est particulièrement adaptée pour des segmentations ne contenant que 2 étiquetages possibles car elle devient difficilement gérable avec un nombre plus important d'étiquettes. En effet, si trois valeurs sont possibles pour un pixel, il sera nécessaire d'appliquer une autre règle afin de pouvoir arbitrer, comme, par exemple donner une priorité à certaines étiquettes par rapport à d'autres.

En l'occurrence, ces deux méthodes sont utilisées dans [Cabria2017] pour réaliser la fusion de multiples segmentations de cerveaux, calculées avec trois méthodes de segmentations différentes, à partir d'images scanner.

## 2.2/ LES MÉTHODES DE FUSION DANS LE CADRE DE LA SEGMENTATION MULTI-ATLAS

La segmentation Multi-Atlas est un principe de segmentation utilisant une étape de fusion, beaucoup utilisée dans le milieu médical ([Aljabar2009, Alkama2015, Nguyen2018]). Cette approche s'appuie sur un ensemble d'images et de segmentations correspondantes appelées Atlas. Le processus de segmentation par Multi-Atlas se déroule ensuite en 3 étapes :

- Tout d'abord, l'image en entrée que l'on souhaite segmenter est comparée aux Atlas pour déterminer les cas les plus proches.
- Ces cas sont alors recalés de manière à correspondre au plus près à l'image en entrée.
- Puis, l'étape de fusion consiste à fusionner toutes les segmentations des Atlas sélectionnés afin d'obtenir la segmentation pour l'image en entrée.

Gorthi Subrahmanyam ([Gorthi2013]) propose un état de l'art sur plusieurs techniques de fusion. La méthode de fusion principalement utilisée est très simple : il s'agit de la fusion par différents types de vote (cf. Section 2.1.1). Cependant, d'autres méthodes de fusion sont apparues dans la segmentation Multi-Atlas, comme la fusion avec la méthode **SBA** et la méthode **STAPLE**. Ces méthodes sont utilisées dans le cadre de la fusion de multiples segmentations similaires, représentant la même information, et non des segmentations complémentaires.

### 2.2.1/ LA MÉTHODE SBA

La méthode SBA pour *Shape Based Averaging* [Rohlfing2007] est une méthode basée sur des fonctions de distances euclidiennes signées. La distance de chaque pixel par rapport à une étiquette est alors calculée.

Notons  $d_{k,l}(\vec{x})$  la distance euclidienne signée d'un pixel à la position  $\vec{x}$  dans la segmentation  $k$  avec l'étiquette  $l$ . La distance pour cette étiquette  $l$  est alors la différence entre deux autres distances euclidiennes non signées :

- La distance, appelée extérieure, qui est, pour chaque pixel, la distance du pixel le plus proche possédant ce label,
- et la distance, appelée intérieure, qui est, pour chaque pixel, la distance du pixel le plus proche possédant un autre label.

Par conséquent, la distance extérieure est nulle pour tous les pixels possédant ce label, tandis que la distance intérieure est nulle pour tous les autres pixels ne possédant pas ce label.

**Définition**

Soit un pixel à la position  $\vec{x}$ , sa distance moyenne  $D_l(\vec{x})$  par rapport à l'étiquette actuelle  $l$  est calculée en additionnant les distances de toutes les segmentations et en divisant par le nombre total de segmentations :

$$D_l(\vec{x}) = \frac{1}{K} \sum_{k=1}^K d_{k,l}(\vec{x}) \quad (19)$$

Où :

- $K$  est le nombre total de segmentations,
- $d_{k,l}(\vec{x})$  est la distance euclidienne signée d'un pixel à la position  $\vec{x}$  dans la segmentation  $k$  avec l'étiquette  $l$ .

Afin d'obtenir la segmentation consensus, le principe de la méthode SBA est de calculer itérativement en minimisant la distance moyenne sur toutes les étiquettes possibles.

**Définition**

Soit un pixel de la segmentation consensus  $S$  à la position  $\vec{x}$ , son étiquette est alors l'étiquette ayant obtenu la plus faible distance  $D_l(\vec{x})$  :

$$S(\vec{x}) = \arg \min_{l \in \varepsilon} D_l(\vec{x}) \quad (20)$$

- où  $\varepsilon$  est l'ensemble des étiquettes possibles de la segmentation.

## 2.2.2/ LA MÉTHODE STAPLE

La méthode STAPLE pour *Simultaneous Truth And Performance Level Estimation* [Mitchell2010b] consiste à estimer de manière probabiliste la segmentation finale en estimant la combinaison optimale des segmentations à fusionner, chaque segmentation étant pondérée en fonction d'un niveau de performance calculé. Initialement, la méthode STAPLE a été créée dans le but de fusionner les différentes segmentations manuelles réalisées par différents experts afin d'obtenir une segmentation consensus plus juste. Il est bien entendu possible de l'utiliser également sur plusieurs segmentations calculées automatiquement.

Cette méthode retourne une carte de probabilités qui estime la vraisemblance de chaque pixel à appartenir aux classes. Cette méthode est dérivée de l'**algorithme de maximisation de l'espérance** (souvent abrégée EM pour *Expectation Maximization*). De manière itérative, la probabilité maximale d'une segmentation consensus par rapport à l'ensemble des segmentations à fusionner en fonction de la sensibilité et la spécificité de chaque segmentation est calculée.

Soit  $p$  et  $q$  la sensibilité et la spécificité, le principe est alors d'effectuer itérativement :

- l'estimation de la segmentation consensus  $S_n$  qui maximise la probabilité  $P(S_{n-1} | \{S\}, p_{n-1}, q_{n-1})$ , avec  $\{S\}$  l'ensemble des segmentations à fusionner,
- puis la mise à jour  $p_n$  et  $q_n$  en fonction de la segmentation consensus calculée  $S_n$ .

## 2.3/ LES MÉTHODES BASÉES SUR UNE OU PLUSIEURS MÉTRIQUES

Les modèles de fusion basés sur une ou plusieurs métriques consistent simplement à modifier les étiquettes des pixels d'une segmentation afin d'optimiser la ou les valeurs de métrique, et ainsi déterminer la segmentation consensus optimale. Cette optimisation peut se faire soit par un algorithme de recuit simulé stochastique [Geman1987] mais qui demeure coûteux, soit par des algorithmes itératifs plus rapides appelés **modèles conditionnels itérés** ou *ICM* [Besag1986]. Techniquement, ICM est un algorithme itératif de descente d'énergie locale la plus raide (*an iterative steepest local energy descent algorithm*), qui cherche à obtenir, pour chaque pixel ou région de la segmentation, l'étiquette d'énergie optimale. Ces énergies sont les différentes métriques à calculer qui sont présentées à la suite de cette section.

L'Algorithme 1 présente un modèle type de réalisation de fusion de segmentations en utilisant l'algorithme ICM avec une métrique :

- Tout d'abord, la plupart des algorithmes possèdent au minimum certaines données afin de mener à bien la fusion : l'ensemble des segmentations à fusionner, l'ensemble des étiquettes possibles d'un pixel dans ces segmentations, le nombre maximal d'itérations, et la métrique utilisée. Dans ce cas, on parle de **métrique moyenne** car la métrique se calcule entre une segmentation et un ensemble d'autres segmentations. Comme la fusion concerne plus de deux segmentations, il est intéressant de calculer la métrique moyenne qui compare alors une segmentation avec l'ensemble de toutes les segmentations à fusionner.
- L'étape d'initialisation permet de choisir la segmentation optimale pour commencer. Cette segmentation servira de base et les étiquettes de ces pixels seront modifiées petit à petit jusqu'à obtenir la segmentation consensus finale. En fonction de la métrique utilisée, la valeur optimale est soit la valeur minimale, soit la valeur maximale de la métrique moyenne.
- Ensuite, si la métrique est de niveau pixel, chaque pixel de la segmentation initiale sera parcouru. Dans le cas où la métrique est de niveau région, chaque région de la segmentation sera parcourue.
- L'étape suivante consiste en l'arbitrage des étiquettes pour chaque pixel/région : pour chaque étiquette possible, le pixel ou la région de la segmentation est associé à cette étiquette, puis la métrique moyenne est calculée par rapport à l'ensemble des segmentations à fusionner. L'étiquette finale de la segmentation consensus sera alors l'étiquette dont la valeur de la métrique est optimale.

- Ce processus est réitéré sur chaque élément à étiqueter dans la segmentation, et ce parcours est réalisé plusieurs fois en fonction du nombre maximal d'itérations désirées.

---

**Algorithme 1** : Modèle type de fusion utilisant une métrique et l'algorithme ICM.

---

**Données :**

$\{S_k\}_{k \leq L}$  : L'ensemble des  $L$  segmentations à fusionner

$\varepsilon$  : L'ensemble des étiquettes possibles d'un pixel

$T_{max}$  : Le nombre maximal d'itérations

$\overline{M}$  : La métrique moyenne utilisée

1 **1. Initialisation**

2  $S_M^{[0]} = \arg \min / \max_{S \in \{S_k\}_{k \leq L}} \overline{M}(S, \{S_k\}_{k \leq L})$

3 **2. Steepest local energy descent**

4 **tant que**  $i < T_{max}$  **faire**

5     **pour** chaque pixel/région  $p$  avec l'étiquette  $l$  dans la segmentation  $S$  **faire**

6          $V_M^o = \arg \min / \max_{x \in \varepsilon} \overline{M}(S, \{S_k\}_{k \leq L})_{l_s^p \rightarrow x}$

7         Remplacer l'étiquette  $l_s^p$  par l'étiquette  $x$  de  $V_M^o$  dans la segmentation  $S$

8     **fin**

9      $i = i + 1$

10 **fin**

---

L'utilisation de cet algorithme couplé aux différentes métriques, présentées ci-dessous, permet la fusion de multiples segmentations représentant la même information. Chaque pixel (ou région si une approche par région est privilégié) de l'image est alors parcourue et traité, et son étiquette potentiellement modifiée. De plus, chaque segmentation pouvant avoir un nombre différent de régions, c'est à l'algorithme de déterminer le nombre final de régions en en ajoutant ou en en retirant si nécessaire. Ces méthodes ne sont donc pas prévues à la base pour la fusion de segmentations complémentaires, ainsi que pour traiter seulement les pixels en conflits.

### 2.3.1/ LA MÉTHODE BASÉE SUR F-MESURE

Une des méthodes de fusion de segmentations réalisée par Mignotte et al. en 2014 ([Mignotte2014b]) utilise la métrique d'évaluation F-Mesure (cf. Équation 7) comme critère de fusion. Cette métrique se base sur les contours des segmentations. Dans le modèle de fusion, la précision et le rappel sont pondérés par une valeur  $\alpha$  positive.

**Définition**

La métrique F-mesure est alors définie par :

$$F_\alpha = \frac{PR}{\alpha R + (1 - \alpha)P} \quad (21)$$

Où :

- $P$  est la mesure de précision,
- $R$  est la mesure de rappel (également appelée mesure de sensibilité),
- $\alpha$  est le coût relatif entre la précision et le rappel.

Cette méthode permet dans son calcul de déterminer le nombre de région ou classe de la segmentation consensus. Ainsi, plus  $\alpha$  est proche de 0 et plus le modèle de fusion favorise une sur-segmentation de la segmentation consensus. À l'inverse, plus la valeur est proche de 1 et plus le modèle de fusion favorise une segmentation consensus unifiée possédant un faible nombre d'étiquettes. De même, si le calcul est basé entre une segmentation calculée et une vérité terrain,  $\alpha$  peut être alors fixé à 0.5 (ce qui donne  $F = (2 * PR)/(P + R)$ ) pour évaluer objectivement les deux segmentations.

Les définitions de la précision et du rappel peuvent également être révisées pour l'utilisation de la métrique F-Mesure dans la fusion de segmentations. La précision représente alors la fraction des détections qui sont de véritables contours. C'est-à-dire pour une région donnée, l'ensemble des pixels appartenant au contour de cette région dans la segmentation calculée et dans la vérité terrain, le tout sur l'ensemble des pixels appartenant au contour de la région de la segmentation calculée (dont ceux qui n'appartiennent pas au contour dans la vérité terrain). Le rappel représente la fraction des véritables contours détectés. Dit autrement, pour une région donnée, c'est l'ensemble des pixels appartenant au contour de cette région dans la segmentation calculée et dans la vérité terrain, le tout sur l'ensemble des pixels appartenant au contour de la région dans la vérité terrain.

**Définitions**

Soit la segmentation calculée  $S^a = \{C_1^a, C_2^a, \dots, C_{R^a}^a\}$  comportant un nombre  $R^a$  de régions  $C$ , et la segmentation  $S^b = \{C_1^b, C_2^b, \dots, C_{R^b}^b\}$  comportant un nombre  $R^b$  de régions  $C$ , la précision  $P$  et le rappel  $R$  sont alors définis par :

$$P = \frac{|B_{c^a} \cap B_{c^b}|}{|B_{c^a}|} \quad (22)$$

$$R = \frac{|B_{c^a} \cap B_{c^b}|}{|B_{c^b}|} \quad (23)$$

Où :

- $B_{C^a}$  est l'ensemble des pixels appartenant au contour de la région  $C^a$  de la segmentation  $S^a$ ,
- $B_{C^b}$  est l'ensemble des pixels appartenant au contour de la région  $C^b$  de la segmentation  $S^b$ .

La mesure moyenne de F-mesure est définie par :

$$\overline{F}_\alpha(S^t, \{S_k^g\}_{k \leq L}) = \frac{1}{L} \sum_{k=1}^L F_\alpha(S^t, S_k^g) \quad (24)$$

Où :

- $S^t$  est une segmentation,
- $\{S_k^g\}_{k \leq L}$  est un ensemble de  $L$  segmentations.

$\overline{F}_\alpha$  sera alors élevée si la segmentation  $S^t$  est cohérente avec l'ensemble  $\{S_k^g\}_{k \leq L}$ .

La maximisation de  $\overline{F}_\alpha$  permet ainsi d'obtenir la segmentation consensus optimale, définie par :

$$\hat{S}_{\overline{F}_\alpha} = \arg \max_{S \in S_n} \overline{F}_\alpha(S, \{S_k\}_{k \leq L}) \quad (25)$$

Où :

- $S_n$  est l'ensemble de toutes les segmentations possibles en utilisant  $n$  pixels.

Afin d'améliorer le résultat de la segmentation consensus  $\hat{S}_{\overline{F}_\alpha}$ , une pondération de l'importance de chaque segmentation calculée est ajoutée. La segmentation consensus est alors calculée comme suit :

$$\hat{S}_{\overline{F}_\alpha} = \arg \max_{S \in S_n} \left\{ \frac{1}{L} \sum_{k=1}^L w_k * F_\alpha(S, \{S_k\}_{k \leq L}) \right\} \quad (26)$$

L'algorithme du modèle de fusion est alors l'algorithme ICM (cf. Algorithme 1) permettant la maximisation de  $\overline{F}_\alpha$  de manière itérative.

Le modèle de fusion se basant sur *F-Mesure* réalise donc des fusions par région et non par pixel. Idéalement, un grand nombre de régions est important dans les segmentations calculées pour obtenir une bonne fusion avec cette métrique, et des méthodes de segmentation par superpixels sont préférables.

## 2.3.2/ LA MÉTHODE BASÉE SUR VOI

Un protocole de fusion de multiples segmentations dérivé de la méthode basée sur la *F-Mesure* mais utilisant cette fois le Voi (cf. Équation 16 du chapitre précédent) comme critère pour la fusion des multiples segmentations a été introduit dans [Mignotte2014a]. En considérant cette Équation 16 et l'ensemble de  $L$  segmentations  $\{S_k^g\}_{k \leq L}$  à fusionner, le Voi moyen ( $\overline{Voi}$ ) est alors utilisé pour comparer une segmentation  $S^t$  avec l'ensemble  $\{S_k^g\}_{k \leq L}$ .

$$\overline{Voi}(S^t, \{S_k\}_{k \leq L}) = \frac{1}{L} \sum_{k=1}^L Voi(S^t, S_k) \quad (27)$$

Plus le résultat de  $\overline{Voi}$  sera faible, plus la segmentation  $S^t$  sera proche de l'ensemble de segmentations. Pour réaliser la fusion des segmentations en une segmentation consensus, l'Équation 28 est appliquée :

$$\hat{S}_{\overline{Voi}} = \arg \min_{S \in S_n} \overline{Voi}(S, \{S_k\}_{k \leq L}) \quad (28)$$

Où :

- $S_n$  est l'ensemble de toutes les segmentations possibles à  $n$  pixels.

Ce modèle de fusion est alors un modèle générateur de segmentations probables dans lequel la segmentation consensus est déterminée en fonction de la valeur la plus faible de  $\overline{Voi}$ . L'algorithme ICM (cf. Algorithme 1) permet la minimisation de  $\overline{Voi}$  en parcourant pixel par pixel la segmentation. Cependant, le coût de l'utilisation de la métrique  $\overline{Voi}$  peut s'avérer excessif. En conséquence, son expression locale dérivée  $\Delta \overline{Voi}$  peut être employée :

$$\begin{aligned} \Delta \overline{Voi}(S, \{S_k\}_{k \leq L})_{s:m \rightarrow x} = & L \cdot \left\{ -\frac{n_m}{n} \log\left(\frac{n_m}{n}\right) - \frac{n_x}{n} \log\left(\frac{n_x}{n}\right) \right. \\ & \left. + \frac{n_m - 1}{n} \log\left(\frac{n_m - 1}{n}\right) + \frac{n_x + 1}{n} \log\left(\frac{n_x + 1}{n}\right) \right\} \\ & - 2 \cdot \sum_{l=1}^L \left\{ \frac{n_m \mathcal{L}_s^l}{n} \log\left(\frac{n_m \mathcal{L}_s^l}{n} \cdot \frac{n}{n_m} \cdot \frac{n}{n_{\mathcal{L}_s^l}}\right) \right. \\ & \left. + \frac{n_x \mathcal{L}_s^l}{n} \log\left(\frac{n_x \mathcal{L}_s^l}{n} \cdot \frac{n}{n_x} \cdot \frac{n}{n_{\mathcal{L}_s^l}}\right) \right. \\ & \left. - \frac{(n_m \mathcal{L}_s^l - 1)}{n} \log\left(\frac{(n_m \mathcal{L}_s^l - 1)}{n} \cdot \frac{n}{n_m - 1} \cdot \frac{n}{n_{\mathcal{L}_s^l}}\right) \right. \\ & \left. - \frac{(n_x \mathcal{L}_s^l + 1)}{n} \log\left(\frac{(n_x \mathcal{L}_s^l + 1)}{n} \cdot \frac{n}{n_x + 1} \cdot \frac{n}{n_{\mathcal{L}_s^l}}\right) \right\} \end{aligned} \quad (29)$$

Où :

- $n$  est le nombre total de pixels,
- $L$  est le nombre de segmentations à fusionner,
- $\mathcal{L}_s^l$  est l'étiquette du pixel à la position  $s$  dans la segmentation  $S_l \in \{S_k\}_{k \leq L}$ ,
- $n_m, n_x$  et  $n_{\mathcal{L}_s^l}$  sont le nombre de pixels possédant l'étiquette  $m$  ou  $x$  ou  $\mathcal{L}_s^l$ ,
- $n_{m\mathcal{L}_s^l}$  est le nombre de pixels possédant la même étiquette  $m$  dans  $S$  ainsi que l'étiquette  $\mathcal{L}_s^l$  à la position  $s$  dans la segmentation  $S_l$ ,
- et  $s : m \rightarrow x$  est le pixel à la position  $s$  avec l'étiquette  $m$  qui est remplacée par l'étiquette  $x$  dans la segmentation  $S$ .

Contrairement à la métrique du Vol moyen ( $\overline{Vol}$ ), la valeur maximale de  $\Delta\overline{Vol}$  est alors recherchée dans l'algorithme d'optimisation ICM.

Nguyen et al. [Nguyen2018] utilisent également la métrique Vol dans le cadre de la segmentation Multi-Atlas du pelvis, de la rotule et du talus (os du pied ou astragale) pour la fusion des différents Atlas sélectionnés. Leur processus de segmentation Multi-Atlas est réalisé en deux étapes :

- La première permet d'estimer les contours externes de la structure osseuse à segmenter.
- La deuxième permet l'affinement de ce contour, la sélection des segmentations dans l'Atlas étant les plus similaires au contour calculé, et surtout l'estimation de la segmentation en fusionnant les segmentations sélectionnées par l'optimisation de la métrique Vol, comme c'est le cas dans les travaux de Max Mignotte [Mignotte2014a].

### 2.3.3/ LA MÉTHODE BASÉE SUR PRI

La mesure de PRI (*Probabilistic Rand Index*) proposée par [Unnikrishnan2007] est une nouvelle définition de l'indice de Rand (cf. Équation 9) qui permet de comparer une segmentation calculée à un ensemble d'autres segmentations. Elle permet de compter le nombre de paires de pixels ayant les mêmes étiquettes, le tout pondéré par la probabilité de leur apparition.

**Définitions**

Soit  $S^t$  une segmentation et  $\{S_k^g\}_{k \leq L}$  l'ensemble des  $L$  segmentations à fusionner, le calcul de la mesure PRI est le suivant :

$$PRI(S^t, \{S_k^g\}_{k \leq L}) = \frac{1}{\binom{n}{2}} \sum_{\substack{i,j \\ i < j}} [c_{i,j} \cdot p_{i,j} + (1 - c_{i,j})(1 - p_{i,j})] \quad (30)$$

Où

- $c_{i,j}$  est le nombre de paires de pixels  $(p_i, p_j)$  ayant la même étiquette dans  $S^t$ ,
- $p_{i,j}$  est la probabilité moyenne qu'une paire de pixels ait la même étiquette sur l'ensemble des vérités terrain, elle est définie comme suit :

$$p_{i,j} = \frac{1}{L} \sum_{k=0}^L \delta(l_i^{S_k^{gt}}, l_j^{S_k^{gt}}) \quad (31)$$

Où

- $\delta$  est la fonction de Kronecker : il s'agit d'une fonction de deux variables qui est égale à 1 si les valeurs des deux variables sont identiques, et 0 sinon,
- $l_i^{S_k^{gt}}$  et  $l_j^{S_k^{gt}}$  les étiquettes des pixels  $i$  et  $j$  dans la  $k^{\text{ième}}$  segmentation.

De cette manière, la mesure de PRI est simplement la moyenne de l'indice de *Rand* par rapport à l'ensemble des segmentations données.

M. Mignotte utilise alors dans l'article [Mignotte2010] un modèle de fusion basé sur la mesure de PRI, exprimé en tant que modèle de Gibbs basé sur l'énergie, dans lequel l'estimation de PRI est maximisée par l'algorithme ICM (cf. Algorithme 1).

### 2.3.4/ LA MÉTHODE BASÉE SUR GCE

Un modèle développé en 2017 par Khelifi et al. ([Khelifi2017]) utilise le critère d'erreur de la cohérence globale (*GCE*). Afin d'une part, d'éviter les problèmes de cette métrique quand les segmentations à fusionner possèdent un nombre de régions différent et d'autre part, de pouvoir utiliser cette métrique dans un modèle de fusion basée sur l'énergie, l'auteur modifie l'équation de la mesure CGE (cf. Équation 18) par :

$$GCE^*(S^t, S^g) = \frac{1}{2n} \left\{ \sum_{i=1}^n LRE(S^t, S^g, p_i) + \sum_{i=1}^n LRE(S^g, S^t, p_i) \right\} \quad (32)$$

La mesure GCE peut également être utilisée pour comparer une segmentation calculée avec un ensemble d'autres segmentations de la même image. Notons  $\{S_k^g\}_{k \leq L} =$

$\{S_1^g, S_2^g, \dots, S_L^g\}$  l'ensemble des  $L$  segmentations. La mesure GCE moyenne ( $\overline{GCE}$ ) est définie par :

$$\overline{GCE}(S^t, \{S_k^g\}_{k \leq L}) = \frac{1}{L} \sum_{k=1}^L GCE(S^t, S_k^g) \quad (33)$$

De même, la mesure GCE moyenne de la nouvelle équation ( $\overline{GCE}^*$ ) de la mesure GCE (cf. Équation 32) est :

$$\overline{GCE}^*(S^t, \{S_k^g\}_{k \leq L}) = \frac{1}{L} \sum_{k=1}^L GCE^*(S^t, S_k^g) \quad (34)$$

Une stratégie possible pour trouver une segmentation consensus  $\hat{S}$  en fonction d'un ensemble de segmentations est de déterminer la segmentation minimisant  $\overline{GCE}^*$  :

$$\hat{S}_{\overline{GCE}^*} = \arg \min_{S \in S_n} \overline{GCE}^*(S, \{S_k^g\}_{k \leq L}) \quad (35)$$

Cependant, si l'on souhaite obtenir une segmentation consensus contenant un nombre limité de régions, il est indispensable d'ajouter des contraintes à l'Équation 35. Dans cette optique, le terme de *régularisation* est ajouté à l'équation, permettant de fusionner ou de diviser certaines régions, en fonction des différentes régions présentes dans la segmentation consensuelle résultante. Le terme utilisé est le terme de régularisation par région, dérivé de la théorie de l'information [Khelifi2017].

La segmentation consensus est alors calculée en appliquant à l'Équation 35 le terme de régularisation et l'application de l'algorithme ICM (cf. Algorithme 1).

Avec le critère GCE, il est impossible dans l'algorithme de fusion d'utiliser une stratégie d'optimisation par pixel. Une stratégie de fusion plus générale se basant sur les régions des segmentations à fusionner est mise en place. Cette stratégie réduit le temps de calcul mais, comme pour le modèle basé sur la métrique F-Mesure, il est important d'avoir un nombre important de régions dans les segmentations à fusionner.

## 2.4/ LES AUTRES MÉTHODES BASÉES SUR LES DISTANCES ET LES SIMILARITÉS

### 2.4.1/ LA MÉTHODE *Neighbours and Intensity* : NANDI

Une méthode basée sur les informations des pixels voisins dans une image est utilisée dans [Feng2017] pour la fusion de segmentations d'images du cerveau humain. Une seule méthode de segmentation est réalisée, mais appliquée sur une coupe du cerveau humain ainsi que sur cette même coupe préalablement lissée, afin d'obtenir deux segmentations provenant de la même coupe mais contenant des informations potentiellement différentes. La fusion de ces deux segmentations laisse apparaître des pixels en conflit devant alors être résolus. La méthode de fusion présentée utilise les informations de niveaux de gris des pixels voisins dans les images scanner ainsi que les étiquettes

dans les segmentations calculées. Cette méthode, que nous avons nommée *NandI* (pour *Neighbours and Intensity*), calcule un **indice de similarité** entre deux pixels d'une même image scanner.

### Définition

Soit deux pixels  $p$  et  $q$  d'une même image scanner, le calcul de l'indice de similarité de la méthode *NandI* est :

$$Sim(p, q) = \exp\left(-\frac{d^2(p, q)}{2\alpha^2} - \frac{dif^2(p, q)}{2\beta^2}\right) \quad (36)$$

Où

- $d(p, q) = \|p - q\|$  est la distance euclidienne entre  $p$  et  $q$ ,
- $dif(p, q) = |I(p) - I(q)|$  est la différence d'intensité des valeurs des pixels entre eux,
- $\alpha$  et  $\beta$  sont les deux poids.

Un ensemble de voisins cohérents doit ensuite être déterminé pour chaque pixel en conflit  $p$  afin de déterminer son étiquette finale. Soit  $E_p^r$  l'ensemble d'un **voisinage effectif** de  $p$ . Un pixel  $q$  doit satisfaire les trois conditions suivantes pour devenir un voisin effectif de  $p$  :

1.  $d(p, q) < r$  : la distance euclidienne entre  $p$  et  $q$  doit être inférieure au rayon  $r$ .
2.  $l^q \in S = l^q \in S'$  : soient deux segmentations de la même image  $S$  et  $S'$  calculées de manières différentes, l'étiquette de  $q$  doit être la même dans les deux segmentations, afin de ne pas avoir de conflit sur un pixel.
3.  $l^q \in \mathcal{E}$  : l'étiquette de  $q$  appartient à l'ensemble des étiquettes possibles.

La nouvelle étiquette de  $p$  est alors obtenue par la formule suivante :

$$l^p = \begin{cases} l_a & \text{si } \sum_{q \in E_p^r, l^q = l_a} Sim(p, q) > \sum_{q \in E_p^r, l^q = l_b} Sim(p, q), \\ l_b & \text{sinon} \end{cases} \quad (37)$$

Où

- Si la somme totale de la similarité (cf. Équation 36) des pixels effectifs avec l'étiquette de la classe  $l_a$  est supérieure à celle des pixels effectifs avec l'étiquette de la classe  $l_b$ , alors  $l^p = l_a$ , sinon  $l^p = l_b$ .
- Et pour chaque pixel en conflits de l'image, la Formule 37 est alors appliquée.

Comme nous l'avons expliqué, cette méthode permet de gérer seulement les pixels en conflit, par rapports aux méthodes utilisant des métriques. De plus, elle est appliquée

sur des coupes transversales du cerveau humain. Cependant, elle ne prend pas en considération les informations de profondeur avec les coupes adjacentes, potentiellement similaires et pouvant permettre d'ajouter plus d'informations dans le processus de fusion.

#### 2.4.2/ LA MÉTHODE BASÉE SUR K-MOYENNE

La méthode des k-moyennes [Macqueen1967] est une méthode de **clustering** permettant de regrouper des individus ayant des caractéristiques similaires, en se basant sur des mesures de similarité ou de distance. Cette méthode est utilisée dans divers domaines ainsi que dans la segmentation d'images [Bandhyopadhyay2013] [Islam2013].

Son principe est le suivant :

1. Un ensemble de régions (k) appelé *centre de cluster* est choisi arbitrairement ou aléatoirement.
2. Tant que les centres de clusters sont modifiés :
  - La distance entre chaque pixel considéré et chaque centre de cluster est calculée dans l'image. Le pixel appartient alors au centre de cluster le plus proche.
  - Dans la grande majorité des cas, la distance est calculée en prenant la distance euclidienne. Il est cependant à noter que la distance de Bhattacharya est également utilisée pour comparer 2 histogrammes.
  - Chaque centre de cluster est ensuite recalculé en prenant en compte les nouveaux pixels intégrés. La valeur d'un centre de cluster peut être tout simplement la moyenne des valeurs de tous les pixels appartenant au cluster.

Une méthode de fusion de segmentations est basée sur l'algorithme des k-moyennes [Mignotte2008], en récupérant les histogrammes locaux de chaque pixel des segmentations à fusionner. Le type des images qui sont fusionnées est le même que celles utilisées pour les méthodes basées sur les métriques. Il s'agit de multiples segmentations de scène diverses, contenant un nombre varié de région. R. Harrabi et al., dans [Harrabi2012], suivent la même idée de fusion, mais en utilisant l'algorithme des k-moyennes floues sur des images de cellules de cancer du sein. Cette méthode peut être adaptée pour gérer les conflits existants lors de la fusion de différentes segmentations complémentaires. En effet, les segmentations déjà réalisées (sans prendre en compte les conflits) peuvent être utilisées pour définir les centres de clusters. Il est alors possible de déterminer la distance de chaque pixel en conflit avec chacun des clusters et lui affecter l'étiquette du centre de cluster le plus proche.

## 2.5/ LES MÉTHODES MULTICRITÈRES

Lors de la fusion de segmentations basée sur un critère unique, le résultat est naturellement orienté vers une famille particulière de solutions possibles, en accord avec le critère choisi. Cela peut donc réduire les performances d'un système de fusion. Pour pallier

cet inconvénient, il est possible de combiner différents critères complémentaires via une approche multi-objectifs.

### 2.5.1/ LA MÉTHODE EFA-BMFM

Un modèle de fusion multicritères pondéré par une mesure basée sur l'entropie, EFA-BMFM (*Multi-criteria Fusion Model Based on the Entropy-Weighted Formula Approach*), est introduit en 2017 [Khelifi2017a]. Ce modèle combine et optimise simultanément deux critères différents et complémentaires, à savoir les critères Vol et F-Mesure, basés sur les régions et les contours des segmentations. L'optimisation multi-objectifs de la fusion se base alors sur l'approche de la **formule pondérée WFA** (*Weighted Formula Approach*) dans laquelle chacun des critères est pondéré, puis ces derniers sont additionnés. D'après les Formules 24 et 27 de la mesure moyenne de F-Mesure et de Vol, la fonction multi-objective moyenne entre une segmentation  $S$  et un ensemble de segmentations  $\{S_k\}_{k \leq L}$  est définie par :

$$\overline{MO}(S, \{S_k\}_{k \leq L}) = w_{Vol} \overline{Vol}(S, \{S_k\}_{k \leq L}) + \frac{w_{F_\alpha}}{\overline{F}_\alpha(S, \{S_k\}_{k \leq L})} \quad (38)$$

Où :

- $w_{Vol}$  et  $w_{F_\alpha}$  sont les pondérations des deux métriques.
- Ces pondérations sont alors estimées grâce à la valeur d'entropie qui permet de mesurer la quantité d'informations contenue dans les segmentations par rapport à chaque critère. L'entropie  $e_{Vol}$  en fonction du critère Vol et l'entropie  $e_{F_\alpha}$  en fonction du critère F-Mesure sont définies par :

$$e_{Vol} = -\frac{1}{\log(L)} \sum_{i=1}^L \left\{ \frac{\overline{Vol}(S_i, \{S_k\}_{k \leq L})}{S_{\overline{Vol}}} \log \frac{\overline{Vol}(S_i, \{S_k\}_{k \leq L})}{S_{\overline{Vol}}} \right\} \quad (39)$$

$$e_{F_\alpha} = -\frac{1}{\log(L)} \sum_{i=1}^L \left\{ \frac{\overline{F}_\alpha(S_i, \{S_k\}_{k \leq L})}{S_{\overline{F}_\alpha}} \log \frac{\overline{F}_\alpha(S_i, \{S_k\}_{k \leq L})}{S_{\overline{F}_\alpha}} \right\} \quad (40)$$

Où :

$$S_{\overline{Vol}} = \sum_{j=1}^L \overline{Vol}(S_j, \{S_k\}_{k \leq L}) \quad (41)$$

et

$$S_{\overline{F}_\alpha} = \sum_{j=1}^L \overline{F}_\alpha(S_j, \{S_k\}_{k \leq L}) \quad (42)$$

et finalement, le poids de chaque critère est défini par :

$$w_{Vol} = \frac{1 - e_{Vol}}{(1 - e_{Vol}) + (1 - e_{F_\alpha})} \quad (43)$$

$$w_{F_\alpha} = \frac{1 - e_{F_\alpha}}{(1 - e_{Vol}) + (1 - e_{F_\alpha})} \quad (44)$$

Comme pour les modèles de fusion par les métriques Vol et F-Mesure, ce modèle est optimisé par l'algorithme ICM dans le but de minimiser  $\overline{MO}$ . Cette minimisation se fait également en parcourant chaque superpixel des segmentations, et non chaque pixel pour un gain de temps et de performance.

### 2.5.2/ LA MÉTHODE GCE+F-MESURE

Khelifi et al., dans [Khelifi2017b], proposent également une autre méthode de fusion combinée, appelée *MOBFM* pour *Multi-objective Optimization Based-Fusion Model*. Cette méthode associe ainsi le critère GCE (cf. Équation 18) au critère F-Mesure (cf. Équation 7), tout en utilisant également l'algorithme ICM pour l'optimisation par itération de la segmentation fusionnée. La combinaison de ces deux méthodes se fait simplement par l'intersection de leurs résultats. Une pondération pour chaque segmentation est également ajoutée au critère F-mesure afin d'améliorer la précision des résultats. Enfin, un terme de régularisation est également ajouté à la métrique GCE.

**Synthèse du chapitre****SYNTHÈSE**

Cette section, nous a permis de montrer que si les méthodes de fusion par vote sont intuitives, elles sont très vite limitées dans le cas de fusion de segmentations plus complexes. Nous avons également pu déterminer que l'utilisation de métrique pour la fusion impose de réaliser un algorithme d'optimisation du résultat de la métrique, comme, par exemple les algorithmes appelés ICM.

Sont également présentées dans la littérature des méthodes de fusion différentes qui sont utilisées dans le cadre de la segmentation Multi-Atlas, ainsi que d'autres méthodes utilisant des calculs de distances et de similarités au niveau des intensités des pixels. Pour finir, Afin de combiner différentes méthodes de fusion, une première approche est d'additionner les critères pondérés entre eux afin d'obtenir une nouvelle fonction à optimiser. La seconde approche consiste à réaliser l'intersection des résultats de deux métriques.

Cependant, à part pour les méthodes dites intuitives qui peuvent être utilisées dans de très nombreux domaines, les autres méthodes présentées sont utilisées dans le cadre de la fusion de plusieurs segmentations, comportant les mêmes éléments mais segmentés différemment, et provenant d'une même image source. Tous les pixels de l'image sont alors traités et fusionnés (sauf dans le cas de la méthode Nandl). Ces méthodes n'ont jamais été adaptées et testées pour la fusion de segmentations dites complémentaires d'une même image source, c'est-à-dire comportant des éléments différents, et pour la gestion de seulement certains pixels se trouvant en conflits. De plus, dans le cas d'images médicales avec l'acquisition de plusieurs coupes 2D, il serait intéressant de récupérer les informations de profondeur avec les coupes adjacentes lors du processus de fusion, ce qui n'a pas encore été réalisé.

Dans le cas précis de notre étude et du projet SAIAD, comme nous voulons résoudre certains pixels en conflits, les méthodes agissant au niveau des pixels sont intéressantes à utiliser. Ainsi comme nous le verrons dans la suite de ce document, nous nous sommes orientés vers l'utilisation de métriques au niveau pixel, comme la métrique Vol, avec l'utilisation d'algorithmes itératifs de type ICM, ainsi que l'utilisation des informations de distances et de similarités entre pixels comme la méthode nommée Nandl.

Dans le chapitre suivant, nous nous intéresserons à l'utilisation des réseaux de neurones dans le cadre de la fusion de segmentations.

Nous présenterons le fonctionnement général des réseaux de neurones, ainsi que des réseaux profonds. Nous nous attarderons sur les réseaux de neurones convolutifs et les différentes couches qui sont utilisées dans ces réseaux principalement appliqués au traitement d'images. Un état de l'art sur différents réseaux pour la fusion de segmentations sera également présenté.

## LA FUSION DE SEGMENTATIONS PAR RÉSEAUX DE NEURONES

Le *machine learning* ou l'apprentissage automatique est l'étude d'algorithme capable d'apprendre et d'évoluer à partir de données. Les réseaux de neurones et leur évolution en paradigme appelée *Deep Learning* sont des méthodes d'interpolation basées sur l'apprentissage. Depuis leur création, et encore plus depuis l'avènement des GPU (*Graphical Processing Unit*), ils font l'objet d'une attention particulière et d'un engouement du monde scientifique car ils font maintenant partie des méthodes les plus performantes dans les domaines de l'interpolation, de la classification et de la segmentation d'images. Aujourd'hui, leur performance et certainement le mystère de leur fonctionnement (effet "boîtes noires") en font des outils de plus en plus populaires. Utilisés dans de nombreux domaines, les réseaux de neurones dans le contexte de l'imagerie médicale constituent également un point important dans la littérature. En témoigne l'étude réalisée par Litjens et al. en 2017 sur l'ensemble des réseaux de neurones profonds utilisés dans le traitement d'images médicales [Litjens2017].

Les CNNs (*Convolutional Neural Networks*) sont des réseaux de neurones spécialement conçus pour le traitement d'images. Grâce à leur grande capacité à extraire les caractéristiques des images à l'aide d'opérations linéaires, ils sont un véritable atout pour la classification, la détection d'objets ainsi que la segmentation d'images. Cependant, les réseaux de neurones pour la fusion de multiples segmentations sont moins répandus dans la littérature, mais ils peuvent être vus comme des CNNs particuliers effectuant des segmentations. En effet, le principe des réseaux de fusion de plusieurs segmentations est de fournir en sortie une segmentation supplémentaire à partir des segmentations données en entrée, qui correspondra alors à la segmentation consensus.

Dans un premier temps, les notions générales des réseaux de neurones ainsi que des CNNs seront étudiées dans ce chapitre. Puis, les différents réseaux de neurones utilisés dans la littérature pour la fusion de segmentations seront présentés.

### 3.1/ LE PRINCIPE DES RÉSEAUX DE NEURONES

#### 3.1.1/ LES NOTIONS GÉNÉRALES

##### 3.1.1.1/ LE PERCEPTRON

Le neurone biologique est une cellule du système nerveux. Il reçoit, traite et envoie des informations sous forme de signaux électriques à ses neurones voisins. Les réseaux de neurones artificiels sont inspirés des neurones biologiques et ont pour objectif d'imiter la stimulation des neurones du cerveau humain. Le perceptron est le réseau de neurones artificiels le plus simple, inventé en 1957 par Franck Rosenblatt [Rosenblatt1958]. Il est représenté comme un système à  $N$  entrées et une sortie, comme illustré par la Figure 3.1. Chaque entrée possède un poids  $w$ . La somme des entrées pondérées est réalisée avant d'envoyer le résultat dans une **fonction d'activation**.

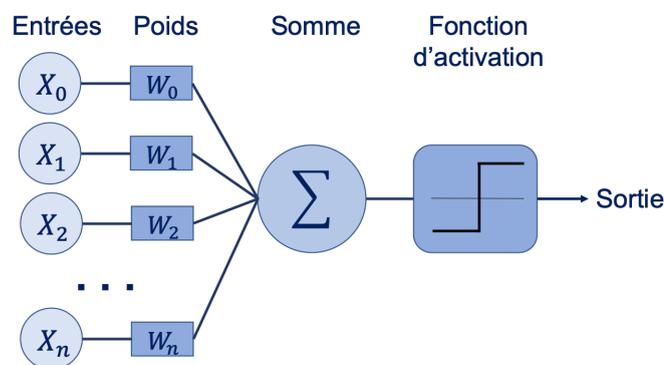


FIGURE 3.1 – Le perceptron.

#### Définitions

Soit un perceptron, le résultat est calculé par la fonction d'activation  $f$  sur la somme des entrées pondérées  $\Sigma$  telle que :

$$f(s) = f(\Sigma + b) \quad \text{avec} \quad \Sigma = \sum_{i=1}^m x_i w_i \quad (45)$$

Où :

- $b$  est le biais,
- $x_i$  est le neurone  $i$ ,
- $w_i$  le poids du neurone  $i$ ,
- et  $m$  le nombre de neurones.

Le biais de la fonction d'activation correspond au seuil d'activation de la fonction. Lorsque la somme pondérée dépasse ce seuil, la fonction renvoie la valeur 0 ou 1. De plus, elle permet d'appliquer une non-linéarité en sortie du neurone.

#### Définitions

Différentes fonctions d'activation peuvent être appliquées :

- La fonction Threshold :  $\sigma(x) = \begin{cases} 1 & \text{if } \sum > \text{threshold} \\ 0 & \text{else} \end{cases}$
- La fonction Sigmoid :  $\sigma(x) = \frac{1}{1+e^{-x}}$
- La fonction Tanh :  $\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- La fonction ReLU :  $\sigma(x) = \max(0, x)$

La fonction majoritairement utilisée est la fonction **ReLU** car elle est très performante dans les réseaux profonds. Les autres fonctions peuvent engendrer des problèmes d'apprentissage, notamment l'accentuation du problème de la disparition du gradient (également appelé *the vanishing gradient problem*).

#### 3.1.1.2/ LE PERCEPTRON MULTICOUCHE

Le perceptron multicouche a été inventé dans les années 1980. Il s'agit d'une succession de couches contenant plusieurs neurones, chaque neurone d'une couche étant relié à la totalité des neurones des couches adjacentes (cf. Figure 3.2). Ainsi, le terme de couche *Fully Connected* ou FC est utilisé pour désigner les couches de ce type. Un perceptron multicouche possède une couche d'entrée, une couche de sortie ainsi qu'une ou plusieurs couches cachées. Chaque résultat des fonctions d'activations des neurones à la couche  $N-1$  sera alors transmis à tous les neurones de la couche  $N$ . Chaque neurone de la couche  $N$  effectue la somme de toutes leurs entrées pondérées, transmet ce résultat à la fonction d'activation et envoie la sortie à tous les neurones de la couche  $N+1$ , et ainsi de suite jusqu'à la dernière couche.

#### 3.1.1.3/ L'APPRENTISSAGE PROFOND

L'apprentissage profond ou *Deep Learning* ne se concrétise qu'au début des années 2010. Cet apprentissage fait partie du domaine du *Machine Learning*. Le *Deep Learning* consiste à réaliser un réseau de neurones profond, sur le même modèle que les perceptrons multicouches, mais avec de nombreuses couches cachées. Ce réseau est alors capable d'extraire des abstractions complexes de haut niveau des données lui étant présentées. Nous nous intéressons ici à l'utilisation du *Deep Learning* dans le cadre d'un apprentissage supervisé, c'est-à-dire que le réseau a besoin d'une base d'apprentissage pour s'entraîner et ainsi trouver une solution optimale.

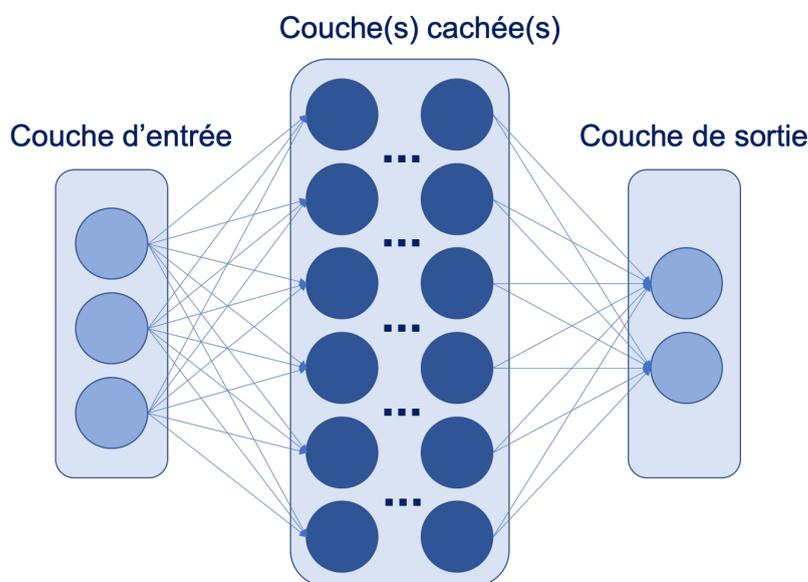


FIGURE 3.2 – Le perceptron multicouche.

**Principe de l'apprentissage** Les réseaux de neurones ont besoin dans un premier temps d'une phase d'apprentissage durant laquelle les **poïds** et les **biais** des neurones sont ajustés jusqu'à ce que la sortie prédite soit similaire à la sortie désirée. La Figure 3.3 montre le principe général de l'apprentissage, dans lequel les entrées sont d'abord transmises au système, c'est-à-dire au réseau de neurones. Le système calcule alors une sortie qui est comparée à la sortie désirée grâce à une fonction de coût. En fonction de la différence entre la sortie calculée et la sortie désirée, les paramètres seront ajustés avant de repasser dans le réseau pour une nouvelle prédiction, afin de minimiser l'erreur mesurée en sortie. Il s'agit alors d'un processus itératif minimisant l'erreur mesurée afin d'obtenir la sortie optimale.

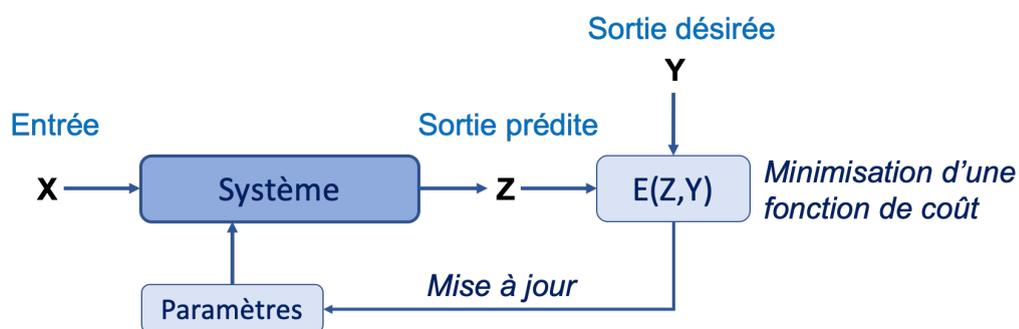


FIGURE 3.3 – Le principe général de l'apprentissage.

**La rétropropagation** L'ajustement des paramètres se fait généralement grâce à un algorithme de descente du gradient stochastique, souvent abrégé par SGD (*Stochastic Gradient Descent*). Il s'agit d'un algorithme d'optimisation du gradient permettant de converger de manière itérative vers une configuration optimisée. La rétropropagation du gradient (*backpropagation*) s'appuie sur l'algorithme SGD dans le cas de réseaux plus

profonds et est utilisée pour la première fois en 1988 d'une part par Rumelhard et al. [Rumelhart1988] et d'autre part par Yann Lecun [Lecun1988].

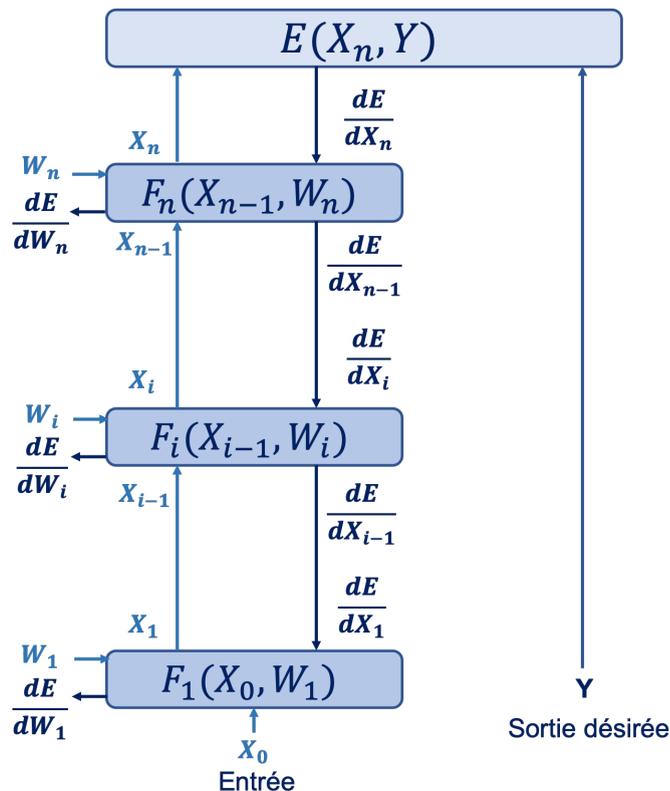


FIGURE 3.4 – Le principe de rétropropagation du gradient.

Comme nous pouvons le voir sur la Figure 3.4, le principe d'apprentissage par rétropropagation du gradient est le suivant :

- Les paramètres de chaque couche du système sont dans un premier temps initialisés de façon aléatoire à travers l'ensemble du réseau.
- L'entrée est ensuite propagée le long du réseau (on parle alors de propagation vers l'avant ou *forward propagation*).
- Avec la sortie prédite  $X_n$  et la sortie désirée  $Y$ , l'indicateur de performance avec la fonction de coût est alors calculé.
- L'algorithme de rétropropagation permet ensuite à l'information du coût de retourner en amont du réseau (on parle alors de propagation vers l'arrière ou *backward propagation*), afin de calculer le gradient.

Pour cela, le gradient de cette fonction de coût par rapport à la sortie  $X_n$  est calculé ( $\frac{dE}{dX_n}$ ).

- On peut alors remonter à l'entrée du réseau en calculant la dérivée de chaque couche. Pendant la remontée, afin de permettre l'ajustement des paramètres de chaque couche, la dérivée de la fonction de coût par rapport aux paramètres  $W_i$  de la couche  $i$  est calculée.

- Les paramètres sont alors incrémentés du gradient calculé.
- Arrivé à l'entrée du réseau, ce processus incluant le calcul et la mesure d'écart entre sortie attendue et sortie interpolée est réitéré. Cet écart se réduit alors grâce aux paramètres précédemment optimisés : le but étant d'itérer jusqu'à ce que l'indicateur de performance soit satisfaisant.

### 3.1.2/ LES RÉSEAUX DE NEURONES CONVOLUTIFS

Un **réseau de neurones convolutif** ou **CNN** pour *Convolutional Neural Network* [Lecun1989] est un type de réseau de neurones très utilisé en traitement d'images, comme, par exemple pour la classification, la détection et la segmentation d'objets. Les multiplications matricielles classiques dans les différentes couches du réseau sont remplacées par des opérations linéaires appelées convolutions dans un réseau de neurones convolutif. De plus, chaque neurone n'est connecté qu'à une partie des données d'entrée, et c'est l'ensemble des neurones dans une couche qui permettra de traiter les données en entier. Un CNN possède différents types de couches, ayant chacun un rôle particulier et permettant de modifier et de moduler l'information afin de parvenir au résultat final.

#### 3.1.2.1/ LA COUCHE DE CONVOLUTION

L'une des couches les plus importantes dans un CNN est la couche de convolution. Une convolution en traitement d'images permet d'appliquer un filtre à l'image en entrée dans le but de la modifier. Le filtre, également appelé noyau de convolution (ou *kernel*), généralement de taille  $3 \times 3$  ou  $5 \times 5$ , est superposé à l'image en entrée. La sortie de la convolution est alors la multiplication des valeurs des pixels de l'image avec les valeurs associées du filtre, puis l'addition de toutes ces multiplications (cf. Figure 3.5). Le résultat est alors la valeur du pixel de l'image convoluée se trouvant au centre du filtre. Le filtre est ensuite déplacé sur toute l'image afin de calculer les valeurs de tous les pixels. La convolution possède également d'autres paramètres, comme le pas (ou *stride*) qui représente la distance (en pixels) à appliquer pour chaque déplacement du filtre le long de l'image, et la valeur de remplissage (ou *padding*) qui consiste à ajouter un certain nombre de pixels de valeurs nulles autour de l'image. Ces paramètres peuvent modifier les dimensions de l'image, et l'ajout d'un remplissage, avec un pas de 1, permet alors de pallier la réduction des dimensions de l'image induite par l'opération de convolution.

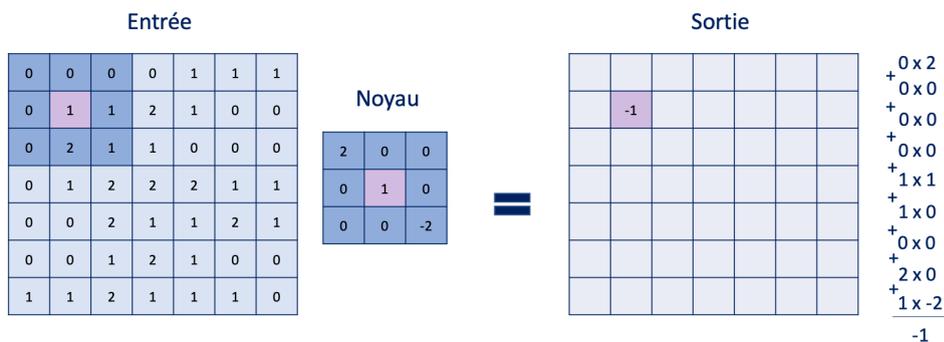


FIGURE 3.5 – Le principe de la convolution.

Au sein d'un CNN, une couche de convolution permet d'appliquer plusieurs filtres à l'image, permettant ainsi d'obtenir autant d'images convoluées en sortie que de filtres appliqués. Ces images convoluées sont appelées **cartes de caractéristiques** ou *features maps*. Lors de l'entraînement, ce sont les valeurs de ces filtres qui sont ajustées afin d'obtenir au mieux la sortie désirée (les valeurs des filtres étant les poids des neurones).

### 3.1.2.2/ LA COUCHE DE CORRECTION

Cette couche possède un rôle complémentaire à la couche de convolution, car il s'agit de la fonction d'activation à appliquer à la convolution. Différentes fonctions peuvent être appliquées, comme la fonction *threshold* ou *sigmoid*, mais c'est principalement la fonction ReLU qui est la plus couramment utilisée compte tenu de son efficacité [Glorot2011] (les définitions de ces fonctions d'activations ont été présentées précédemment en Section 3.1.1.1).

### 3.1.2.3/ LA COUCHE DE DÉCONVOLUTION

La couche de déconvolution peut être vue comme l'opération inverse de la convolution et comme une étape de sur-échantillonnage. Cependant, le terme de déconvolution est un abus de langage et il est plus juste de parler de convolution transposée. Une déconvolution possède également plusieurs paramètres comme le filtre, le pas et le remplissage.

### 3.1.2.4/ LA COUCHE DE *pooling*

Le *pooling* est un type de sous-échantillonnage de l'image. L'image d'entrée est découpée en plusieurs régions et chaque région sera désignée par une seule valeur. Il existe différents types de *pooling* comme le *max pooling* ([Zhou1988]), illustré par la Figure 3.6, qui consiste à choisir la valeur maximale de tous les pixels d'une région ou encore le *average pooling* qui calcule la moyenne de tous les pixels d'une région. Notons que le *pooling* possède également plusieurs paramètres comme la taille du filtre qui correspond à la taille des régions souhaitées et le pas appliqué qui est généralement de la taille du filtre afin que chaque pixel ne puisse pas se trouver dans plusieurs régions en même temps.

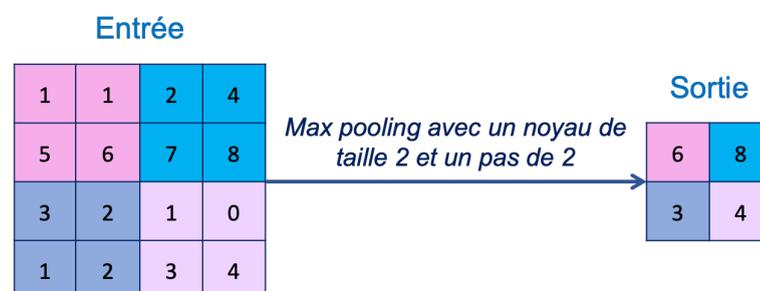


FIGURE 3.6 – Le principe du *max pooling*.

### 3.1.2.5/ LA COUCHE DE PERTE ET LA COUCHE *SoftMax*

Principalement, la dernière couche du CNN est une couche de perte permettant de calculer l'erreur obtenue lors de l'entraînement selon une fonction de coût. Cette couche permet donc d'orienter l'apprentissage afin de diminuer l'erreur. Différentes fonctions de coût peuvent être appliquées comme la fonction *multinomial logistic loss* ou la *cross entropy*. Enfin, la couche *SoftMax* est également couramment utilisée en couplage avec la couche de perte, dans le but de normaliser la sortie. Elle transforme les scores calculés par le réseau en probabilités avec des valeurs comprises entre 0 et 1. Dans le cas de la segmentation d'images, chaque pixel aura alors une probabilité d'appartenir à chaque classe, la somme des probabilités étant égale à 1. Afin de calculer le résultat final à travers les probabilités, la fonction  $\arg \max$  peut simplement être utilisée pour déterminer l'étiquette possédant la plus forte probabilité pour chaque pixel.

### 3.1.2.6/ L'ARCHITECTURE VGG16

Compte tenu de ses performances, une grande partie des CNNs pour la segmentation se base sur l'architecture **VGG16** [Simonyan2014] (VGG pour *Visual Geometry Group*, et 16 pour les 13 couches de convolution ainsi que les 3 couches de *Fully Connected*). Cette architecture a gagné la deuxième place du concours ILSVRC 2014<sup>1</sup> pour la classification d'images, derrière GoogLeNet [Szegedy2015], avec néanmoins un nombre de couches relativement plus faible que ce dernier (et par conséquent, avec un nombre de paramètres moins important) : au total, l'architecture **VGG16** est composée de 13 couches de convolution avec un noyau de taille 3, 5 couches de *pooling* avec un noyau de taille 2, 3 couches *Fully Connected* (FC) et une couche *Softmax*. La Figure 3.7 représente cette architecture avec également le nombre de cartes de caractéristiques en sortie de chaque couche. À la fin du réseau, les 1000 sorties correspondent au nombre d'éléments pouvant être classifiés dans le concours ILSVRC 2014.

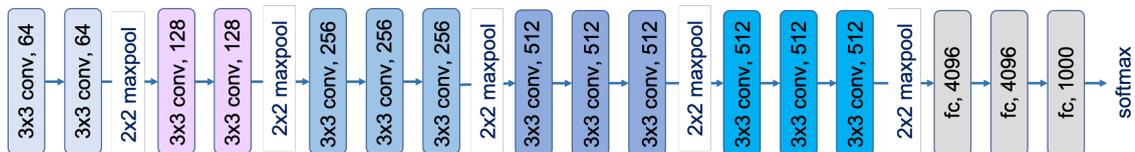


FIGURE 3.7 – L'architecture VGG16.

### 3.1.3/ LES CNNs POUR LA SEGMENTATION

Les premiers CNNs pour la segmentation d'images apparaissent en 2015 avec la méthode FCN, *Fully Convolutional Networks* [Long2015]. Les couches FC d'un CNN pour la classification permettent de passer de plusieurs cartes de caractéristiques à des données numériques sous forme de vecteurs afin de sortir pour une image, le pourcentage d'appartenance à une classe. Dans le cas d'un CNN pour la segmentation, ces dernières couches FC du réseau sont alors remplacées par des couches de convolution,

1. ILSVRC est un concours de traitement d'images dans la détection d'objets et la classification d'images, fournissant une base importante de données d'images.

suivies de l'ajout d'un sur-échantillonnage avec une couche de déconvolution, afin d'obtenir en sortie une segmentation de la même taille que l'image en entrée. La plupart de ces réseaux de segmentations ont une base se rapprochant du réseau VGG16 qui est la référence compte tenu de son efficacité par rapport à d'autres architectures.

Long et al. proposent alors trois réseaux obtenant des segmentations de plus en plus précises : FCN-32s, FCN-16s et FCN-8s. Leur architecture est présentée dans la Figure 3.8. Celle-ci est basée sur le réseau VGG16, mais par mesure de clarté seules les cinq couches de pooling sont représentées. Pour obtenir le réseau FCN-32s, une simple déconvolution est ajoutée à la fin des couches de convolution remplaçant les couches FC. La déconvolution est effectuée avec un pas (*stride*) de 32 et un noyau de 64, d'où son nom : FCN-32s. Afin d'obtenir des contours plus précis, les dernières couches peuvent alors être combinées avec des couches utilisant un pas plus petit. Le réseau FCN-16s est alors obtenu en divisant le flux en deux au niveau du *pool4*. Une couche de convolution avec un noyau de taille  $1 \times 1$  est alors ajoutée dans le second flux, tandis qu'une plus petite déconvolution avec un pas de 2 est ajoutée à la place de la déconvolution 32 dans le premier flux. Le tout est alors agrégé et une déconvolution avec un pas de 16 et un noyau de 32 est ajoutée, ce qui nous donne le réseau FCN-16s. Enfin, un nouveau flux peut encore être ajouté à ce réseau, dans lequel une autre convolution avec un noyau de taille 1 est ajoutée à la sortie du *pool3*, tandis qu'une déconvolution avec un pas de 2 est ajoutée après l'agrégation des précédents résultats. Ces deux résultats sont également combinés et une dernière déconvolution avec un pas de 8 et un noyau de 16 est ajoutée, ce qui nous donne le réseau le plus précis des 3 : FCN-8s.

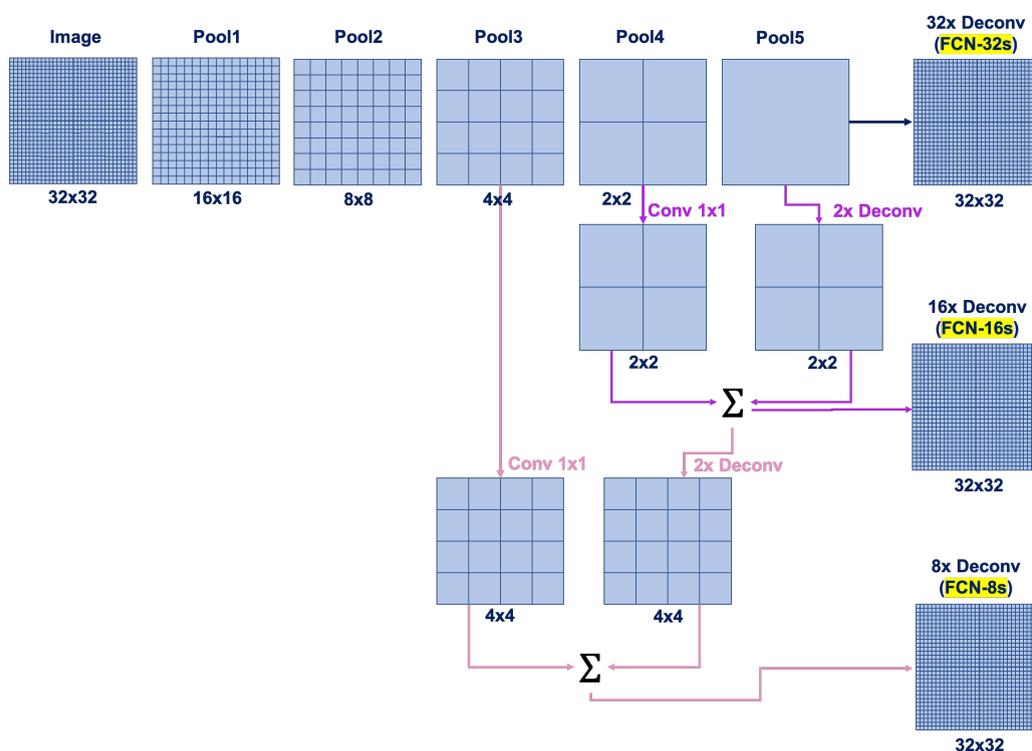


FIGURE 3.8 – L'architecture du réseau de segmentations FCN-32s, FCN-16s et FCN-8s.

D'autres réseaux sont ensuite apparus, avec un nombre de couches plus important, et accentuant le principe *encoder-decoder*. La partie *encoder* permet d'extraire les

caractéristiques, comme c'est le cas de l'architecture VGG16, et la partie *decoder* suréchantillonne les caractéristiques obtenues de la première partie. FCN-8s possède une partie *decoder* en une couche de déconvolution, mais il est possible d'obtenir une partie *decoder* avec autant de couches permettant le suréchantillonnage que de couches dans la partie *encoder*, comme avec l'architecture DeconvNet [Noh2015].

### 3.2/ DIFFÉRENTES APPLICATIONS DES RÉSEAUX DE NEURONES À LA FUSION DE SEGMENTATIONS

La fusion de multiples segmentations par réseau de neurones s'avère être un domaine encore peu exploré par rapport au processus de fusion avec des méthodes dites plus classiques. Cependant, plusieurs méthodes appliquent la fusion dans le contexte de la **détection de saillances** ou *saliency map*. La détection de saillances vise à mettre en évidence et à segmenter les objets ou régions les plus importants ou visuellement distinctifs dans une image en extrayant ses caractéristiques discriminantes, puis à calculer leur importance dans l'image. Concrètement, la saillance est un aspect du stimulus qui fait qu'un objet, une personne ou un pixel se démarque du reste dans une scène ou une image. Un objet est dit saillant si nos yeux y sont facilement attirés. Il existe deux types de saillances :

- la saillance *involontaire* dans laquelle un objet dans une scène attire involontairement notre attention (cela est souvent dû à ses propriétés stimulantes) et
- la saillance *volontaire* qui est le résultat d'une recherche volontaire d'un élément dans une scène.

La saillance involontaire peut être utilisée pour déterminer les régions saillantes d'une image à partir de ses caractéristiques photométriques ou colorimétriques (contrastes de luminosité, de couleurs, contours d'un objet, ...). Cette approche imite la perception visuelle de l'homme. Mais, elle présente deux inconvénients : elle ne peut pas incorporer d'informations sémantiques à partir d'une image et elle n'est pas capable d'estimer la saillance d'un objet car elle se base uniquement sur les pixels d'une image pour déterminer ses régions saillantes.

Pour déterminer la saillance des objets dans une image, un processus de segmentation est nécessaire. Cette segmentation des objets saillants dans les images est appelée détection de saillances ou prédiction de saillances. La détection de saillances d'une image est également un problème de segmentation d'images car il s'agit d'étiqueter également les pixels d'une image.

Le résultat d'une détection de saillances est appelé carte de saillances et contient la segmentation des objets saillants dans une image. Un certain nombre de méthodes de détection de saillances incorpore une fusion de plusieurs types de cartes de saillances afin d'améliorer le résultat de la segmentation dans la carte de saillances finale. Cette fusion de plusieurs cartes de saillances est la partie qui nous intéresse dans le cadre de nos recherches. Ensuite, d'autres algorithmes sont appliqués aux images générées précédemment pour calculer leurs scores de saillance et mesurer l'importance des éléments dans l'image.

La détection de saillances par réseau de neurones et plus particulièrement par réseau de neurones convolutifs est utilisée en raison de sa capacité à extraire des caractéristiques de haut niveau et sa capacité à imiter le mécanisme du cerveau humain lors du traitement des informations visuelles. Il existe deux approches de détection de saillances :

- l'approche au niveau pixels, dans laquelle les caractéristiques de chaque pixel sont extraites,
- et l'approche au niveau régions, dans laquelle l'image est d'abord segmentée en régions, et les caractéristiques de chaque région sont ensuite extraites.

Plusieurs méthodes de détection de saillances qui incorporent la fusion de segmentations sont présentées par la suite. Nous nous intéressons plus particulièrement à leurs méthodes de fusion et non aux méthodes de segmentations utilisées, ni aux méthodes de quantification du score de saillance. Pour plus de clarté, l'architecture de chaque réseau pour la fusion est décrite avec les couches de convolution  $conv(N, K)$ , les couches de pooling  $pool(T, K)$  et les couches FC (Fully Connected)  $fc(N)$ , dans lesquelles  $N$  représente le nombre de cartes de caractéristiques désiré en sortie de la couche,  $K$  la taille du noyau et  $T$  le type de pooling réalisé.

### 3.2.1/ LA FUSION PAR UNE COUCHE DE CONVOLUTION

Li et al. [Li2016] proposent pour la détection de saillances, en 2016, un réseau utilisant les approches au niveau pixels et au niveau régions (cf. Figure 3.9).

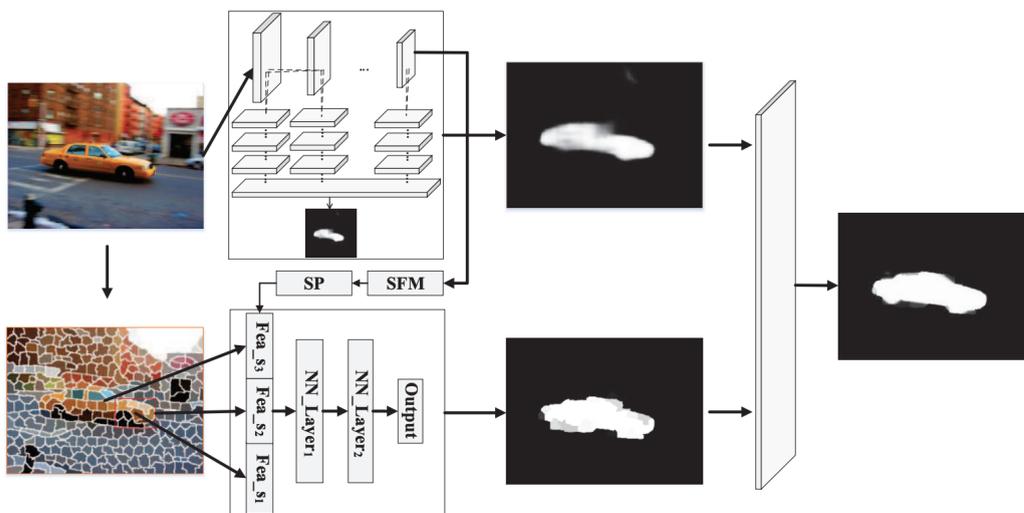


FIGURE 3.9 – Le réseau pour la détection de saillances avec une couche de fusion pour la concaténation des résultats des deux sorties [Li2016].

Un premier CNN est utilisé sur l'image d'entrée pour obtenir les informations au niveau pixels, permettant ainsi d'obtenir une carte de saillances avec une exactitude au niveau pixels. Un second CNN est également utilisé en parallèle sur l'image d'entrée qui a été préalablement segmentée en superpixels : une segmentation en superpixels permet une

segmentation au niveau régions, dans laquelle les pixels proches ayant les mêmes caractéristiques sont regroupés en une région. Ce deuxième réseau permet alors d'obtenir une carte de saillances avec une exactitude au niveau régions. Ces deux cartes sont alors fusionnées à l'aide d'une couche de convolution avec un noyau  $1 \times 1$  afin de ne pas réduire la taille de l'image :  $conv1(1, 1)$ .

Une autre fusion de plusieurs cartes de caractéristiques est également réalisée dans le premier réseau dans l'approche au niveau pixels. Pour rappel, une carte de caractéristiques est une image produite en sortie d'une couche de convolution. Ce réseau profond se base sur l'architecture VGG16 dans laquelle les couches FC sont modifiées en couches de convolution avec un noyau  $1 \times 1$ . Le réseau est également modifié pour être multi-échelles : à chacune des quatre premières couches de pooling du VGG16 sont reliées trois couches de convolution, le tout en parallèle. Ces quatre flux donnent alors chacun une carte de caractéristiques, en plus de celle calculée par la sortie du VGG16. Ces 5 cartes sont ensuite concaténées en une carte à cinq canaux, et fusionnées par une couche de convolution de noyau  $1 \times 1$  produisant la première segmentation à un canal :  $conv1(1, 1)$ .

La même couche de convolution ( $conv1(1, 1)$ ) est également utilisée dans le système de fusion proposé par Tang et al. ([Tang2016a]) afin de fusionner cinq cartes de saillances qui sont préalablement concaténées en une image à cinq canaux. Le réseau proposé est également un réseau basé sur l'architecture VGG16, mais il s'agit d'un réseau *Deeply-Supervised Recurrent Convolutional Neural Network* (DSRCNN), qui est la combinaison entre un réseau de neurones récurrent [Liang2015] et un réseau profond supervisé [Lee2015].

### 3.2.2/ LA FUSION PAR TROIS COUCHES DE CONVOLUTION

Tang et al. réalisent également un autre système de détection de saillances [Tang2016b]. Comme pour le système de segmentation réalisé par Li et al. [Li2016], ils utilisent deux CNNs :

- un réseau basé sur l'architecture VGG16 pour une segmentation au niveau pixels de l'image en entrée,
- et un autre pour une segmentation au niveau régions.

Le réseau de fusion est cependant plus profond, comme le montre la Figure 3.10. Afin d'apporter plus d'informations, l'image d'origine est également utilisée dans le processus de fusion. Les trois images (l'image d'origine en RGB, la carte de saillances au niveau pixels et la carte de saillances au niveau régions) sont concaténées en une seule image à 5 canaux. Trois couches de convolution sont alors appliquées afin de réaliser la fusion :  $conv1(64, 3) - conv2(128, 3) - conv3(1, 1)$ .

### 3.2.3/ LA FUSION PAR FCM

Xiao et al., dans [Xiao2018], introduisent un Module de Fusion par Convolution (FCM présenté dans la Figure 3.11), afin de fusionner différentes cartes de saillances

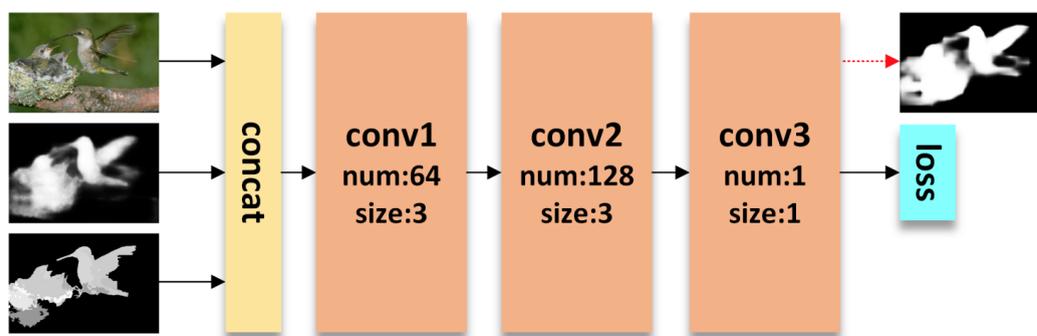


FIGURE 3.10 – Le réseau de fusion à trois couches pour la détection de saillances [Tang2016b].

préalablement obtenues. Les cartes de saillances sont obtenues avec un réseau de neurones convolutionnel récurrent (RCNN) ainsi que l'utilisation de plusieurs couches de déconvolution empilées. Ce module FCM consiste à concaténer les quatre cartes de saillances obtenues en une image à quatre canaux. Le réseau de fusion est ensuite similaire au réseau présenté dans [Tang2016b], puisque trois couches de convolution sont appliquées avec un remplissage (*padding*) de 1, avec pour chaque convolution une couche ReLU :  $conv1(32, 3) - relu1 - conv2(64, 3) - relu2 - conv3(1, 3) - relu3$ .

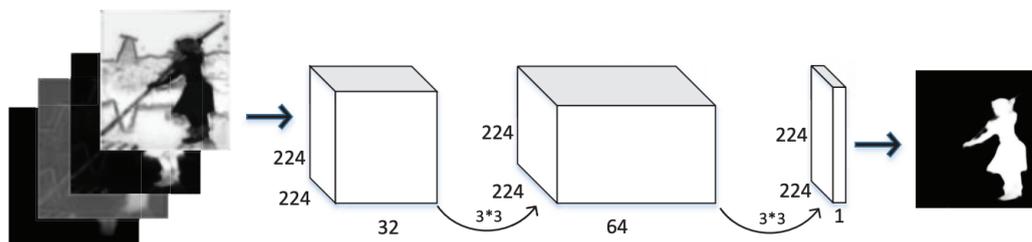


FIGURE 3.11 – Le module de fusion FCM pour la détection de saillances [Xiao2018].

### 3.2.4/ LA FUSION PAR UN RÉSEAU PLUS PROFOND

La solution de détection de saillances proposée par Qu et al. [Qu2017] est composée de plusieurs parties :

- La première consiste à générer les superpixels de l'image en entrée.
- Plusieurs vecteurs de caractéristiques de saillances, *saliency feature vectors*, sont alors générés pour chaque région.
- Puis le réseau de fusion, présenté dans la Figure 3.12, combine les cartes de saillances multiples.
- Et enfin la dernière partie consiste à utiliser la propagation laplacienne aidant à obtenir la carte de saillances finale.

Le réseau de fusion consiste en trois couches de convolution, chacune suivie par une couche de régression logistique avec la fonction non linéaire sigmoïde, ainsi que d'une

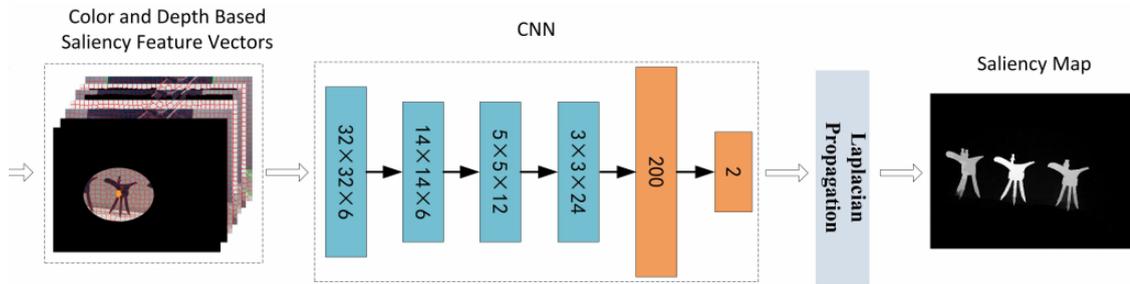


FIGURE 3.12 – Le réseau de fusion inclus dans le réseau pour la détection de saillances [Qu2017].

couche de *pooling* pour les deux premières couches de convolution. Puis deux couches FC avec une couche ReLU entre les deux sont mises en place. Plus précisément, l'architecture est :  $conv1(6, 5) - sig1 - pool1(MEAN, 2) - conv2(12, 5) - sig2 - pool2(MEAN, 2) - conv3(24, 3) - sig3 - fc4(200) - relu4 - dropout4 - fc5(2)$ .

Cependant, la fusion des segmentations ou des cartes de caractéristiques n'est pas seulement liée à la détection de saillances. La fusion de segmentation proposée par Hu et al. intervient dans le domaine de la segmentation d'images interactives ([Hu2019b]). Le réseau de segmentation présenté dans la Figure 3.13, appelé TSLFN, est composé de deux entrées : l'une est l'image d'origine, et l'autre les cartes d'interactions dites positive et négative des utilisateurs. Pour ces deux entrées, le réseau est composé de deux flux similaires constitués des 10 premières couches de convolution du réseau VGG16 (cf. Section 3.1.2.6). Puis, les cartes de caractéristiques résultantes des deux flux sont concaténées. Le réseau de fusion est ensuite utilisé. Il est composé de 6 couches de convolution, chacune étant suivie d'une couche ReLU, sauf pour la dernière couche de convolution qui est suivie d'une couche de déconvolution. Une couche de *pooling* est également appliquée avant la quatrième couche de convolution :  $conv1(512, 3) - relu1 - conv2(512, 3) - relu2 - conv3(512, 3) - relu3 - pool3(MAX, 2) - conv4(4096, 7) - relu4 - conv5(4096, 1) - relu5 - conv6(1, 1) - deconv(1, 64)$ .

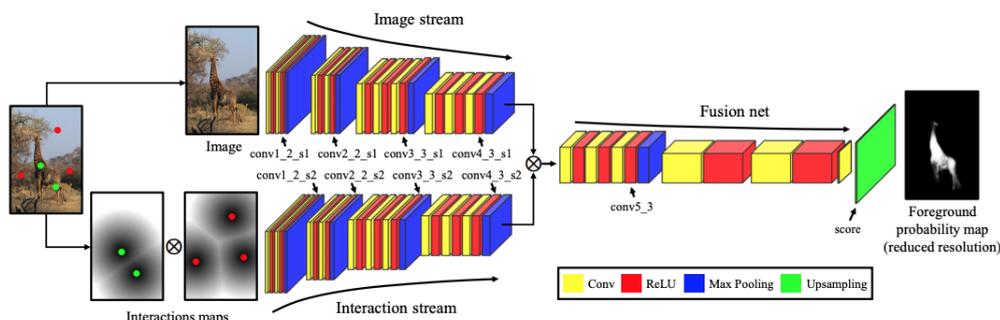


FIGURE 3.13 – Le réseau pour la segmentation interactive incluant un réseau de fusion profond [Hu2019b].

De nombreux autres réseaux utilisent une simple technique d'addition par élément pour fusionner plusieurs cartes de caractéristiques provenant de différents flux dans leur réseau de segmentation. Il n'y a alors pas de réseau ni de couche spécifique à la fusion. C'est le cas pour un réseau de segmentations connu, appelé *FuseNet*, créé en 2016 [Hazirbas2016], et qui segmente à partir d'une image RVB et de son image de

profondeur. Le réseau utilise le principe des réseaux *encoder-decoder*, comme pour les réseaux DeconvNet [Noh2015] ou SegNet [Badrinarayanan2017]. Cependant, comme le montre la Figure 3.14 présentant le réseau FuseNet, il y a alors deux flux encodeurs (inspirés de l'architecture VGG16), un pour chaque image en entrée, dont les différentes cartes de caractéristiques du flux profond sont fusionnées par addition avec le flux RVB.

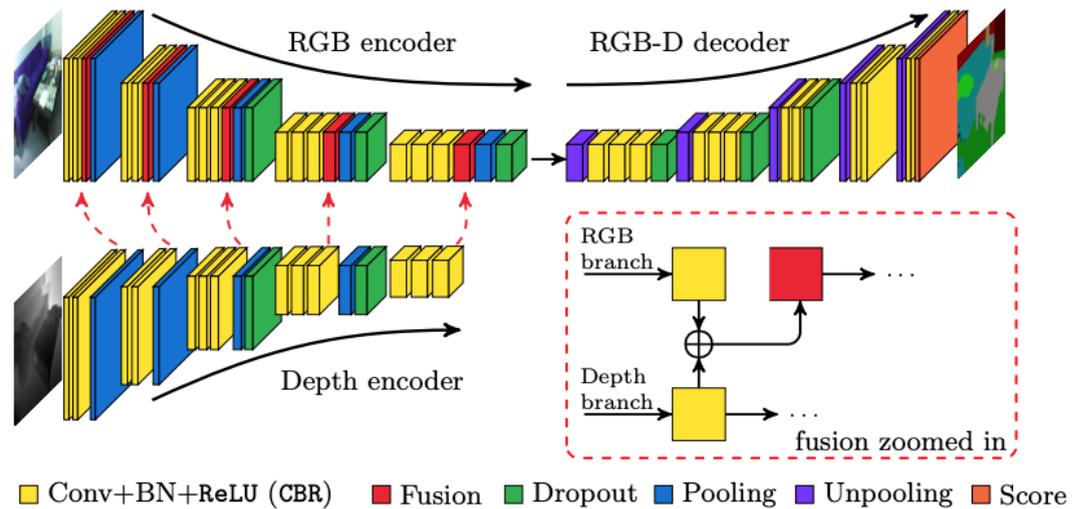


FIGURE 3.14 – L'architecture de FuseNet, réseau de type encodeur-décodeur avec deux flux encodeur qui sont fusionnés [Hazirbas2016].

Plusieurs travaux ont ensuite suivi ce principe pour fusionner deux types d'entrées différentes à travers deux flux encodeurs [Jiang2018, Hu2019a, Zhang2018a]. Récemment, Sun et al [Sun2019] ont réalisé un réseau profond dans le cadre de la segmentation des scènes urbaines pour les véhicules autonomes, et fonctionnels dans des conditions d'éclairage insatisfaisantes. Ce réseau, basé sur le réseau ResNet [He2016], utilise le même principe de fusion par addition en utilisant une image RVB de la scène et une image thermique afin d'apporter de plus amples informations non visibles dans l'image couleur à cause de l'éclairage.

**Synthèse du chapitre**

Dans ce chapitre nous avons vu les différentes notions liées aux réseaux de neurones et à leur fonctionnement, avec notamment le perceptron, le perceptron multicouche et les principes de l'apprentissage pour les réseaux, profonds ou non, grâce à la rétropropagation du gradient et à l'algorithme SGD (algorithme de descente du gradient stochastique). Les CNNs, principalement utilisés pour le traitement d'images, ont également été présentés, ainsi que les différentes couches qui les composent, comme les couches de convolution et de pooling. Enfin, nous avons également détaillé les différents réseaux utilisés dans la littérature pour la fusion de multiples segmentations, et nous avons précisé qu'il s'agit d'un domaine encore assez peu exploré. Cependant, ces réseaux de fusion sont très présents dans le domaine de la détection de saillances afin d'améliorer le résultat de la segmentation de l'objet le plus saillant dans l'image. Les réseaux utilisés sont, pour la plupart, encore assez peu profonds, et certains réseaux de fusion sont directement intégrés aux réseaux de segmentations fournissant les images, ou *features maps*, à fusionner. De plus, aucun n'a été créé ni testé sur la fusion de segmentations complémentaires.

Cette première partie était consacrée à l'exposé du contexte de ce travail de thèse ainsi qu'à l'état de l'art de différentes méthodes dites "classiques" utilisées dans la littérature pour la fusion de multiples segmentations et la gestion de conflits.

Nous avons également présenté un état de l'art approfondi sur les réseaux de neurones, en précisant particulièrement les types de réseaux utilisés dans le traitement d'images, ainsi que différents réseaux utilisés pour la fusion de multiples segmentations dans la littérature.

La seconde partie de ce document constitue le cœur de nos contributions dans le domaine de la fusion de segmentations multiples et de la gestion des conflits qui en résulte.



## CONTRIBUTION



Cette seconde partie est consacrée à nos contributions qui s'articulent principalement sur des techniques de fusion de segmentations complémentaires et la gestion des conflits en résultant : par des méthodes dites classiques et des méthodes d'Intelligence Artificielle (IA). Ces techniques de fusion de segmentations sont intégrées dans le cadre du projet SAIAD, réalisant préalablement la segmentation par IA de différentes structures de patients atteints de néphroblastome. Chacun de ces chapitres présente dans un premier temps la contribution réalisée, puis dans un second temps les résultats obtenus et l'évaluation des performances sur l'ensemble des patients disponibles dans le projet, ainsi qu'une discussion générale.

Tout d'abord, le quatrième chapitre introduit notre première contribution pour laquelle nous proposons deux modèles qui sont des combinaisons d'un ensemble de méthodes dites classiques (nouvelles ou existantes dans la littérature) pour la fusion d'images et la gestion de pixels en conflit. Ces méthodes utilisent de multiples informations présentes dans les différentes images, qui sont des images scanner ou des segmentations, et se basent sur différents critères. Le résultat de chacune de ces méthodes est alors agrégé, de différentes manières possibles afin de comparer et de déterminer le meilleur système d'agrégation des résultats.

Le cinquième chapitre présente nos contributions dans l'utilisation pour la fusion de méthodes dites d'IA, et plus particulièrement l'utilisation de réseaux de neurones de type convolutifs (CNNs) ainsi que de Raisonnements à Partir de Cas (RàPC) :

- Une première méthode de fusion de segmentations complémentaires se base sur le couplage d'un CNN avec un RàPC, dans lequel le RàPC permet de choisir dans sa base les cas les plus proches à utiliser pour l'entraînement du CNN, avec l'utilisation d'entraînements successifs en fonction de chaque cas.
- Une seconde méthode utilise ce même CNN mais y applique un entraînement spécifique, appelé OVASSION, permettant le sur-apprentissage par patient. Nous adaptons ici pour la fusion de segmentations OVASSION qui avait été initialement prévue dans le cadre de la segmentation d'images.



## AGRÉGATION DE DIFFÉRENTES MÉTHODES CLASSIQUES POUR LA FUSION DE SEGMENTATIONS ET LA GESTION DE CONFLITS

Comme nous l'avons précédemment expliqué dans les chapitres consacrés à l'état de l'art, la fusion de différentes segmentations peut engendrer des conflits qu'il faut alors résoudre. Dans cette première contribution, deux modèles de gestion des conflits sont proposés pour la fusion de segmentations complémentaires. Cette procédure de gestion des conflits et de fusion, intégrée dans le projet SAIAD, permet de fusionner les segmentations du néphroblastome et du rein pathologique en utilisant la **combinaison de six méthodes indépendantes**. Ces méthodes sont basées sur différents critères, comme l'utilisation des coupes segmentées adjacentes, la variation de l'information, le score de Dice, les étiquettes voisines, l'intensité des pixels dans l'image scanner, ainsi que les champs aléatoires conditionnels. L'ensemble des résultats nous permet alors de collecter les différentes informations 3D provenant des scanners ainsi que des segmentations du patient pour arbitrer et résoudre les conflits.

### 4.1/ LES DIFFÉRENTES MÉTHODES DE GESTION DE CONFLITS UTILISÉES ET ADAPTÉES

Pour cette première contribution, nous partons de la fusion réalisée en amont du processus de gestion des conflits. Cette fusion, présentée dans la Figure 4.1, regroupe la segmentation de la tumeur et la segmentation du rein pathologique, segmentations réalisées de manière triviale au sein de la plateforme SAIAD, dans laquelle une nouvelle étiquette temporaire, présentée en couleur bleue claire, est assignée aux pixels se trouvant en conflit, c'est-à-dire les pixels étiquetés comme appartenant à différentes structures dans les différentes segmentations.

L'idée du processus de gestion des conflits est de lancer 6 méthodes distinctes, utilisant différentes informations sur les images scanner et/ou les segmentations. Leurs résultats sont ensuite combinés afin d'obtenir la segmentation consensus sans conflit.

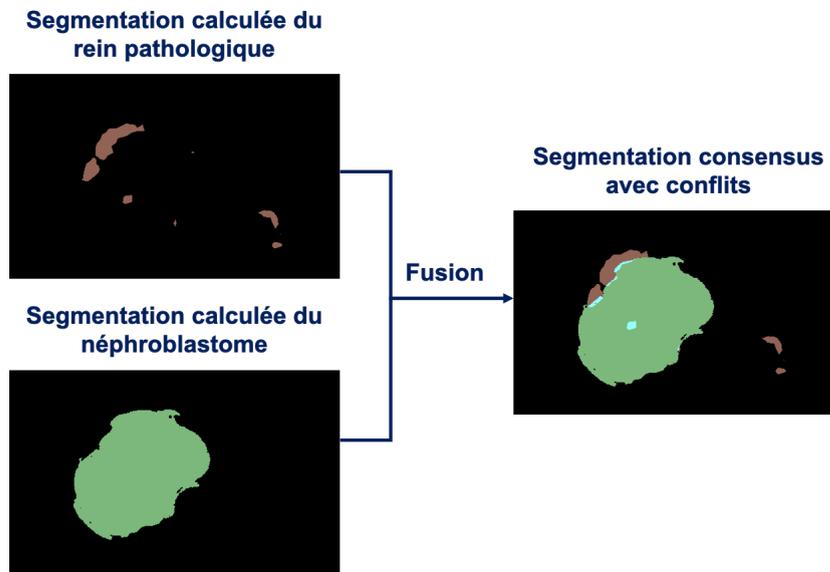


FIGURE 4.1 – Fusion de segmentations calculées du rein pathologique et de la tumeur, appartenant à une coupe de l’abdomen d’un patient.

#### 4.1.1/ NOTRE NOUVELLE MÉTHODE DE GESTION DE CONFLITS NOMMÉE ADS2

Les segmentations des coupes adjacentes peuvent donner des informations sur les pixels en conflit de la segmentation actuelle. En effet, nous avons supposé que les segmentations des coupes adjacentes étaient relativement similaires, puisque les structures représentées ont plus ou moins les mêmes formes et positions. Pour cette raison, nous avons conçu la méthode des coupes segmentées adjacentes ADS2 (*AD*joint *S*egmented *S*lices) pour la résolution des conflits. Cette méthode examine les étiquettes de chaque pixel en conflit dans les segmentations des coupes adjacentes. Le pseudo-code de la méthode ADS2 est présenté dans l’algorithme 2.

---

**Algorithme 2** : Méthode de résolution de conflits basée sur ADS2.

---

```

1 pour chaque coupe segmentée  $S$  faire
2   pour chaque pixel en conflit  $p$  possédant l'étiquette  $l_S^p$  faire
3     si  $l_{S-1}^p$  et  $l_{S+1}^p$  ne sont pas en conflit alors
4       si  $l_{S-1}^p == l_{S+1}^p$  alors
5          $l_S^p = l_{S-1}^p$ ;
6       fin
7     fin
8   fin
9 fin

```

---

Considérant un pixel conflictuel  $p$  étiqueté par  $l_S^p$  dans la coupe  $S$ , si ce pixel dans les coupes adjacentes  $S - 1$  et  $S + 1$  n’est pas en conflit et que ses étiquettes  $l_{S-1}^p$  et  $l_{S+1}^p$  possèdent la même valeur (c’est-à-dire que  $l_{S-1}^p = l_{S+1}^p$ ), alors nous supposons que son étiquette à la segmentation actuelle  $S$  est la même ( $l_S^p = l_{S-1}^p = l_{S+1}^p$ ).

Malheureusement, cette méthode ne peut pas résoudre tous les conflits. En effet, ADS2 ne peut être appliquée que dans des situations spécifiques dans lesquelles les pixels

adjacents ne sont pas en conflit. C'est pourquoi nous avons également conçu et mis en œuvre d'autres méthodes.

#### 4.1.2/ L'ADAPTATION DE LA MÉTHODE BASÉE SUR VOI ET LA NOUVELLE MÉTHODE BASÉE SUR DICE

##### 4.1.2.1/ LE CRITÈRE VOI POUR LA GESTION DES CONFLITS

Le critère de variation de l'information, ou Vol, a également été intégré dans nos modèles. À la base, le Vol peut être utilisé dans la fusion de multiples segmentations, provenant d'une même image et comportant les mêmes éléments, en parcourant tous les pixels et en fusionnant l'ensemble des segmentations, avec la méthode présentée dans la Section 2.3.2, comme c'est le cas dans [Mignotte2014a] et [Nguyen2018]. Il y a alors dans ce cas une étape dans l'algorithme de descente d'énergie locale qui détermine le nombre de régions et le nombre d'étiquettes possibles dans la segmentation consensus. Dans notre cas, nous réalisons la fusion de segmentations dites complémentaires, comportant donc des éléments différents. De plus, le nombre d'étiquettes dans notre cas est déjà défini, car nous savons que nous avons deux éléments dans la segmentation : la tumeur et le rein pathologique, ce qui porte à 3 le nombre d'étiquettes (en ajoutant l'étiquette de l'arrière-plan pour les pixels n'appartenant pas à ces deux éléments.)

Nous avons donc adapté cette méthode pour que la métrique ne puisse gérer que les pixels en conflit de notre segmentation fusionnée. De plus, comme nous travaillons sur plusieurs coupes segmentées d'un patient, nous avons voulu profiter des informations 3D disponibles dans les segmentations du patient.

La méthode de gestion de conflits basée sur la métrique Vol est présentée par l'algorithme 3. Le Vol moyen dérivé ( $\Delta \overline{Vol}$ ), présenté dans l'équation 29 et permettant d'attribuer une étiquette à un pixel, est maximisé par un algorithme ICM (*Iterated Conditional Modes*, cf. Section 2.3), réalisant une descente d'énergie locale itérative.

---

**Algorithme 3** : Méthode de résolution de conflits basée sur la métrique Vol.

---

**Données :**

$\{S_k\}_{k \leq L}$  : L'ensemble de  $L$  coupes segmentées d'un patient

$n_a$  : Le nombre de coupes adjacentes prises en compte

$\mathcal{E}$  : L'ensemble des étiquettes possibles pour un pixel

$T_{max}$  : Le nombre maximal d'itérations

```

1 pour chaque coupe segmentée  $S$  dans  $\{S_k\}_{k \leq L}$  faire
2    $\{S_k\}_{k \in \{k-n_a, \dots, k+n_a\}}$  : L'ensemble des segmentations fusionnées voisins à  $S$  d'un
   patient
3   tant que  $i < T_{max}$  faire
4     pour chaque pixel en conflit  $p$  avec l'étiquette  $l_S^p$  dans  $S$  faire
5       Calculer  $VoIMax = \arg \max_{x \in \mathcal{E}} \Delta \overline{Vol}(S, \{S_k\}_{k \in \{k-n_a, \dots, k+n_a\}})_{l_S^p \rightarrow x}$ 
6       Remplacer l'étiquette  $l_S^p$  par l'étiquette  $x$  de  $VoIMax$  dans  $S$ 
7     fin
8      $i = i + 1$ 
9   fin
10 fin

```

---

Sachant que pour un patient, les coupes adjacentes sont similaires, que ce soit sur les images scanner ou sur les segmentations, cette méthode prend en compte les coupes adjacentes à la segmentation courante, jusqu'à une certaine distance notée  $n_a$ . Par exemple, pour la gestion des pixels en conflit de la quatrième coupe segmentée du patient avec une distance  $n_a = 2$ , les coupes segmentées  $\{2, 3, 4, 5, 6\}$  sont prises en compte. Notons que la segmentation courante se retrouve également dans l'ensemble des segmentations, car elle peut également posséder des informations utiles à la résolution. L'algorithme itère alors pour chaque segmentation, et récupère l'ensemble des segmentations voisines, de distance  $n_a$  par rapport à la segmentation courante. Ensuite, pour chaque pixel en conflit présent dans la segmentation, on calcule le Vol moyen dérivé entre la segmentation courante et l'ensemble des segmentations voisines pour chaque étiquette possible. L'étiquette finale (de l'itération en cours) sera l'étiquette ayant obtenu la valeur la plus élevée de  $\overline{\Delta Vol}$ . Cette étape est répétée tant que  $i < T_{max}$ ,  $T_{max}$  étant décidé arbitrairement, adapté suivant l'utilisation pour faciliter la convergence.

#### 4.1.2.2/ LA MÉTHODE DE GESTION DE CONFLITS BASÉE SUR LE CRITÈRE DICE

La métrique Dice, présentée dans la Section 1.5.1.2, est couramment utilisée en imagerie médicale pour la comparaison entre deux segmentations médicales. Il est alors pertinent de l'ajouter dans nos modèles de gestion des conflits. La méthode utilisant ce critère est presque identique à la méthode basée sur le critère Vol (cf. Section 4.1.2.1). De la même manière, la métrique moyenne, en l'occurrence ici le Dice moyen, est calculée pour une segmentation  $S_i$  en fonction d'un ensemble de  $L$  segmentations  $\{S_k\}_{k \leq L}$ . En complément, un poids est associé à chaque coupe afin de donner un Dice plus important aux coupes les plus proches.

Le Dice moyen est alors défini par :

$$\overline{Dice}(S_i, \{S_k\}_{k \leq L}) = \frac{1}{L} \sum_{k=1}^L Dice(S_i, S_k) * r_{i,k} \quad (46)$$

Où :

- $r_{i,k}$  est le poids attribué à chaque segmentation, défini par :

$$r_{i,k} = 1 - \frac{|i - k|}{n_a * 2 + 1} \quad (47)$$

Où :

- $n_a$ , comme pour la méthode basée sur le critère Vol, est le nombre de coupes segmentées voisines à prendre en compte pour chaque segmentation  $S_i$ .

La segmentation fusionnée avec les bonnes étiquettes est alors celle obtenant la plus grande valeur de  $\overline{Dice}$  :

$$\hat{S}_{i \overline{Dice}} = \arg \max_{S_i \in S_n} \overline{Dice}(S_i, \{S_k\}_{k \leq L}) \quad (48)$$

Où :

- $S_n$  est l'ensemble de toutes les segmentations possibles à  $n$  pixels.

L'algorithme de résolution des pixels en conflit est le même que l'algorithme 3 avec l'utilisation d'un algorithme de type ICM, dans lequel un nombre déterminé de segmentations voisines est pris en compte. Le Dice moyen ( $\overline{Dice}$ ) est alors à maximiser et les étiquettes choisies seront celles permettant d'obtenir le  $\overline{Dice}$  le plus élevé.

#### 4.1.3/ NOUVELLE MÉTHODE 3D-NANDI ISSUE DE NANDI

La méthode de fusion Nandi, présentée dans la Section 2.4.1, calcule la meilleure étiquette possible pour chaque pixel en conflit de l'image en fonction des pixels voisins. Ainsi, il est facile de l'adapter à notre situation. Les conditions de fonctionnement de Nandi, pour obtenir un voisinage effectif d'un pixel dans le but de calculer la similarité, suppose qu'il y ait deux segmentations  $S$  et  $S'$  calculées à partir d'une même image source. Ces deux segmentations sont nécessaires pour vérifier qu'un pixel  $q$  que l'on souhaite ajouter au voisinage effectif d'un pixel  $p$  ne soit pas en conflit dans ces segmentations. Dans notre cas, n'ayant pas plusieurs segmentations d'une même image, nous avons considéré sur ce point que  $q$  ne devait simplement pas être un pixel en conflit dans la segmentation actuelle.

---

**Algorithme 4 :** La méthode Nandi pour la résolution des pixels en conflit.

---

**Données :**

$\{S_k\}_{k \leq L}$  : L'ensemble de  $L$  coupes segmentées d'un patient  
 $r$  : Le rayon

```

1 pour chaque coupe segmentée  $S$  dans  $\{S_k\}_{k \leq L}$  faire
2   | pour chaque pixel en conflit  $p$  dans  $S$  faire
3   |   | Calculer le voisinage effectif  $E_p^r$ 
4   |   | Calculer  $l_s^p$  en fonction du calcul de similarité  $Sim$  et  $E_p^r$ 
5   |   fin
6 fin
```

---

La méthode utilisée pour la gestion des conflits avec Nandi est présentée dans l'algorithme 4. Le voisinage de chaque pixel en conflit pour chaque coupe segmentée est calculé dans un premier temps. L'étiquette finale du pixel en conflit est attribuée suivant la similarité calculée avec ce voisinage (cf. Équation 37).

---

**Algorithme 5 :** La méthode 3D-Nandi pour la résolution des pixels en conflit.

---

**Données :**

$\{S_k\}_{k \leq L}$  : L'ensemble de  $L$  coupes segmentées d'un patient  
 $r$  : Le rayon

```

1 pour chaque coupe segmentée  $S$  dans  $\{S_k\}_{k \leq L}$  faire
2   | pour chaque pixel en conflit  $p$  dans  $S$  faire
3   |   | Calculer le voisinage effectif 3D  $E_p^r$ 
4   |   | Calculer  $l_s^p$  en fonction du calcul de similarité 3D  $Sim$  et  $E_p^r$ 
5   |   fin
6 fin
```

---

Cependant, la méthode Nandi utilise uniquement les informations de la coupe actuelle. Comme nous travaillons sur un ensemble de coupes segmentées correspondant à l'ab-

domen d'un patient, les informations 3D peuvent également être prises en compte. Ainsi, nous avons développé la méthode 3D-Nandl (cf. Algorithme 5) qui est une évolution de la méthode Nandl, dans laquelle un voisinage effectif 3D et une similarité 3D sont calculés. Comparée à la similarité classique définie dans l'Équation 36, la similarité 3D utilise simplement une distance euclidienne 3D au lieu d'une distance euclidienne en 2D. Les coupes adjacentes prises en compte sont celles situées dans un rayon défini à l'initialisation de l'algorithme. Les conditions déterminant le voisinage d'un pixel en conflit restent inchangées.

#### 4.1.4/ L'ADAPTATION DES CHAMPS ALÉATOIRES CONDITIONNELS POUR LA GESTION DE CONFLITS

Les champs aléatoires conditionnels (CRFs, *Conditional Random Fields*), sont des modèles statistiques utilisés pour la reconnaissance de forme et l'apprentissage automatique [Lafferty2001]. Les CRFs sont également utilisés dans le traitement d'images et la segmentation, et ont comme principal avantage de fournir des informations spatiales sur la corrélation entre les pixels voisins. Krahenbuhl et al. [Krahenbuhl2011] proposent une approche probabiliste des CRFs pour l'amélioration des segmentations et de leur étiquetage, avec un calcul d'inférence très performant, présenté ci-dessous :

##### Définitions

Soit  $\mathbf{I} = \{I_1, I_2, \dots, I_N\}$  une image composée de  $N$  pixels et  $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$  les variables aléatoires associées aux étiquettes possibles de chaque pixel (noté  $\mathcal{L} = \{l_1, l_2, \dots, l_k\}$ ), on peut alors obtenir un CRF( $\mathbf{I}, \mathbf{X}$ ) qui associe à un étiquetage  $\mathbf{x}$  de l'image  $\mathbf{I}$  une énergie de Gibbs  $E$  :

$$E(\mathbf{x}) = \underbrace{\sum_i \psi_u(x_i)}_{\text{Énergie unitaire}} + \underbrace{\sum_{i < j} \psi_p(x_i, x_j)}_{\text{Énergie de Parité}} \quad (49)$$

Où :

- la première partie est une **énergie unitaire** (*Unary Potential*) réalisée par un classificateur,
- et la deuxième partie est une **énergie de parité** (*Pairwise Potential*) pénalisant l'attribution d'étiquettes différentes à des pixels similaires.

Le potentiel de parité étant défini par :

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \underbrace{\sum_{m=1}^K w^{(m)} k^{(m)}(f_i, f_j)}_{k(f_i, f_j)} \quad (50)$$

Où :

- $\mu$  est une fonction de compatibilité des étiquettes entre les pixels  $i$  et  $j$ ,
- et  $k$  est une somme de noyaux gaussiens en fonction des caractéristiques  $f_i$  et  $f_j$ .

Ces caractéristiques sont définies par :

$$k(f_i, f_j) = w^{(1)} \underbrace{\exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right)}_{\text{Noyau d'Apparence}} + w^{(2)} \underbrace{\exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right)}_{\text{Noyau de Lissage}} \quad (51)$$

Où :

- $I_i$  et  $I_j$  sont la valeur des pixels  $i$  et  $j$ ,
- $p_i$  et  $p_j$  leur position,
- $\theta_\alpha$ ,  $\theta_\beta$  et  $\theta_\gamma$  des coefficients pondérant les distances et les intensités des pixels,
- et  $w^{(1)}$  et  $w^{(2)}$  sont des poids associés aux deux noyaux : le noyau dit d'apparence (*Appearance Kernel*) et le noyau de lissage (*Smoothness Kernel*).

L'énergie unitaire est calculée par un classificateur unitaire quelconque. Une étape d'inférence permet ensuite de déduire les paramètres optimaux de l'énergie de parité. L'inférence utilisée dans [Krahenbuhl2011] est une combinaison linéaire de noyaux gaussiens.

Les approches de classification par pixel, comme les CNNs, attribuent une étiquette par pixel, indépendamment des étiquettes des pixels voisins. C'est pourquoi ces approches peuvent avoir certaines limites et incohérences quand les différents éléments à segmenter sont très similaires. Les CRFs peuvent alors être combinés aux CNNs, afin d'apporter des informations spatiales, ce qui aura pour avantage l'amélioration des contours des éléments segmentés et la suppression des éventuels artéfacts. Les résultats des CNNs correspondent alors à l'énergie unitaire, qui sera combinée à l'énergie de parité.

Les CRFs peuvent être utilisés en tant que post-traitement d'un réseau de neurones pour la segmentation, comme c'est le cas pour le réseau DeconvNet [Noh2015], le réseau DeepLab [Chen2017] et bien d'autres ([Kamnitsas2017] [Jiang2017]). Ils peuvent également faire partie intégrante des réseaux ([Zheng2015] [Luo2016] [Lin2016]), ce qui permet à certains paramètres de l'énergie de parité, comme  $w_a$ ,  $w_r$ , et  $\mu$ , d'être appris lors de l'entraînement du réseau, mais surtout d'optimiser les couches de convolution en fonction des résultats du CRF.

Nous avons détourné l'utilisation classique des CRFs de [Krahenbuhl2011] afin de les utiliser pour la gestion des pixels en conflit. La Figure 4.2 montre le processus de cette méthode de fusion basée sur les CRFs.

Les CRFs ont besoin des cartes de probabilité des segmentations pour pouvoir fonctionner (qui seront alors la partie d'énergie unitaire dans l'énergie de Gibbs calculée). Comme nos images scanner ont été préalablement segmentées par un CNN (cf. Section 1.4.2), nous pouvons facilement récupérer les cartes de probabilités de ces segmentations à la fin du réseau de neurones.

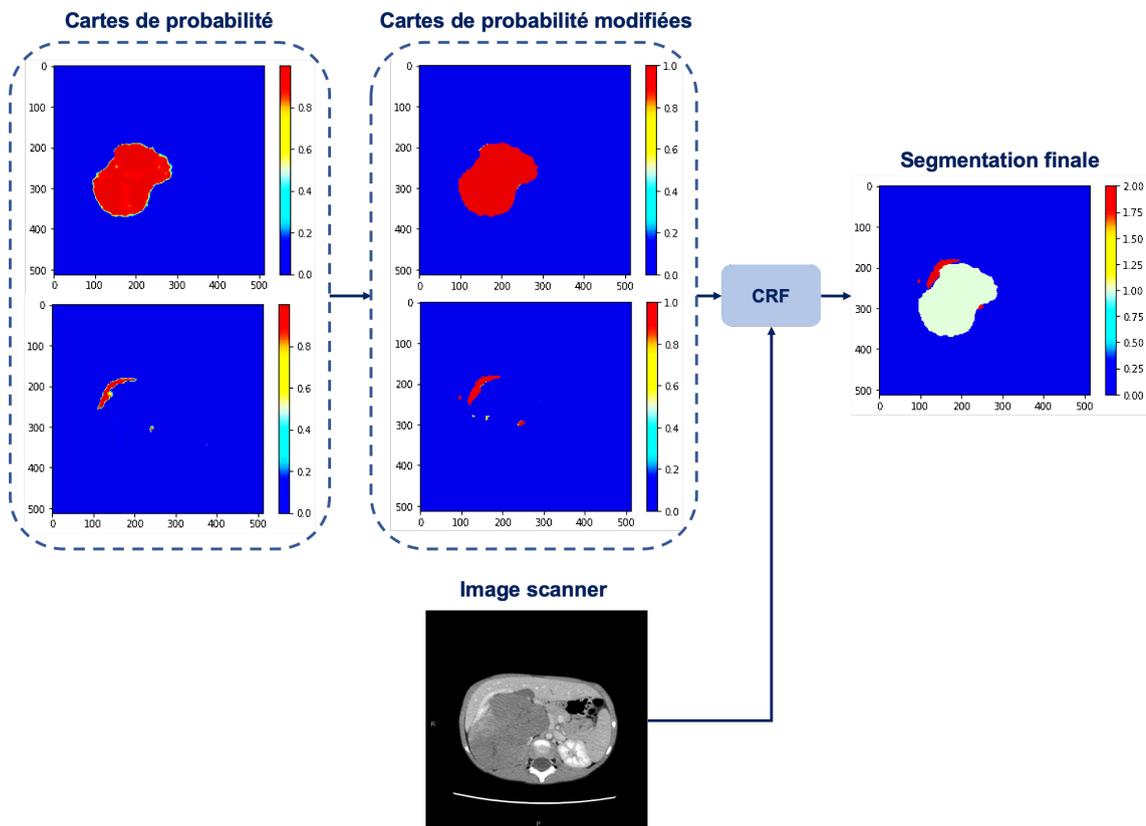


FIGURE 4.2 – La fusion de segmentations et gestion des conflits avec la méthode basée sur les CRFs.

Il y a une carte par étiquette possible et les valeurs des pixels sont comprises entre 0 et 1, et correspondent au pourcentage d'appartenance à l'étiquette de la carte. Ces valeurs sont alors ajustées à 0 ou 1 pour les pixels non conflictuels, ce qui aura pour effet que les étiquettes de ces pixels ne seront pas modifiées par le CRF. Au contraire, les valeurs des pixels conflictuels restent dans cet intervalle. Ces cartes de probabilité modifiées sont données en entrée du CRF, en plus des images correspondantes du scanner, et le CRF déterminera alors les étiquettes des pixels en conflit en produisant une segmentation consensuelle.

## 4.2/ LES MODÈLES DE GESTION DE CONFLITS IMPLÉMENTÉS

### 4.2.1/ LE FONCTIONNEMENT DES MODÈLES

Les Figures 4.3 et 4.4 montrent nos deux versions du système de gestion des conflits basées sur la combinaison des six méthodes décrites dans les sections précédentes. L'entrée du système est composée de l'ensemble des segmentations calculées du rein et de la tumeur qui ont été fusionnées en ajoutant une nouvelle étiquette pour les pixels en conflit, ainsi que les images scanner correspondantes. Dans le premier modèle proposé par la Figure 4.3, les conflits sont traités en parallèle par les 6 méthodes, dont leur résultat sera à fusionner. La deuxième version du système, présentée par la Figure 4.4, utilise la

méthode ADS2 en amont en tant que pré-traitement, ce qui permet d'obtenir par la suite 5 résultats différents fournis par les autres méthodes, à fusionner.

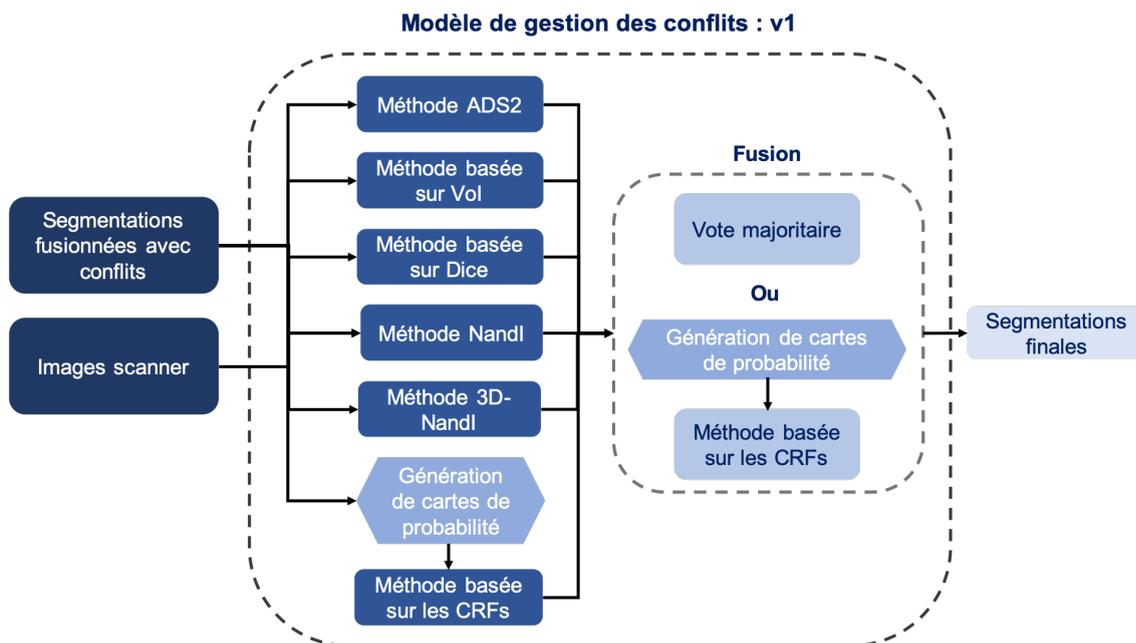


FIGURE 4.3 – Première version du modèle de gestion des conflits pour la fusion de segmentations rénales tumorales.

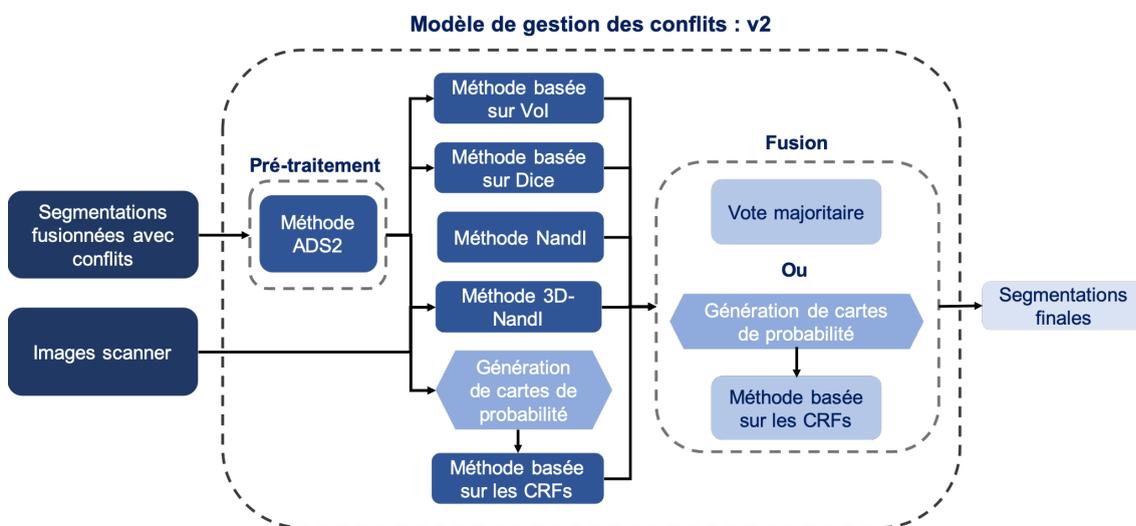


FIGURE 4.4 – Deuxième version du modèle de gestion des conflits, avec un pré-traitement, pour la fusion de segmentations rénales tumorales.

Certaines méthodes, comme la méthode ADS2 et les méthodes basées sur le critère Vol et le score de Dice n'ont besoin que des segmentations pour régler les conflits. Les autres utilisent également les informations contenues dans les images scanner : comme la méthode NandI et celle basée sur le CRF qui utilisent la segmentation courante et son image scanner correspondante, ou bien encore la méthode 3D-NandI qui utilise plusieurs segmentations et images scanner afin de résoudre les pixels en conflit d'une segmenta-

tion. Notons toutefois que la méthode basée sur le CRF n'utilise pas les segmentations telles quelles, mais nécessite des cartes de probabilités de ces segmentations.

Afin de combiner les résultats des différentes méthodes, nous avons utilisé deux méthodes différentes afin de pouvoir les comparer : le vote majoritaire et le CRF.

#### 4.2.2/ LA FUSION DES RÉSULTATS DE LA GESTION DE CONFLITS

Les résultats des méthodes peuvent ensuite être agrégés de deux manières différentes :

- La première solution est basée sur un vote majoritaire entre les étiquettes calculées par chaque méthode pour chaque pixel en conflit.
- La seconde solution est basée sur l'utilisation des CRFs pour fusionner les résultats.

Toutefois, avec le vote majoritaire, il peut y avoir égalité dans le premier modèle pour les conflits pouvant être résolus par la méthode ADS2 car il y a alors un nombre pair de résultats à traiter. Dans ce cas, plusieurs solutions sont possibles, comme déterminer l'étiquette finale par une autre méthode, ou alors décider arbitrairement que ces pixels soient étiquetés comme appartenant à la tumeur ou au rein. Dans le deuxième modèle, ADS2 est réalisée en tant que pré-traitement, il y a alors un nombre impair de résultats à fusionner et donc les problèmes d'égalité pour la fusion par vote majoritaire disparaissent. Cela revient également à accorder la prévalence à la méthode ADS2 sur toutes les autres.

Dans le cas de l'utilisation des CRFs pour la fusion des résultats, des problèmes d'égalités ne peuvent pas apparaître. Les cartes de probabilités voulues par le CRF sont alors créées à partir des résultats des méthodes. Pour chaque pixel en conflit, la valeur de ce pixel dans la carte de probabilité d'une étiquette est le nombre de fois où cette étiquette est choisie, divisé par le nombre total de résultats.

Un exemple de la génération des cartes de probabilités est présenté sur la Figure 4.5. Dans le cas simple d'une petite segmentation  $3 \times 3$  possédant deux étiquettes possibles (la tumeur en vert et le rein en orange) et avec un seul pixel en conflit en bleu, si 4 méthodes choisissent l'étiquette du rein pour le pixel en conflit, et les 2 autres l'étiquette de la tumeur, alors la valeur de ce pixel dans la carte de probabilités de l'étiquette du rein sera de  $4/6$  et de  $2/6$  dans la carte de probabilité de l'étiquette de la tumeur. Dans cet exemple, même si l'étiquette du rein a une probabilité supérieure pour le pixel en conflit, le CRF peut tout de même décider d'attribuer l'étiquette de la tumeur en prenant en considération les informations supplémentaires qu'il aura obtenues de l'image scanner par exemple.

### 4.3/ LES RÉSULTATS SUR NOS DIFFÉRENTES MÉTHODES DE GESTION DE CONFLITS

#### 4.3.1/ L'ACQUISITION DES DONNÉES

Dans cette section, nous présentons le processus d'acquisition des données dont nous avons besoin pour mener à bien la fusion de segmentations et pour pouvoir calculer la

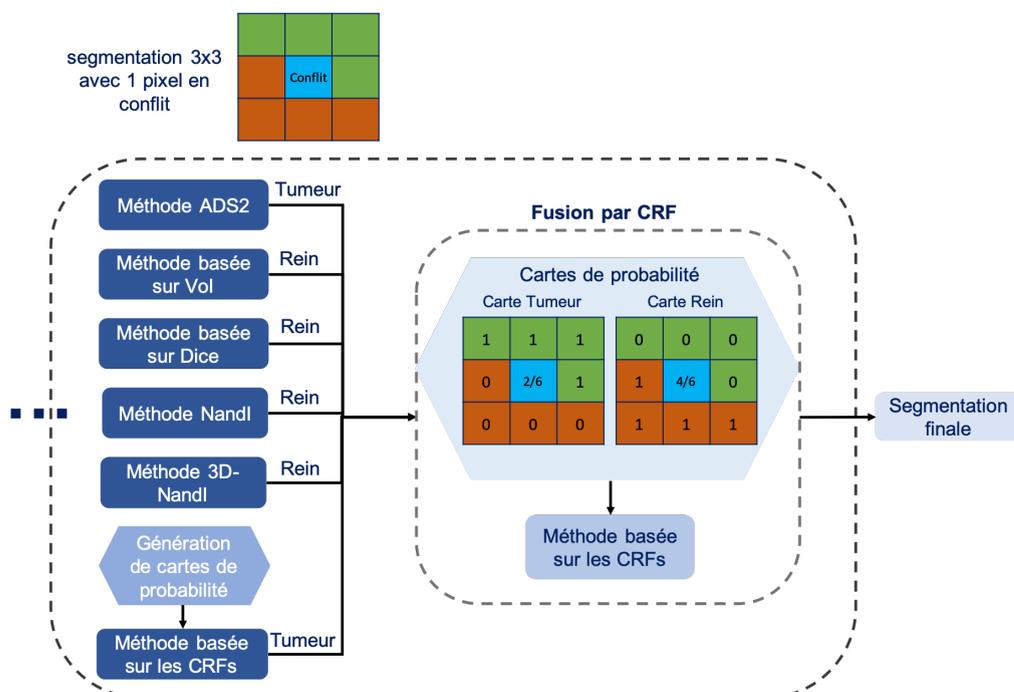


FIGURE 4.5 – Exemple de création de cartes de probabilités pour la fusion par CRF dans le cas d'un pixel en conflit dans une segmentation  $3 \times 3$ .

validité de nos résultats : que l'on peut nommer **consistence** (par anglicisme) par rapport à la vérité terrain obtenue par les médecins.

Notons que cette section nécessaire dans la compréhension de ce chapitre sera également utile dans les différentes autres contributions : les données recueillies servant de base à la fusion de segmentations seront les mêmes pour toutes nos autres contributions décrites dans la suite de ce document.

#### 4.3.1.1/ LES IMAGES MÉDICALES ET LEUR VÉRITÉ TERRAIN

Pour mener à bien les tests de fusion de segmentations, nous avons pu obtenir dans le cadre du projet SAIAD les données anonymisées de 12 patients atteints de néphroblastome, dont l'acquisition des images a eu lieu entre 2005 et 2018. Pour deux de ces patients, deux représentations scanner ont été réalisées à un mois d'intervalle, afin d'obtenir les données avant et après traitement (par chimiothérapie). La chimiothérapie a permis de réduire le volume de la tumeur. Comme la structure de la tumeur et du rein pathologique se retrouvent modifiées, nous avons décidé de prendre en considération chacune de ces acquisitions comme étant un patient unique. Nous avons donc pris la liberté de considérer au total un nombre de **14** patients. La moyenne d'âge de ces patients est de 67 mois, le plus jeune étant âgé de 13 mois et le plus âgé de 179 mois. Ces données comportent les images scanner en temps vasculaire de tous les patients, et pour uniquement 5 patients nous disposons également des images en temps tardif. L'examen en temps tardif n'est pas systématique, afin d'éviter l'irradiation inutile des jeunes patients quand une néphrectomie totale est préconisée. Cependant, si l'on souhaite réaliser une chirurgie conservatrice, il est indispensable de réaliser l'acquisition en temps tardif afin de visualiser s'il existe un envahissement des cavités urinaires (rappelons, cf. Chapitre

1, que si c'est le cas, une chirurgie conservatrice est contre-indiquée en raison du risque trop important de récurrence locale).

Les segmentations correspondant aux images scanner ont également été réalisées par les experts afin de servir de vérités terrain ainsi que de données pour l'entraînement de nos réseaux de neurones. Ces segmentations ont été réalisées semi-manuellement à partir du logiciel 3D-Slicer par les chirurgiens du CHRU de Besançon, en utilisant des techniques classiques mises à disposition par cet outil de segmentation par seuillage ou croissance de régions, puis en corrigeant manuellement ces segmentations. En moyenne, 8,6 heures ont été nécessaires pour la segmentation d'un patient. Nous n'utilisons pas ici les images scanner en temps tardif, compte tenu du fait que nous nous sommes focalisés sur la représentation du rein pathologique et de la tumeur.

De plus, les images scanner de la plupart des patients ont été segmentées une seconde fois par une radiologue, afin de calculer une différence **inter-opérateurs**, c'est-à-dire calculer la variabilité moyenne entre ces deux experts. Cette double segmentation permet d'une part de valider les vérités terrain par deux experts si la variabilité est moindre, et de l'autre de déterminer si la segmentation calculée est assez précise et pertinente pour être médicalement exploitable, avec une variabilité proche de la variabilité inter-opérateurs. La variabilité inter-opérateurs de la segmentation de la tumeur et du rein pathologique de la plupart des patients est présentée dans un rapport technique du CHRU lors d'une réunion SAIAD [CHRU2019] ainsi que dans la Table 4.1.

	<b>Segmentation de la tumeur</b>	<b>Segmentation du rein pathologique</b>	<b>Moyenne</b>
P1	0.9362	0.8277	0.8819
P2	0.9128	0.9573	0.9351
P3	X	X	X
P4	0.9496	0.9241	0.9369
P5	0.9234	0.8795	0.9015
P6	0.9606	0.8901	0.9254
P7	0.9562	0.8231	0.8897
P8	0.9398	0.9187	0.9293
P9	0.9734	0.7903	0.8819
P10	0.9544	0.6932	0.8238
P11	0.9618	0.9228	0.9423
P12	0.9569	0.8810	0.9190
P13	0.9450	0.8758	0.9104
P14	X	X	X
<b>Moyenne</b>	<b>0.9475</b>	<b>0.8653</b>	<b>0.9064</b>
<b>Écart type</b>	<b>0.0172</b>	<b>0.0724</b>	<b>0.0336</b>

TABLE 4.1 – Calcul de la variabilité inter-opérateurs entre deux experts en score de Dice, pour la segmentation de la tumeur et du rein pathologique en fonction des différents patients [CHRU2019].

Notons que les images des patients 3 et 14 n'ont pas encore bénéficié, pour l'heure, d'une double segmentation par les experts : sachant l'aspect chronophage de ce travail. Les résultats nous montrent qu'en moyenne, la variabilité pour la segmentation de la

tumeur est très faible, avec une similitude moyenne de 94,75%, en score de Dice, entre les deux jeux de segmentations. Cependant, la segmentation du rein pathologique est plus complexe à obtenir, la structure étant déformée, généralement petite et les contours de cette structure étant peu perceptibles sur l'image scanner car ses niveaux de gris sont assez similaires à ceux l'environnement. De ce fait, nous obtenons une plus grande variabilité inter-opérateurs pour cette structure, avec une similitude moyenne de 86,53%, ainsi qu'un écart type plus élevé.

#### 4.3.1.2/ L'ACQUISITION DES SEGMENTATIONS PAR INTELLIGENCE ARTIFICIELLE

Le rein pathologique et la tumeur rénale, présents dans les images scanner des 14 patients ont ensuite été préalablement segmentés, séparément, par un CNN de type FCN-8s (cf. Section 3.1.3) avec la méthode d'entraînement particulière par patient, développée au sein de notre équipe de recherche, nommée OV<sup>2</sup>ASSION [Marie2019] (dont le principe général est présenté Section 5.3). Cette méthode d'entraînement permet d'entraîner le réseau sur quelques coupes du patient, afin d'obtenir les segmentations des coupes restantes. Puis, un post-traitement par CRF a été appliqué afin d'améliorer les segmentations calculées. Pour obtenir les segmentations de ces deux structures pour chaque patient, un entraînement a été effectué par patient et pour chaque structure. Le réseau FCN-8s se base sur l'architecture VGG16, et pour mener à bien l'entraînement, tous les paramètres des couches similaires à VGG16 sont initialisés par les paramètres de l'entraînement pour la classification avec la base de données PASCAL VOC 2012<sup>1</sup> [Everingham2015]. Les paramètres des autres couches ont été instanciés par la valeur par défaut, soit 0. La dernière convolution est modifiée pour fonctionner avec deux canaux en sortie afin de prédire deux scores : l'arrière-plan et la structure. Pour les paramètres d'entraînement du réseau, le taux d'apprentissage (*learning rate*) est fixé à  $1e^{-12}$ , le momentum à 0,99 et nous utilisons 10000 itérations pour chaque entraînement.

Côté logiciel, l'implémentation de la segmentation est réalisée en Python à l'aide de la librairie Caffe dédiée au Deep Learning [Jia2014]. Au niveau du matériel, les calculs sont lancés sur un supercalculateur : le Mésocentre de l'Université de Franche-Comté, sur un GPU Tesla Kepler K40.

#### 4.3.1.3/ L'ÉVALUATION DES SEGMENTATIONS RÉSULTANTES

La Table 4.2 présente la **consistence** (par rapport à la vérité terrain obtenue par les médecins) des segmentations calculées du rein pathologique et de la tumeur, pour chaque patient noté de P1 à P14, en fonction de notre première vérité terrain calculée par les chirurgiens. Le score de Dice, une métrique très répandue dans le milieu médical (cf. Section 1.5.1.2), est utilisée pour calculer cette consistance. Ce score de Dice par patient a été obtenu en calculant le Dice global en prenant en compte toutes les coupes segmentées du patient. En moyenne, la segmentation de la tumeur est correcte à 91,54% et la segmentation du rein pathologique est correcte à 80,58%. En moyenne, avec les deux structures, le score de Dice obtenu est alors de 86,06%. Nous pouvons remarquer que le système de segmentation est plus efficace, en moyenne, sur la segmentation des

1. PASCAL VOC (Visual Object Classes Challenge) est un concours en traitement d'images dans la classification, la détection et la segmentation, fournissant une base de données d'images open source à utiliser.

tumeurs que sur le rein, avec une différence d'environ 9%, et que le système de segmentation est alors plus efficace sur des structures imposantes. De plus, la tumeur est souvent de forme ovale et de taille plus importante, ce qui permet une segmentation plus précise, alors que le rein, complètement comprimé, sera de taille plus petite et aura des formes plus aléatoires et moins régulières. Ainsi, les patients 3 et 14 par exemple, étant des patients ayant réalisé une chimiothérapie et dont la tumeur est alors de petite dimension, possède les scores de Dice les plus bas pour la tumeur à 80,66% et 84,13%. Le patient 10, ayant le rein pathologique le plus comprimé par la tumeur (et donc moins visible), possède le score de Dice pour cette structure le plus bas à 61,81%.

	<b>Segmentation de la tumeur</b>	<b>Segmentation du rein pathologique</b>	<b>Moyenne</b>
P1	0.9368	0.8553	0.8961
P2	0.8694	0.9115	0.8904
P3	0.8066	0.9346	0.8706
P4	0.9527	0.9187	0.9357
P5	0.9290	0.8433	0.8862
P6	0.9202	0.8295	0.8748
P7	0.9500	0.7254	0.8377
P8	0.9203	0.9087	0.9145
P9	0.9452	0.6686	0.8069
P10	0.9282	0.6181	0.7731
P11	0.9588	0.7353	0.8471
P12	0.9386	0.6376	0.7881
P13	0.9183	0.7693	0.8438
P14	0.8413	0.9258	0.8835
<b>Moyenne</b>	<b>0.9154</b>	<b>0.8058</b>	<b>0.8606</b>
<b>Écart type</b>	<b>0.0449</b>	<b>0.1127</b>	<b>0.0472</b>

TABLE 4.2 – Résultats de la segmentation (avec le score de Dice) de la tumeur et du rein pathologique, par le réseau FCN-8s avec la méthode d'apprentissage OV<sup>2</sup>ASSION.

En comparaison avec les variabilités inter-opérateurs, nous obtenons un score de Dice proche mais inférieur d'environ 3% pour la segmentation de la tumeur, et d'environ 6% pour la segmentation du rein pathologique. En moyenne pour les deux structures, nous obtenons un score total de 86,06% contre 90,64% en inter-opérateurs. Ces segmentations pourraient alors être améliorées afin de s'approcher encore plus de ces valeurs inter-opérateurs. D'autres solutions sont en cours pour segmenter le rein pathologique avec d'autres méthodes d'IA au sein du projet SAIAD, afin d'améliorer la qualité des segmentations.

Afin d'obtenir la segmentation finale, il reste alors à agréger les segmentations tumorales et rénales en gérant les éventuels conflits.

### 4.3.2/ LE CALCUL DES PARAMÈTRES OPTIMAUX POUR LA GESTION DES CONFLITS

Nous avons tout d'abord déterminé les paramètres optimaux pour chaque méthode de gestion de conflits. En particulier, nous avons déterminé le nombre optimal de coupes adjacentes pour les méthodes basées sur le Vol et le score de Dice, et le rayon à prendre en compte pour les méthodes Nandl et 3D-Nandl. Au début du projet, nous ne disposions des images scanner et des segmentations correspondantes que de quelques patients. C'est pourquoi, pour ces méthodes, nous avons effectué des tests sur trois patients en faisant varier de manière empirique les paramètres.

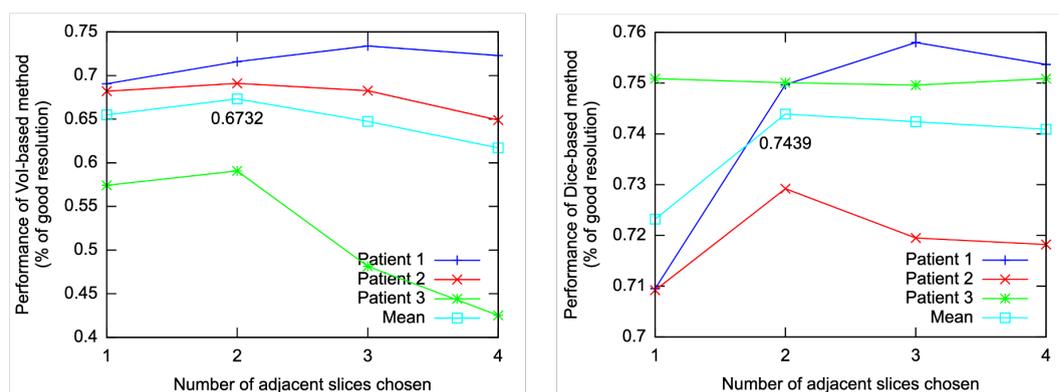


FIGURE 4.6 – Évolution de la performance des méthodes basées sur le Vol et le score de Dice par rapport au nombre de coupes adjacentes considérées.

La Figure 4.6 montre les résultats obtenus pour les méthodes basées sur le Vol et le score de Dice en faisant varier le nombre de coupes adjacentes prises en compte dans les calculs. Pour ces trois patients, la meilleure performance moyenne de ces deux méthodes est obtenue en considérant **2 coupes** adjacentes de chaque côté de la segmentation courante. Ces tests ont également été effectués avec un nombre d'itérations  $T_{max} = 3$  pour leur algorithme de type ICM. En effet, peu d'itérations suffisent pour faire converger le système, car la segmentation initiale pour laquelle il est nécessaire de gérer les conflits est très proche de la segmentation attendue. Les seuls pixels différents entre ces segmentations sont ceux qui sont en conflit : en nombre relativement faible par rapport au nombre total de pixels dans l'image.

Les performances des méthodes Nandl et 3D-Nandl ont été évaluées en faisant varier la taille du rayon (rayon utilisé pour calculer les pixels appartenant au voisinage effectif des pixels en conflit). Les résultats des deux méthodes sont présentés sur la Figure 4.7. En conséquence, **un rayon égal à 3** a été retenu pour la méthode Nandl et un **rayon de 5** pour la méthode 3D-Nandl, pour laquelle sa performance moyenne stagne à partir de cette valeur.

### 4.3.3/ LES RÉSULTATS SUR LES DEUX MODÈLES

Comme expliqué dans la Section 4.1.1, la méthode ADS2 ne peut pas résoudre tous les conflits. Certains pixels sont donc encore étiquetés comme étant en conflit à la suite de l'application d'ADS2. La Table 4.3 présente le pourcentage de pixels en conflit pouvant être résolu par cette méthode, pour chaque patient. En moyenne, seulement 20,58% des pixels en conflit peuvent être résolus par cette méthode sur l'ensemble des patients.

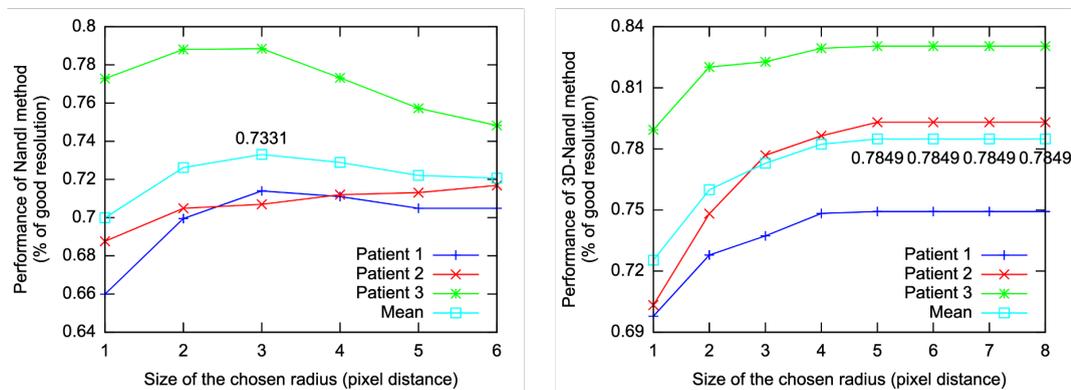


FIGURE 4.7 – Évolution de la performance des méthodes Nandl et 3D-Nandl par rapport à la taille du rayon choisi.

Méthode ADS2	P1	P2	P3	P4	P5	P6	P7	P8
% des pixels résolus	25.95	15.02	15.77	20.62	25.71	30.18	23.57	7.76
	<b>P9</b>	<b>P10</b>	<b>P11</b>	<b>P12</b>	<b>P13</b>	<b>P14</b>	<b>Moyenne</b>	
	10.01	14.34	50.93	11.25	23.07	13.87	<b>20.58</b>	

TABLE 4.3 – Pourcentage de pixels en conflit pouvant être résolus par la méthode ADS2 sur l'ensemble de nos patients.

La suite présente les résultats obtenus pour la bonne gestion des pixels en conflit de notre premier modèle puis du second, ainsi que l'amélioration constatée de la qualité des segmentations.

#### 4.3.3.1/ LES RÉSULTATS DU PREMIER MODÈLE

La Table 4.4 présente les performances des six méthodes utilisées en parallèle avec le premier modèle de gestion des conflits (cf. Figure 4.3), sur nos 14 patients. Les résultats sont composés de deux parties :

- La première correspond au pourcentage de pixels en conflit qui ont été correctement étiquetés,
- et la deuxième, entre parenthèses, correspond au nombre total de pixels en conflit correctement étiquetés sur l'ensemble des pixels en conflit du patient (sauf dans le cas de la méthode ADS2 pour laquelle il s'agit des pixels en conflit pouvant être résolus par cette méthode).

La moyenne de performance pour chaque méthode est calculée par rapport à l'ensemble des pixels en conflit sur tous les patients et non par rapport à la moyenne obtenue pour chaque patient : il en est de même pour les tables qui suivent dans ce document. En moyenne, sur l'ensemble des patients, la méthode ADS2 obtient 87,06% de bonne résolution de conflits, ce qui en fait une méthode performante à son niveau, compte tenu du fait qu'elle ne peut en traiter en moyenne que 20,58%. Les résultats moyens des autres méthodes sont plus faibles, mais avec tout de même 71,01% de bonnes résolutions pour la méthode 3D-Nandl, ce qui en fait la méthode la plus performante pour les méthodes

	<b>ADS2</b>	<b>Vol</b>	<b>Dice</b>	<b>CRF</b>	<b>NandI</b>	<b>3D-NandI</b>
P1	0.8757 (2551/2913)	0.6910 (7757/11226)	0.7483 (8400/11226)	0.6507 (7305/11226)	0.6190 (6949/11226)	0.6806 (7640/11226)
P2	0.7979 (304/381)	0.6803 (1726/2537)	0.6575 (1668/2537)	0.7166 (1818/2537)	0.5956 (1511/2537)	0.6693 (1698/2537)
P3	0.5789 (22/38)	0.5104 (123/241)	0.6971 (168/241)	0.9212 (222/241)	0.6556 (158/241)	0.7220 (174/241)
P4	0.9347 (744/796)	0.6839 (2640/3860)	0.5858 (2261/3860)	0.7663 (2958/3860)	0.6738 (2601/3860)	0.7298 (2817/3860)
P5	0.8173 (1051/1286)	0.7283 (3642/5001)	0.6825 (3413/5001)	0.6517 (3259/5001)	0.6101 (3051/5001)	0.7017 (3509/5001)
P6	0.9042 (1463/1618)	0.6871 (3684/5362)	0.7699 (4128/5362)	0.7072 (3792/5362)	0.7027 (3768/5362)	0.6938 (3720/5362)
P7	0.8267 (544/658)	0.6683 (1866/2792)	0.5670 (1583/2792)	0.6999 (1954/2792)	0.6239 (1742/2792)	0.7400 (2066/2792)
P8	0.7750 (93/120)	0.2937 (454/1546)	0.4386 (678/1546)	0.5640 (872/1546)	0.3959 (612/1546)	0.5045 (780/1546)
P9	0.9519 (297/312)	0.6630 (2066/3116)	0.6694 (2086/3116)	0.7924 (2469/3116)	0.7908 (2464/3116)	0.7994 (2491/3116)
P10	0.8704 (282/324)	0.6038 (1364/2259)	0.5777 (1305/2259)	0.6206 (1402/2259)	0.6029 (1362/2259)	0.7127 (1610/2259)
P11	0.9733 (401/412)	0.7936 (642/809)	0.7874 (637/809)	0.6131 (496/809)	0.6131 (496/809)	0.8022 (649/809)
P12	0.7966 (188/236)	0.5432 (1139/2097)	0.5460 (1145/2097)	0.5994 (1257/2097)	0.6419 (1346/2097)	0.6819 (1430/2097)
P13	0.8932 (2928/3278)	0.6148 (8735/14207)	0.7157 (10167/14207)	0.6866 (9754/14207)	0.6707 (9529/14207)	0.7411 (10529/14207)
P14	0.7863 (905/1151)	0.6612 (5486/8297)	0.6055 (5024/8297)	0.7088 (5881/8297)	0.7029 (5832/8297)	0.7075 (5870/8297)
<b>Moyenne</b>	<b>0.8706</b>	<b>0.6523</b>	<b>0.6734</b>	<b>0.6857</b>	<b>0.6538</b>	<b>0.7101</b>
<b>Écart type</b>	<b>0.0992</b>	<b>0.1203</b>	<b>0.0979</b>	<b>0.0915</b>	<b>0.0868</b>	<b>0.0705</b>

TABLE 4.4 – Performance moyenne des différentes méthodes sur plusieurs patients avec la première version du modèle de gestion des conflits.

traitant l'ensemble des conflits, suivie par la méthode basée sur le CRF avec 68,57%, tous deux ayant également de faibles écarts types par rapport à d'autres méthodes.

La Table 4.5 présente les résultats finaux du premier modèle, avec la fusion des résultats des six méthodes par vote majoritaire (marqué comme "Fusion VM") et la fusion basée sur le CRF. Comme dans ce premier modèle, il est possible d'avoir une égalité au niveau des résultats, il existe deux options de fusion par vote majoritaire : la première option consiste à privilégier la tumeur et donc à attribuer les pixels à l'étiquette de la tumeur en cas d'égalité, et l'autre option consiste à privilégier le rein en attribuant les pixels à l'étiquette du rein pathologique. Nous pouvons constater que d'une part la fusion des résultats par vote majoritaire est plus performante que la fusion basée sur CRF (avec en moyenne 71,96% et 71,85% contre 70,41%, ainsi que des écarts types plus faibles), et que d'autre part nous obtenons de meilleurs résultats en attribuant l'étiquette de la tumeur pour les pixels en conflit se trouvant dans une égalité lors du vote majoritaire. Cette étape de fusion des résultats par vote majoritaire permet alors d'obtenir un pourcentage plus élevé de bonnes résolutions de conflits que par chacune des méthodes individuellement. En complément, nous pouvons remarquer que la méthode 3D-NandI permet d'obtenir de meilleurs résultats (71.01%) que l'ensemble des méthodes le tout fusionné par la méthode basée sur les CRFs (70.41%).

	<b>Fusion VM</b> (Priorité tumeur en cas d'égalité)	<b>Fusion VM</b> (Priorité rein en cas d'égalité)	<b>Fusion basée</b> <b>sur CRF</b>
P1	0.7289 (8183/11226)	0.7289 (8183/11226)	0.6748 (7575/11226)
P2	0.7067 (1793/2537)	0.7016 (1780/2537)	0.6855 (1739/2537)
P3	0.7386 (178/241)	0.7386 (178/241)	0.7635 (184/241)
P4	0.7339 (2833/3860)	0.7324 (2827/3860)	0.7254 (2800/3860)
P5	0.7285 (3643/5001)	0.7231 (3616/5001)	0.7009 (3505/5001)
P6	0.7512 (4028/5362)	0.7507 (4025/5362)	0.7411 (3974/5362)
P7	0.7138 (1993/2792)	0.7045 (1967/2792)	0.7110 (1985/2792)
P8	0.4340 (671/1546)	0.4334 (670/1546)	0.4340 (671/1546)
P9	0.8081 (2518/3116)	0.8097 (2523/3116)	0.8306 (2588/3116)
P10	0.6560 (1482/2259)	0.6547 (1479/2259)	0.6436 (1454/2259)
P11	0.7874 (637/809)	0.7812 (632/809)	0.7763 (628/809)
P12	0.6433 (1349/2097)	0.6381 (1338/2097)	0.6519 (1367/2097)
P13	0.7242 (10289/14207)	0.7247 (10296/14207)	0.7153 (10162/14207)
P14	0.7221 (5991/8297)	0.7236 (6004/8297)	0.7199 (5973/8297)
<b>Moyenne</b>	<b>0.7196</b>	<b>0.7185</b>	<b>0.7041</b>
<b>Écart type</b>	<b>0.0890</b>	<b>0.0891</b>	<b>0.0907</b>

TABLE 4.5 – Performances finales pour le premier modèle sur différents patients avec les méthodes de fusion par vote majoritaire et basée sur le CRF.

#### 4.3.3.2/ LES RÉSULTATS DU SECOND MODÈLE

La Table 4.6 présente les résultats des différentes méthodes de gestion des conflits, avec cette fois-ci le deuxième modèle (cf. Figure 4.4), en utilisant la méthode ADS2 comme pré-traitement. Les autres méthodes gèrent alors moins de pixels en conflit. Les résultats de la méthode ADS2 ne changent donc pas par rapport au premier modèle, mais les résultats de chacune des autres méthodes, individuellement, baissent avec néanmoins toujours un meilleur résultat en moyenne pour la méthode 3D-Nandi avec 67% de bonnes résolutions de conflits, suivie de peu par la méthode basée sur le CRF avec 66,96%. Nous pouvons expliquer la baisse des résultats des autres méthodes par le fait que la méthode ADS2 gère préalablement certains conflits pouvant être facilement correctement étiquetés. Les cinq autres méthodes doivent alors gérer les conflits restants, moins évidents.

La Table 4.7 présente les résultats finaux du second modèle, avec la fusion des résultats

	<b>ADS2</b>	<b>Vol</b>	<b>Dice</b>	<b>CRF</b>	<b>Nandl</b>	<b>3D-Nandl</b>
P1	0.8757 (2551/2913)	0.6391 (5313/8313)	0.7043 (5855/8313)	0.6472 (5380/8313)	0.6148 (5111/8313)	0.6208 (5161/8313)
P2	0.7979 (304/381)	0.6568 (1416/2156)	0.6424 (1385/2156)	0.6735 (1452/2156)	0.5849 (1261/2156)	0.6447 (1390/2156)
P3	0.5789 (22/38)	0.4828 (98/203)	0.7291 (148/203)	0.9212 (187/203)	0.6453 (131/203)	0.7389 (150/203)
P4	0.9347 (744/796)	0.6191 (1897/3064)	0.4951 (1517/3064)	7471 (2289/3064)	0.6345 (1944/3064)	0.6775 (2076/3064)
P5	0.8173 (1051/1286)	0.6958 (2585/3715)	0.6320 (2348/3715)	0.6493 (2412/3715)	0.6027 (2239/3715)	0.6681 (2482/3715)
P6	0.9042 (1463/1618)	0.5946 (2226/3744)	0.7113 (2663/3744)	0.6931 (2595/3744)	0.6819 (2553/3744)	0.6191 (2318/3744)
P7	0.8267 (544/658)	0.6293 (1343/2134)	0.4878 (1041/2134)	0.6813 (1454/2134)	0.6223 (1328/2134)	0.7090 (1513/2134)
P8	0.7750 (93/120)	0.2539 (362/1426)	0.4102 (585/1426)	0.5372 (766/1426)	0.3731 (532/1426)	0.4804 (685/1426)
P9	0.9519 (297/312)	0.6373 (1787/2804)	0.6380 (1789/2804)	0.7728 (2167/2804)	0.7721 (2165/2804)	0.7785 (2183/2804)
P10	0.8704 (282/324)	0.5566 (1077/1935)	0.5318 (1029/1935)	0.6165 (1193/1935)	0.6083 (1177/1935)	0.6879 (1331/1935)
P11	0.9733 (401/412)	0.6145 (240/397)	0.5945 (236/397)	0.4685 (186/397)	0.6121 (243/397)	0.6448 (256/397)
P12	0.7966 (188/236)	0.5094 (948/1861)	0.5121 (953/1861)	0.5852 (1089/1861)	0.6378 (1187/1861)	0.6679 (1243/1861)
P13	0.8932 (2928/3278)	0.5392 (5893/10929)	0.6671 (7291/10929)	0.6565 (7175/10929)	0.6573 (7184/10929)	0.6960 (7607/10929)
P14	0.7863 (905/1151)	0.6437 (4600/7146)	0.5739 (4101/7146)	0.7021 (5017/7146)	0.6920 (4945/7146)	0.6982 (4989/7146)
<b>Moyenne</b>	<b>0.8706</b>	<b>0.5978</b>	<b>0.6210</b>	<b>0.6696</b>	<b>0.6422</b>	<b>0.6700</b>
<b>Écart type</b>	<b>0.0992</b>	<b>0.1104</b>	<b>0.0965</b>	<b>0.1077</b>	<b>0.0865</b>	<b>0.0690</b>

TABLE 4.6 – Performance moyenne des différentes méthodes sur plusieurs patients avec la deuxième version du modèle de gestion des conflits.

des cinq méthodes (basée sur Vol, sur le score de Dice, et le CRF et les méthodes Nandl et 3D-Nandl) par vote majoritaire et basée sur CRF. Contrairement au premier modèle, nous ne pouvons plus avoir d'égalité lors de la fusion par vote majoritaire. Nous obtenons toujours en moyenne de meilleurs résultats par vote majoritaire (71,95% contre 70,85% pour la fusion basée sur le CRF) et ces résultats finaux sont similaires aux résultats finaux du premier modèle, avec seulement 0,01% de différence entre la fusion par vote majoritaire des deux modèles (en prenant en compte le vote majoritaire avec priorité pour l'étiquette de la tumeur dans le premier modèle). La fusion basée sur le CRF est même plus performante pour ce second modèle mais avec un écart type légèrement plus élevé.

#### 4.3.3.3/ L'AMÉLIORATION DES SEGMENTATIONS

La résolution de ces conflits permet également d'améliorer les segmentations. La Table 4.8 présente l'amélioration des segmentations selon le score de Dice (cf. Section 1.5.1.2). Avant la fusion, nous avons calculé le score de Dice obtenu pour chaque structure (tumeur et rein pathologique), le tout pour les coupes segmentées de chaque patient. À cette étape, comme nous n'avons pas encore fusionné les segmentations des deux structures et que les pixels en conflit sont étiquetés comme appartenant à ces structures dans cha-

	<b>Fusion VM</b>	<b>Fusion basée sur CRF</b>
P1	0.7296 (8191/11226)	0.6857 (7698/11226)
P2	0.6989 (1773/2537)	0.6717 (1704/2537)
P3	0.7220 (174/241)	0.7469 (180/241)
P4	0.7332 (2830/3860)	0.7277 (2809/3860)
P5	0.7213 (3607/5001)	0.7069 (3535/5001)
P6	0.7486 (4014/5362)	0.7456 (3998/5362)
P7	0.7138 (1993/2792)	0.7228 (2018/2792)
P8	0.4295 (664/1546)	0.4308 (666/1546)
P9	0.8087 (2520/3116)	0.8280 (2580/3116)
P10	0.6618 (1495/2259)	0.6627 (1497/2259)
P11	0.7911 (640/809)	0.7948 (643/809)
P12	0.6376 (1337/2097)	0.6605 (1385/2097)
P13	0.7282 (10346/14207)	0.7203 (10233/14207)
P14	0.7224 (5994/8297)	0.7154 (5936/8297)
<b>Moyenne</b>	<b>0.7195</b>	<b>0.7085</b>
<b>Écart type</b>	<b>0.0900</b>	<b>0.0912</b>

TABLE 4.7 – Performances finales pour le deuxième modèle sur différents patients avec les méthodes de fusion par vote majoritaire et basée sur le CRF.

<b>Type de fusion</b>	<b>Dice avant Fusion</b>	<b>Dice après le 1er modèle</b>			<b>Dice après le 2ème modèle</b>	
		<b>VM (Tumeur)</b>	<b>VM (Rein)</b>	<b>CRF</b>	<b>VM</b>	<b>CRF</b>
Moyenne	<b>0.8606</b>	0.8611	0.8610	0.8609	0.8611	0.8610
Écart type	<b>0.0472</b>	0.0482	0.0482	0.0483	0.0481	0.0480

TABLE 4.8 – Présentation du Dice moyen (moyenne sur les 14 patients) des segmentations avant et après la gestion des conflits.

cune des différentes segmentations, le score de Dice pour chaque structure peut être calculé sans que cela ne pose de problème. Ensuite, le score de Dice moyen d'un patient est simplement la moyenne entre le Dice calculé de la tumeur et du rein pathologique. Ces résultats préliminaires ont été précédemment présentés dans la Section 4.3.1. Pour des raisons de lisibilité, la Table 4.8 ne représente pas le score de Dice de chaque patient mais représente alors le score de Dice moyen final obtenu de l'ensemble des 14

patients. Ce score de Dice final obtenu avant la fusion pour l'ensemble des patients était alors de 86,06%. Après la fusion et la gestion des conflits, quel que soit le modèle appliqué, le score de Dice a augmenté de 0,03% à 0,05%. Il est important de considérer que le nombre de pixels en conflit est très faible par rapport à l'ensemble des pixels de nos segmentations de taille  $512 \times 512$ , expliquant ainsi que cette amélioration finale soit de cet ordre. En moyenne, en prenant en compte les 14 patients à notre disposition, pour une coupe, seulement 0,02% des pixels seront en conflit. C'est pourquoi, en modifiant seulement les étiquettes des pixels en conflit, le score de Dice ne peut pas être considérablement modifié. Il est également à noter que l'écart type a faiblement augmenté, mais reste toujours dans des valeurs convenables. L'amélioration de ces segmentations n'est alors pas encore assez significative pour permettre de s'approcher des résultats inter-opérateurs, afin de rendre ces segmentations potentiellement exploitables du point de vue médical.

#### 4.3.3.4/ LA PRÉSENTATION VISUELLE D'UN RÉSULTAT

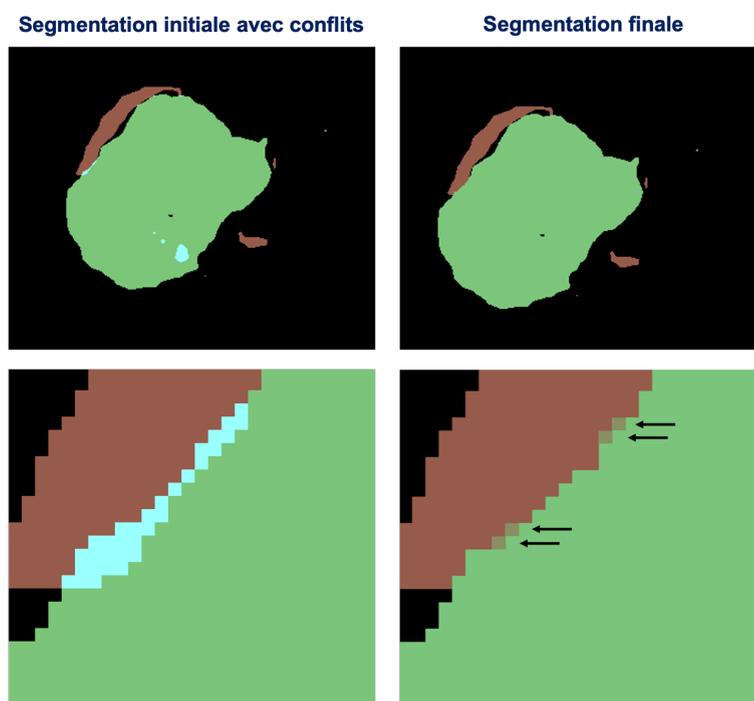


FIGURE 4.8 – Exemple d'un résultat de fusion de segmentations d'une coupe d'un patient, avec un avant/après de la coupe ainsi qu'un zoom sur une zone conflictuelle.

Le résultat de la gestion des conflits d'une segmentation d'un patient est présenté dans la Figure 4.8. Les deux premières images correspondent à la segmentation avant et après la gestion des conflits, et les deux images en dessous représentent les mêmes segmentations zoomées au niveau de la frontière problématique entre le rein et la tumeur. La segmentation initiale à gauche présente des conflits (en couleur bleue claire) à l'intersection entre le rein et la tumeur et à l'intérieur de la tumeur. Ces derniers conflits sont relativement évidents à résoudre, car ils appartiennent généralement à la structure qui les entoure. Les deux modèles de gestion de conflits donnent alors le même résultat pour ces conflits. Cependant, les conflits se trouvant à l'intersection des deux structures sont plus

difficiles à résoudre, et c'est là que les propositions divergent en fonction de la méthode de fusion appliquée. Quatre pixels (indiqués par les flèches) ont été étiquetés comme appartenant à la tumeur avec la fusion par vote majoritaire dans les deux modèles, et étiquetés comme appartenant au rein par la fusion avec la méthode basée sur CRF dans les deux modèles également.

## 4.4/ DISCUSSIONS

### 4.4.1/ L'EFFICACITÉ ET LA COMPARAISON DES DIFFÉRENTES MÉTHODES UTILISÉES

Les résultats montrent qu'en moyenne la combinaison de plusieurs méthodes de fusion donne de meilleurs résultats que chaque méthode prise individuellement. En effet, le fait d'obtenir des méthodes complémentaires exploitant différents types d'informations (comme la localisation, l'utilisation de statistiques et de probabilités, des calculs de similarités, ...) et différents types et nombre d'images (comme l'utilisation de la segmentation en conflit uniquement, ou des segmentations adjacentes avec leur image scanner, ...) améliorent les performances.

	La seg. Actuelle	Les seg. Adjacentes	Scan Actuel	Scans Adjacents	Carte de Probabilités
ADS2	✓	✓ (1)	✗	✗	✗
Vol	✓	✓ (2)	✗	✗	✗
Dice	✓	✓ (2)	✗	✗	✗
CRF	✗	✗	✓	✗	✓
NandI	✓	✗	✓	✗	✗
3D-NandI	✓	✓ (5)	✓	✓ (5)	✗

TABLE 4.9 – Comparaison des différentes images prises en compte dans le processus de gestion des conflits pour chaque méthode.

Entre chaque méthode, nous avons pu constater des différences de performances. La Table 4.9 présente les différents types d'image pris en compte pour la gestion des conflits en fonction des méthodes. Les chiffres entre parenthèses représentent le nombre de coupes adjacentes (de chaque côté) prises en compte. Nous avons constaté que la méthode basée sur le score de Dice est plus performante que celle basée sur le Vol, alors qu'elles exploitent les mêmes images. Ceci est dû au fait que la méthode basée sur le score de Dice associe en plus une pondération aux différentes segmentations. Les méthodes basées sur le CRF, NandI et 3D-NandI utilisent également les images scanner et exploitent donc plus d'informations. Elles sont donc, comme attendu, plus performantes, à l'exception de la méthode NandI dans certains cas, car elle n'exploite que les informations de la coupe à traiter. Comme pour la méthode NandI, la méthode basée sur le CRF utilise seulement les informations de la coupe actuelle. Elle est cependant plus performante car elle utilise la carte de probabilités de la segmentation, qui constitue une source d'informations complémentaires utiles et pertinentes. La méthode 3D-NandI est celle exploitant le plus d'images, de segmentations et d'images scanner, et est alors la méthode gérant la totalité des pixels en conflit la plus performante. La méthode ADS2

est une méthode particulière car même si elle n'utilise que peu d'images, elle donne de très bons résultats. Elle ne peut être comparée aux autres méthodes car elle ne peut être appliquée à l'ensemble des pixels en conflit.

Il pourrait alors être intéressant de pondérer les différentes méthodes en privilégiant celles utilisant un nombre plus élevé d'images, ou alors en pondérant en fonction de la qualité des images utilisées. Par exemple, si les images scanner sont trop bruyantes, il serait préférable d'accorder plus d'importance aux méthodes qui n'exploitent que les segmentations. Inversement, avec des segmentations très éloignées des vérités terrain, il serait préférable de donner une pondération plus importante aux méthodes exploitant les images scanner.

#### 4.4.2/ LE MEILLEUR MODÈLE DE GESTION DES CONFLITS À UTILISER

Les performances des deux modèles de gestion des conflits proposés sont plus ou moins équivalentes : aucun modèle ne se démarque véritablement sur les 14 dossiers patients que nous avons utilisés pour nos tests. Deux méthodes de fusion sont proposées dans les modèles mais la fusion par vote majoritaire se démarque de la fusion par CRF. Cependant, l'utilisation de la méthode basée sur le CRF permet de ne pas rencontrer de cas d'indécision en cas d'égalité avec un nombre pair de résultats (comme c'est le cas en utilisant le vote majoritaire). Le fait de privilégier la méthode ADS2 dans le deuxième modèle ne semble pas avoir d'impact significatif sur les résultats finaux. Enfin, ces modèles proposés sont flexibles et peuvent être adaptés en fonction des besoins. Il est alors facilement possible d'ajouter ou de supprimer des méthodes de gestion de conflits, de les pondérer et d'utiliser une méthode de fusion adéquate.

#### 4.4.3/ LES DIFFÉRENTS DEGRÉS DE DIFFICULTÉ DE RÉOLUTION DE CONFLITS

Nous avons remarqué que certains conflits étaient plus évidents à résoudre que d'autres. En effet, il est plus facile d'arbitrer l'étiquette des pixels en conflit se trouvant à l'intérieur d'une structure, car ces pixels appartiennent en général à la structure qui les entoure. Ils résultent principalement d'artefacts obtenus par le processus de segmentation. Ces artefacts sont en général uniquement présents sur la coupe considérée lors du processus de fusion et il est possible de déterminer simplement les étiquettes correctes de ces pixels par leur intensité de gris dans l'image scanner. En revanche, les conflits relevant de pixels se trouvant à l'intersection entre deux structures sont moins simples à résoudre. Parfois, les informations sur les images scanner et sur les segmentations ne suffisent pas pour trouver une délimitation claire. Dans ces derniers cas, même les praticiens (radiologues et chirurgiens) utilisent leur expérience pour délimiter ces zones lors de la réalisation de segmentations manuelles. Malheureusement, ce sont précisément les zones dans lesquelles se trouvent la plupart des pixels en conflit. C'est pourquoi il est difficile d'obtenir un pourcentage plus élevé de bonne résolution. Les méthodes que nous avons conçues et présentées ont toutefois permis d'améliorer la qualité des segmentations des dossiers des 14 patients sur lesquelles elles ont été testées.

Comme nous l'avons remarqué dans nos échanges avec les chirurgiens et radiologues du CHRU de Besançon, les experts se basent sur leur propre expérience pour arbitrer les conflits, en particulier pour les pixels les plus compliqués à étiqueter. Nous nous sommes donc également intéressés à des méthodes du paradigme de l'Intelligence Artificielle,

Deep Learning et Raisonnement à Partir de Cas en l'occurrence, pour améliorer encore cette étape de fusion. Ces outils ont en effet l'avantage d'être basées sur l'expérience et intègrent d'autres types de connaissances.

**Synthèse du chapitre****SYNTHÈSE**

Dans ce chapitre, nous avons présenté notre première contribution pour la fusion de segmentations complémentaires et la gestion des conflits en résultant. Deux modèles différents de gestion de pixels en conflit ont été implémentés et sont composés de deux étapes : la première étape est l'utilisation de six méthodes différentes qui ont été créées ou adaptées pour déterminer les étiquettes des pixels, et la seconde étape est l'utilisation d'une méthode de fusion des résultats obtenues par ces différentes méthodes, permettant ainsi d'obtenir la segmentation consensus dépourvue de conflits.

Les méthodes de gestion des conflits sont basées sur différents critères et utilisent différentes informations afin d'être complémentaires.

1. Ainsi, la méthode ADS2 utilise les informations des segmentations adjacentes.
2. Les méthodes basées sur le Vol et le score de Dice utilisent ces deux métriques avec l'utilisation d'un algorithme ICM.
3. La méthode NandI et 3D-NandI utilisent un calcul de similarité.
4. Et enfin, la méthode basée sur CRF utilise l'approche probabiliste des CRFs avec les cartes de probabilités pour gérer les conflits.

C'est sur l'utilisation de ces méthodes que diffèrent les deux modèles. Le premier modèle utilise les six méthodes en parallèle, alors que le deuxième utilise la méthode ADS2 en amont, suivie des cinq autres.

Nous avons ensuite proposé deux méthodes de fusion de ces résultats : une méthode par vote majoritaire, et une méthode basée sur le CRF. Même si la méthode de fusion par vote majoritaire est plus performante, il est tout de même intéressant d'utiliser dans certains cas la méthode par le CRF afin d'éviter les cas d'égalités pouvant survenir par vote majoritaire.

Ces deux modèles ont été testés sur la fusion de segmentations de rein pathologique et de néphroblastome, qui ont été calculées par Deep Learning avec la méthode d'apprentissage particulière appelée OV<sup>2</sup>ASSION (développée dans la thèse de F. Marie [Marie2019]). Les résultats démontrent une bonne gestion des pixels en conflit (presque 72% de bonnes résolutions de conflits) ainsi qu'une légère amélioration de la qualité des segmentations.

Dans le chapitre suivant, nous présentons une deuxième contribution sur la fusion de segmentations complémentaires basée sur des techniques d'Intelligence Artificielle, et plus particulièrement par le Deep Learning et le Raisonnement à Partir de Cas.



## MÉTHODES DE FUSION DE SEGMENTATIONS BASÉES SUR L'INTELLIGENCE ARTIFICIELLE

Dans ce chapitre, nous proposons deux approches originales basées sur l'utilisation de techniques d'Intelligence Artificielle (IA), et plus particulièrement sur l'apprentissage profond (*Deep Learning*) avec l'utilisation de réseaux de type CNN (*Convolutional Neural Networks*), pour réaliser la fusion de segmentations de reins déformés et des tumeurs. Dans notre première méthode, nous combinons au CNN une autre méthode d'IA : le Raisonnement à Partir de Cas (RàPC), ce qui nous permet d'apporter de la connaissance avec une base de cas. Notre deuxième approche se concentre sur le type d'entraînement effectué par le CNN avec l'utilisation de la méthode OVASSION : cette approche initialement prévue pour la segmentation d'images, est appliquée ici dans le cadre de la fusion de segmentations complémentaires.

### 5.1/ L'ARCHITECTURE DU RÉSEAU DE NEURONES CRÉÉE POUR LA FUSION

Deux approches basées sur l'apprentissage profond ont été réalisées dans le but spécifique de fusionner des segmentations complémentaires. Ces deux approches utilisent le même réseau de neurones, spécialement conçu, dont l'architecture générale est présentée par la Figure 5.1. Ce CNN est inspiré de celui utilisé pour la segmentation d'images et plus particulièrement des réseaux FCN (*Fully Convolutional Networks*) utilisés par Long et al. [Long2015]. Ce nouveau réseau est composé de 8 couches de convolution et de 2 couches de *pooling*. Il consiste en l'application de deux convolutions  $3 \times 3$ , chaque convolution étant suivie d'une couche de correction ReLU, et d'une couche de *max pooling*  $2 \times 2$  avec un pas de 2 pour le sous-échantillonnage, le tout répété deux fois.

Puis, deux autres convolutions  $3 \times 3$  sans remplissage (*padding*) avec ReLU sont appliquées et suivies d'une convolution  $1 \times 1$  avec une réduction des *features maps*, ou cartes de caractéristiques, à 3 afin d'obtenir le nombre de classes souhaité. Ici, nous avons besoin de trois types d'étiquettes, correspondant à l'arrière-plan de l'image (*background*), au rein pathologique et à la tumeur. Enfin, une couche de déconvolution  $20 \times 20$  est utilisée pour obtenir un résultat de même taille que l'entrée du réseau. À la fin de

la phase de prédiction, une fonction  $\arg \max$  est utilisée afin d'obtenir une segmentation finale fusionnée à un seul canal.

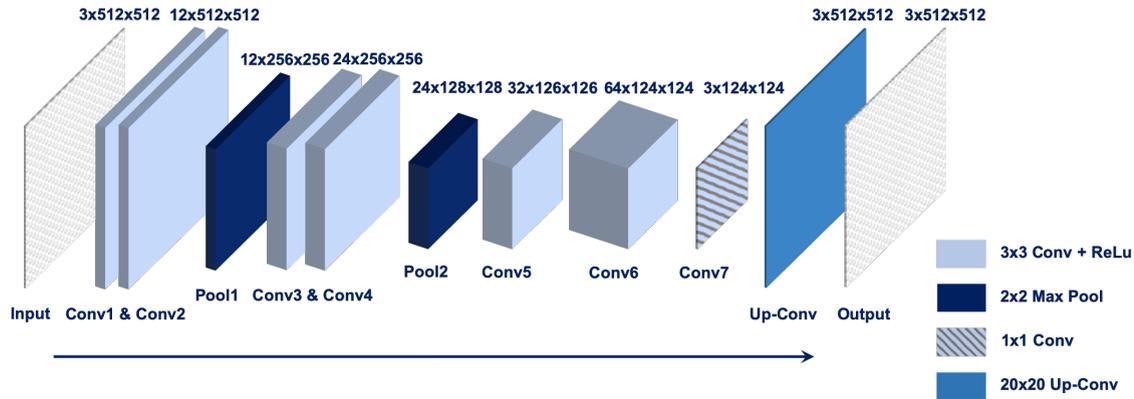


FIGURE 5.1 – L'architecture générale du réseau de neurones proposée pour la fusion de segmentations complémentaires.

Nous avons également choisi d'initialiser les poids des couches de convolution avec l'algorithme Xavier [Glorot2010] qui permet de déterminer automatiquement l'échelle d'initialisation en fonction du nombre de neurones d'entrée et de sortie. De manière générale, si les poids dans un réseau sont trop petits, alors la propagation en avant dans le réseau ne sera pas efficace car le signal diminuera jusqu'à être trop faible. De la même manière, si les poids sont trop élevés, le signal deviendra trop important pour être également utilisable. Concrètement, à partir d'une distribution gaussienne avec une moyenne nulle et une variance spécifique, l'algorithme Xavier s'assure d'obtenir des valeurs de poids correctes, en gardant le signal dans une plage de valeurs raisonnable à travers les différentes couches.

L'entraînement de ce réseau est par la suite utilisé de différentes manières en fonction des deux approches implémentées.

## 5.2/ LE SYSTÈME DEEP LEARNING COUPLÉ AU RÀPC POUR LA FUSION DE SEGMENTATIONS

L'hypothèse initiale que nous avons émise, dans le domaine d'application particulier du projet SAIAD, est que des segmentations similaires conduisent à des segmentations fusionnées similaires. Par conséquent, nous avons évalué la possibilité d'utiliser une technique de RàPC afin d'améliorer la fusion par un Deep Learning (DL), et d'introduire un schéma hiérarchique sur l'ensemble des cas à utiliser pendant la phase d'apprentissage du CNN. Concrètement, le RàPC est présent pour rechercher et fournir au DL les cas les plus similaires qu'il trouve dans sa base de cas. Les cas sélectionnés sont alors utilisés par le CNN pour son entraînement, afin d'avoir un réseau entraîné pour la fusion de segmentations avec des cas très proches du nouveau cas que l'on veut fusionner.

Comme le résume la Figure 5.2, le système DL couplé au système RàPC (DL-RàPC)

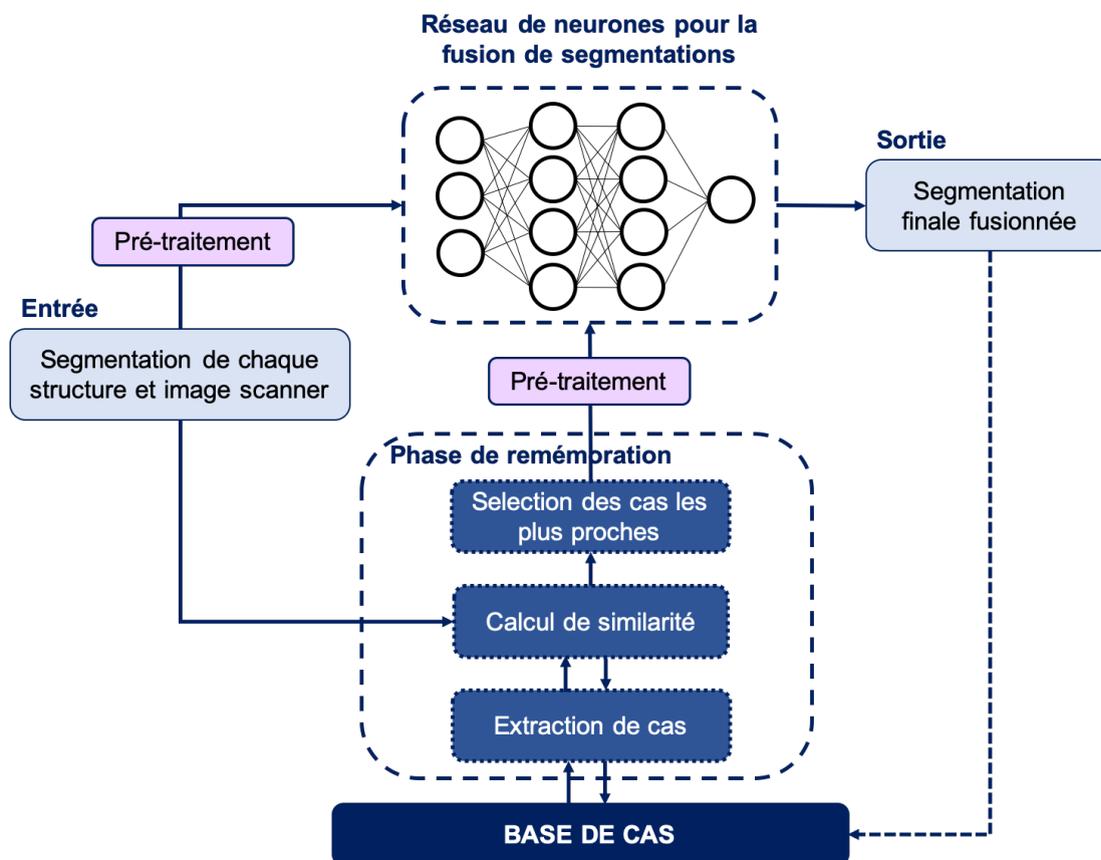


FIGURE 5.2 – Présentation générale du système DL-RàPC pour la fusion de segmentations.

comporte plusieurs parties. Le système RàPC, lui-même, est une méthode de résolution de problèmes basée sur l'analogie qui s'appuie sur ses expériences passées pour résoudre un nouveau problème. Notre système de fusion possède une base de cas permettant le stockage des cas connus, une phase de remémoration permettant le rappel de cas similaires, une phase de pré-traitement spécifique pour un cas avant d'être transmis à l'entrée du réseau de neurones et enfin le réseau de neurones prévu pour le processus de fusion. L'architecture du réseau de neurones utilisé pour la fusion a été détaillée ci-dessus, sur la Figure 5.1. Les détails de chacune des autres parties sont présentés ci-dessous. Le principe du système est le suivant :

1. Lorsqu'un nouveau cas est ajouté, le système calcule la valeur de similarité entre ce nouveau cas et chaque cas stocké dans la base.
2. Pendant la phase de remémoration, un certain nombre de cas source présentant les similarités les plus élevées sont extraits.
3. Puis, ces cas extraits sont préparés pour l'entrée du réseau de neurones et utilisés pour sa phase d'apprentissage.
4. Une fois le réseau formé, le nouveau cas, qui a été également pré-traité, est fourni à l'entrée du réseau pour le calcul du résultat final, qui est alors la segmentation finale dans laquelle la segmentation de la tumeur et celle du rein pathologique sont concaténées.

5. Enfin, ce nouveau cas, avec la solution, peut alors être stocké dans la base de cas si la segmentation qui en résulte est jugée consistante, afin d'enrichir la base.

L'entrée du système (le nouveau cas) est l'information d'une coupe transversale d'un patient. Plusieurs coupes transversales doivent être traitées afin d'obtenir toutes les informations sur le néphroblastome du patient. Ainsi, afin de réaliser la fusion des segmentations du rein et du néphroblastome dans leur intégralité, notre processus doit être répété sur chaque coupe segmentée du patient. En effet, les techniques de segmentation utilisées permettent de segmenter les structures coupe par coupe. Cela nous permet, en superposant ces segmentations, d'obtenir la segmentation 3D contenant toutes ces structures.

### 5.2.1/ LA MODÉLISATION D'UN CAS

De manière générale, **un cas** au sein d'un RàPC est composé d'une partie problème et d'une partie solution :

- la partie problème décrit les caractéristiques du problème à résoudre,
- et la seconde partie est la solution associée.

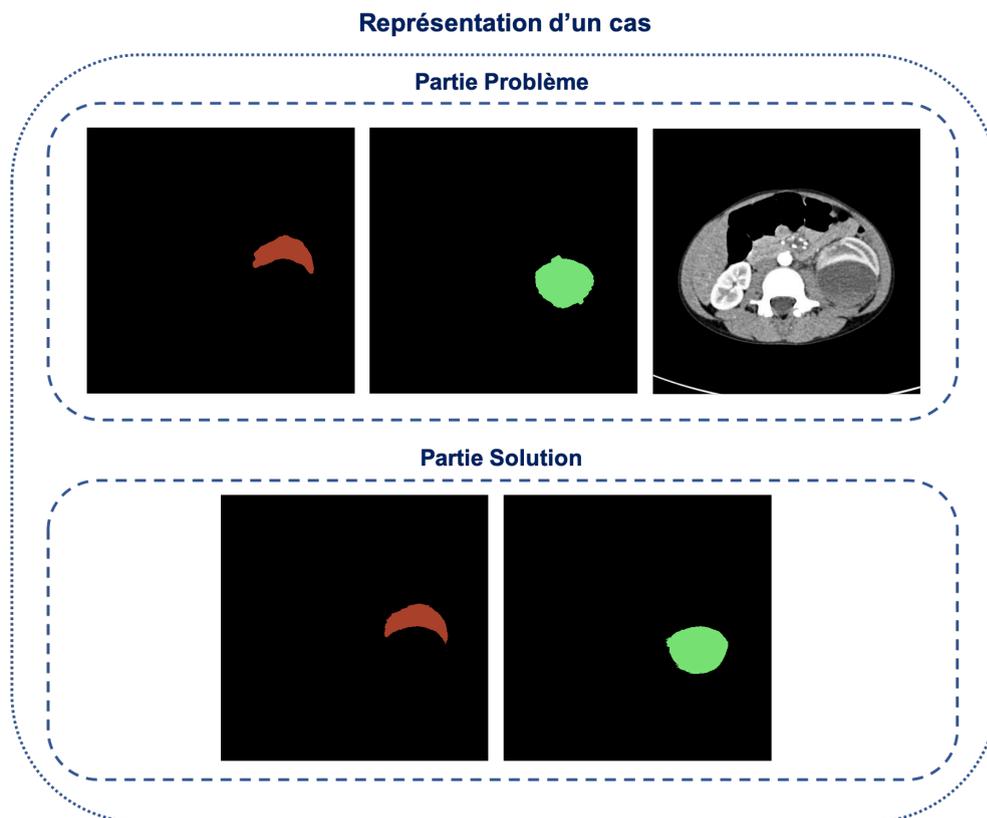


FIGURE 5.3 – Représentation des données définissant un cas de la base du RàPC.

Dans notre approche, un cas correspond aux informations présentes sur une coupe transversale d'un patient atteint de néphroblastome, ainsi que les vérités terrain. Comme le

montre la Figure 5.3, la partie problème d'un cas contient l'image scanner ainsi que les segmentations de chaque structure (rein pathologique en marron et tumeur en vert), pour lesquelles leur superposition peut conduire à l'apparition de pixels en conflit. La partie solution contient les vérités terrain des segmentations calculées. L'ensemble de ces données représentant un cas sera utilisé pour l'entraînement du réseau de neurones.

### 5.2.2/ LA PHASE DE REMÉMORATION

Afin de pouvoir extraire des cas similaires à la fusion à effectuer, un calcul de similarité doit être réalisé. La base de cas est composée de deux types de données : les images scanner et les segmentations. La formule de similarité utilisée dans notre phase de remémoration du RàPC est présentée dans l'Équation 52. Il s'agit de la moyenne entre le score Dice, qui calcule la similarité entre deux segmentations  $S$  et  $S'$  (décrites dans la Section 1.5.1.2), et la MSSIM (*Mean Structural SIMilarity*), présentée dans l'article de Wang et al. [Wang2004]), et calculant la similarité entre deux images scanner  $I$  et  $I'$  comme dans l'Équation 53. La métrique de Dice est couramment utilisée dans le domaine médical pour la comparaison entre différentes segmentations. La MSSIM quant à elle, est une amélioration de la SSIM (*Structural Similarity*) couramment utilisée dans la compression d'images. Comparée à la SSIM, la MSSIM utilise un fenêtrage itératif pour augmenter la capacité de comparaison structurelle sur les images, c'est-à-dire que chaque fenêtre décrit une région d'intérêt qui est comparée indépendamment.

$$s(S, S', I, I') = \frac{1}{2}(\text{Dice}(S, S') + \text{MSSIM}(I, I')) \quad (52)$$

#### Définitions

Le calcul de similarité MSSIM est la moyenne des SSIM calculées pour chaque fenêtre, et est définie par :

$$\text{MSSIM}(I, I') = \frac{1}{M} \sum_{i=1}^M \text{SSIM}(I_i, I'_i) \quad (53)$$

Où

- M est le nombre de fenêtres,
- SSIM est la mesure de similarité.

#### Définitions

SSIM est définie par :

$$\text{SSIM}(I, I') = \frac{(2\mu_I\mu_{I'} + C_1) + (2\sigma_{II'} + C_2)}{(\mu_I^2 + \mu_{I'}^2 + C_1)(\sigma_I^2 + \sigma_{I'}^2 + C_2)} \quad (54)$$

Où

- $\mu$  est la moyenne,
- $\sigma$  est l'écart-type,
- $C_1 = K_1 \times L$  et  $C_2 = K_2 \times L$ , avec  $L = 255$  correspondent à la plage de valeurs possibles des pixels.  $K_1 = 0,01$  et  $K_2 = 0,03$  sont les valeurs des variables issues de l'article de Wang et al. ([Wang2004]).

Comme les segmentations de différentes structures sont présentes dans un cas, il n'est pas possible d'utiliser directement l'Équation 52, car elle ne prend en entrée qu'une seule segmentation calculée qui sera comparée à sa vérité terrain. Les deux segmentations des différentes structures sont alors combinées en une seule, seulement pour le calcul de similarité, avec l'ajout d'une nouvelle étiquette pour les pixels en conflit. De même, leurs vérités terrain sont également combinées pour obtenir une seule vérité terrain qui inclut les deux structures. Le score Dice est alors calculé globalement sur toutes les structures.

De manière à limiter le temps de calcul nécessaire, le calcul de MSSIM se fait dans une fenêtre  $128 \times 128$ , ce qui fait un total de 16 fenêtres pour une image de taille  $512 \times 512$ . Les **trois cas** connus les plus similaires seront considérés, récupérés de la base de cas et utilisés pour la formation du réseau de neurones. Ce nombre a été défini expérimentalement et il est le résultat d'un compromis : il permet de réaliser un surapprentissage et d'expérimenter cette approche en un temps raisonnable.

### 5.2.3/ LE PRÉ-TRAITEMENT ET LA PHASE D'ENTRAÎNEMENT DU RÉSEAU

Pour que le réseau de neurones puisse traiter toutes les données présentes dans un cas, les images sont pré-traitées afin de pouvoir être données au réseau. Les deux segmentations calculées, ainsi que l'image scanner, présentes dans la partie solution sont alors concaténées en une image à trois canaux, dans laquelle chaque canal correspond à une image. Les vérités terrain de la partie solution serviront lors de l'entraînement du réseau à le guider pour le paramétrer correctement, afin d'obtenir la sortie désirée.

L'entraînement du réseau de neurones utilisé (présenté dans la Section 5.1) est également spécifique à notre système DL-RàPC. Pour chaque cas choisi lors de la phase de remémoration, l'entraînement est effectué avec un calcul automatique du nombre d'itérations pour l'apprentissage en fonction de la similarité calculée entre le nouveau cas et le cas choisi (cf. Équation 52). La taille de lot (*batch size*) représente le nombre de cas fournis à l'entrée du réseau lors de l'apprentissage pour une itération : dans notre cas la valeur est de 1. Les entraînements sont effectués dans l'ordre croissant des similarités obtenues. Ainsi, le cas choisi avec la similarité la plus faible sera utilisé pour le premier apprentissage, et le cas le plus similaire sera utilisé à la fin du processus d'apprentissage. Nous avons choisi d'utiliser le cas connu le plus similaire pour le dernier entraînement de notre réseau, afin de spécialiser et de concentrer notre entraînement principalement sur ce cas le plus similaire à notre nouveau cas, après avoir été aiguillé par les entraînements sur les autres cas choisis moins similaires.

Plus précisément, la phase de remémoration donne une liste de trois cas connus ( $C_1$ ,  $C_2$  et  $C_3$ ), Où :

- $C_1$  est le cas le plus similaire,
- $C_2$  le deuxième,
- et  $C_3$  le troisième cas le plus similaire récupéré.

Ces cas sont fournis au module DL de notre méthode avec les valeurs de similarité respectivement égales à  $S_1$ ,  $S_2$  et  $S_3$ . Ces valeurs sont calculées selon l'Équation 52. Puis, l'apprentissage du réseau de neurones est effectué en utilisant les images de  $C_3$  avec le nombre d'itérations fixé à  $S_3 \times NBMaxIt$ , où  $NBMaxIt$  correspond à un nombre constant d'itérations fixé à 3000. Ensuite, ce réseau de neurones formé est à nouveau entraîné en utilisant les images de  $C_2$ , avec  $S_2 \times NBMaxIt$  itérations. Enfin, ce réseau formé en deux temps est entraîné une troisième fois en utilisant les images de  $C_1$ , avec  $S_1 \times NBMaxIt$  itérations.

Cette méthode d'entraînement générale nous permet d'ajouter indirectement la notion de poids sur l'ensemble de l'apprentissage en accentuant l'apprentissage sur les cas les plus similaires et en pénalisant donc ceux l'étant moins. Plus un cas connu est similaire au nouveau cas à fusionner, plus le nombre d'itérations de l'entraînement du réseau sur ce cas connu sera élevé.

## 5.3/ LA MÉTHODE OV<sup>2</sup>ASSION POUR LA FUSION DE SEGMENTATIONS

### 5.3.1/ LE PRINCIPE DE LA NOUVELLE MÉTHODE D'ENTRAÎNEMENT

Pour former un réseau de neurones profond il est essentiel de disposer d'un nombre suffisamment important de données représentatives de toutes les données possibles. La segmentation manuelle d'un patient est un processus qui prend beaucoup de temps (6 à 8 heures pour la segmentation par un expert). De plus, le néphroblastome n'étant pas une tumeur très commune, à l'échelle d'un hôpital, il n'y a alors environ qu'une dizaine de segmentations abdominales de patients atteints de cette tumeur. Ce nombre peut ne pas être suffisant pour l'apprentissage classique, freiné encore plus par l'hétérogénéité des structures. Chaque tumeur et chaque rein pathologique est unique et varie fortement d'un patient à l'autre. De plus, les niveaux de gris des structures sur les images fournies par le scanner sont variables d'un patient à l'autre car elles dépendent entre autres du délai entre l'injection du produit de contraste et l'examen et de la vitesse à laquelle l'organisme du patient évacue ce produit de contraste. Beaucoup de paramètres sont donc patient-dépendants. Néanmoins, pour un patient donné, ces structures ont les mêmes nuances de gris et sont généralement homogènes d'une coupe à une autre.

Nous avons alors fait le choix, au sein du projet SAIAD, de nous orienter vers un apprentissage semi-automatique d'un CNN par patient, afin d'obtenir des résultats probants. Cette nouvelle méthode d'apprentissage, appelée OV<sup>2</sup>ASSION pour *Overlearning Vector for Valid Sparse SegmentatIONS*, est basée sur le sur-apprentissage de certaines coupes choisies d'un patient : ces coupes ayant été, par exemple, segmentées manuellement, et séparées par un écart (*gap*) afin de calculer automatiquement la segmentation de l'ensemble des coupes non segmentées de ce patient. Le système de *gap* est présenté par la Figure 5.4, et cet écart entre chaque coupe choisie est le même pour récupérer les informations de manière homogène à différents niveaux.

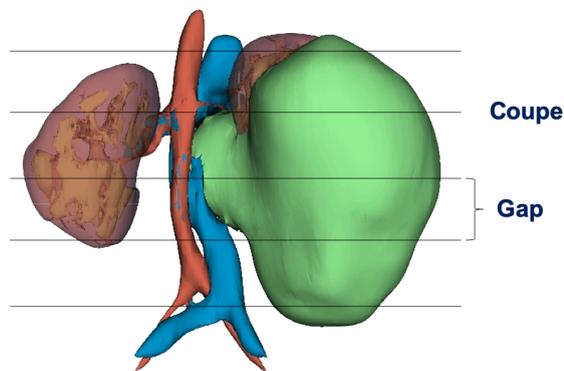


FIGURE 5.4 – La représentation 3D du néphroblastome : chaque ligne noire représente les coupes sélectionnées pour l'entraînement du CNN, séparées par un *gap* similaire entre chaque coupe choisie.

Dans le cas de la fusion de différentes segmentations issues du néphroblastome, les différentes structures n'ayant toujours qu'une faible différence d'une coupe voisine à une autre, cette méthode d'entraînement peut également être appliquée à un réseau effectuant la fusion. Comme les segmentations que l'on veut fusionner ont été réalisées avec cette méthode d'apprentissage, nous pouvons également tenir compte des segmentations manuelles réalisées pour la segmentation, et les utiliser dans le cadre de la fusion par apprentissage OV<sup>2</sup>ASSION, en gardant le même écart.

### 5.3.2/ FORMALISATION

La formalisation de l'ensemble d'apprentissage du réseau de neurones avec la méthode OV<sup>2</sup>ASSION a été présentée dans l'article de Marie et al. [Marie2019] pour la segmentation du rein pathologique par RàPC couplé à la croissance région, ainsi que la segmentation de la tumeur par DL avec cet entraînement spécifique. Dans cette méthode d'entraînement particulière, nous devons déterminer l'ensemble des coupes à prendre en considération pour l'entraînement. Cet ensemble est formalisé sous la forme d'un vecteur noté  $V_g$  qui détermine pour chaque coupe d'un patient, les coupes prises en compte pour l'entraînement et celles ne l'étant pas, le tout en fonction de l'écart (*gap*)  $g$ . Les valeurs possibles pour une coupe sont :

- 1 si elle est prise en compte,
- 0 sinon.

L'écart choisi est constant et doit être représentatif de l'ensemble des structures, ce qui permet alors d'exploiter les informations présentes de manière homogène et sur différents niveaux, afin de parvenir à calculer les résultats des autres coupes de manière fidèle. De plus pour un écart donné, il existe plusieurs possibilités d'ensemble de coupes choisies pour l'apprentissage, selon la première coupe prise en considération.

Soit un vecteur  $V_g$  avec un écart  $g$  et faisant partie de  $V$  (correspondant à l'ensemble de toutes les combinaisons possibles), tel que :

$$V_g \in V = \{(\{0, 1\}_1, \dots, \{0, 1\}_n)\} \quad (55)$$

Où :

- $n$  est le nombre de coupes du patient, variant généralement entre 30 et 150 selon l'épaisseur des coupes et la taille de la tumeur.

L'ensemble d'apprentissage  $LS$  (*Learning Set*) du réseau de neurones, que ce soit un réseau pour la segmentation ou pour la fusion, contient alors toutes les coupes utilisées pour l'entraînement qui ont été déterminées par un vecteur, et est défini comme :

$$LS = \{S_j, \dots, S_k\} \quad (56)$$

Où :

- $S_j, \dots, S_k$  représentent l'ensemble des coupes sélectionnées.

Cependant, il existe différents vecteurs possibles pour un écart donné, en fonction de la première coupe incluse dans  $LS$ .  $V_g$  correspond alors à l'ensemble des vecteurs possibles en fonction de  $g$ , tel que :

$$V_g = \bigcup_{i=1}^h (V_g)_i \quad (57)$$

Où :

- $h$  est le nombre de vecteurs possibles pour l'écart  $g$ ,
- et  $(V_g)_i$  le vecteur numéro  $i$  de l'écart  $g$ .

Ainsi, le vecteur  $(V_g)_1$  correspond à un vecteur commençant par inclure la première coupe à  $LS$  avant d'inclure toutes les  $g + 1$  coupes, le vecteur  $(V_g)_2$  commençant par inclure la seconde coupe à  $LS$ , et ainsi de suite.

Un exemple de ces différents vecteurs sur 60 coupes pour un écart de 2 est présenté dans le Tableau 5.1. Trois vecteurs sont alors possibles.

	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	...	$S_{58}$	$S_{59}$	$S_{60}$
$(V_2)_1$	1	0	0	1	0	...	1	0	0
$(V_2)_2$	0	1	0	0	1	...	0	1	0
$(V_2)_3$	0	0	1	0	0	...	0	0	1

TABLE 5.1 – Exemple des vecteurs possibles pour un *gap* de 2  $(V_2)_i$  et pour 60 coupes d'un patient ( $n = 60$ ).

La formalisation de l'ajout de coupes dans  $LS$  peut être réalisée par une fonction  $f$  qui permet de sélectionner une coupe ayant pour valeur 1 dans un vecteur  $(V_g)_i$  :

$$f : LS \times \{0, 1\} \rightarrow LS | \forall i \in \{1, \dots, n\} \begin{cases} f(S_i, 0) = \emptyset \\ f(S_i, 1) = S_i \end{cases} \quad (58)$$

Et enfin nous pouvons définir l'ensemble  $LS$  final, en introduisant une fonction  $F$  qui utilise la fonction  $f$  suivant le vecteur  $(V_g)_i$ , en parcourant tous les éléments du vecteur.

$$F : V \rightarrow LS | F(V_g)_i = \bigcup_{j=1}^n \{f(S_j, ((V_g)_i)_{j-1})\} \quad (59)$$

## 5.4/ LES RÉSULTATS

Toutes les données dont nous avons besoin pour cette contribution (images scanner, vérités terrain, segmentations calculées par IA) sont présentées dans la première contribution, Section 4.3.1 sur l'acquisition des données. Les segmentations calculées des dossiers anonymisés des 14 patients mis à notre disposition pour nos tests sont fusionnées par ces deux nouvelles méthodes de fusion par IA : la méthode DL-RàPC et la méthode DL-OV<sup>2</sup>ASSION. Pour ces deux méthodes, le réseau de neurones présenté dans la Section 5.1 est appliqué. D'un point de vue logiciel et matériel, la partie réseau de neurones de nos deux méthodes est développée en Python à l'aide de la librairie Caffe dédiée au DL [Jia2014], et les implémentations sont lancées dans un supercalculateur : le Mésocentre de l'Université de Franche-Comté, sur un GPU Tesla Kepler K40.

### 5.4.1/ LES PARAMÈTRES D'ENTRAÎNEMENT

Pour la fusion des segmentations d'un patient avec la méthode DL-RàPC, nous avons testé deux versions différentes de la base de cas du RèPC. La première version (**DL-RàPC v1**) contient l'ensemble des coupes de tous les autres patients, ainsi que quelques coupes du patient actuel qui ont été utilisées pour l'entraînement lors de la segmentation grâce au protocole OV<sup>2</sup>ASSION. La seconde version (**DL-RàPC v2**) contient dans sa base de cas uniquement les quelques coupes segmentées du patient actuel. Le réseau est ensuite entraîné sur un maximum de 3000 itérations par cas (le nombre d'itérations étant calculé par rapport à la similitude entre le cas récupéré de la base et le nouveau cas à fusionner) avec un total de 3 cas choisis pour son entraînement.

Dans le cas de la méthode de fusion **DL-OV<sup>2</sup>ASSION**, le réseau s'entraîne selon la méthode particulière pour chaque patient, sur toutes les coupes sélectionnées du patient avec un *gap* de 4, en commençant par la première coupe. En effet, un écart de 4 s'est avéré être le paramètre donnant le meilleur compromis entre la précision et le temps consommé pour la segmentation selon les travaux menés par Marie et al. [Marie2019]. Ces coupes avec un *gap* de 4 pour chaque patient (coupes utilisées précédemment pour l'entraînement dans l'étape de segmentation) ont également été segmentées, afin de servir de base d'entraînement à notre méthode de fusion. Le réseau est ensuite entraîné sur 30000 itérations, et pour la phase de prédiction, toutes les autres coupes du patient considéré qui n'ont pas été utilisées pour l'entraînement sont données au réseau.

Pour les deux méthodes, le taux d'apprentissage utilisé est de  $1e^{-9}$  et le momentum de 0,99.

### 5.4.2/ L'ÉVALUATION DES DIFFÉRENTES MÉTHODES

#### 5.4.2.1/ COMPARAISON AVEC DES RÉSEAUX DE NEURONES

Les résultats de nos deux méthodes de fusion par Intelligence Artificielle sont dans un premier temps comparées avec les résultats d'autres réseaux de neurones dans la littérature. Plus particulièrement, nous les comparons, d'une part, avec le réseau pour la fusion proposé par Tang et al. [Tang2016b], qui est composé de trois couches de convolution (cf. Section 3.2.2) et d'autre part, avec le réseau proposé par Xiao et al. [Xiao2018] composé de trois couches de convolution chacune suivie d'une couche ReLU (cf. Section 3.2.3). Enfin, le réseau proposé par Qu et al. [Qu2017], plus profond (cf. Section 3.2.4), est également utilisé pour la comparaison mais il est alors adapté à notre situation afin de pouvoir le tester. En effet, ce réseau utilise des couches FC (*Fully Connected*) en fin de réseau suivies d'une propagation laplacienne afin d'obtenir le résultat. Dans le but de ne tester que le CNN proposé, les deux couches FC ont été remplacées par des couches de convolution avec un noyau de taille 3 et 1, et la valeur du remplissage (*padding*) de chaque convolution dans le réseau est ajustée afin que l'image en sortie ne soit pas de taille réduite par rapport à l'image donnée en entrée du réseau. L'entraînement réalisé pour ces trois réseaux de fusion est un entraînement classique (entraînement sur toutes les coupes de tous les patients sauf un, puis validation sur ce patient non entraîné), sur 20000 itérations avec un taux d'apprentissage de  $1e^{-10}$  et un momentum de 0,99. Comme nous utilisons notre propre architecture de réseau (cf. Figure 5.1), et que nous l'utilisons dans nos méthodes avec des entraînements spécifiques, nous l'avons également utilisée dans un entraînement classique afin de comparer ces résultats. Les mêmes paramètres que pour l'entraînement des autres réseaux de la littérature sont également appliqués.

La Table 5.2 présente le pourcentage de pixels en conflit correctement résolus, sur l'ensemble de nos 14 patients, pour nos deux nouvelles méthodes de fusion par IA : DL-OV<sup>2</sup>ASSION et DL-RàPC (DL-RàPC v1 contenant dans sa base quelques coupes du patient à fusionner ainsi que l'ensemble des coupes des autres patients et DL-RàPC v2 contenant seulement les coupes utilisées pour la segmentation du patient à fusionner), le tout comparé aux résultats d'autres réseaux. Les valeurs en gras correspondent au meilleur résultat pour chaque patient et la moyenne des résultats pour l'ensemble des patients ainsi que l'écart type noté  $\sigma$  sont également exposés. Nous pouvons observer que la méthode DL-OV<sup>2</sup>ASSION est la plus efficace pour la gestion des pixels en conflit parmi les autres méthodes d'IA. Elle fait néanmoins partie des méthodes ayant un écart-type élevé (14,23%), notamment à cause du patient P11 pour lequel le réseau montre des difficultés à résoudre les conflits par rapport aux autres méthodes. On peut supposer qu'il s'agit d'un manque d'entraînement pour les données de ce patient P11. En effet, les paramètres d'entraînement ont été fixés à 3000 itérations pour tous les patients. Les résultats des tests avec la méthode DL-RàPC sont dans la moyenne de tous les résultats obtenus. Cependant, la méthode DL-RàPC v2 est plus performante que la version v1, alors que sa base de cas est moins fournie. Enfin, la gestion des conflits avec les méthodes DL-OV<sup>2</sup>ASSION, DL-RàPC v1 et DL-RàPC v2 est plus performante qu'en utilisant le réseau implémenté dans ces méthodes avec un entraînement classique.

La Table 5.3 présente la qualité des segmentations obtenues en score Dice pour chaque méthode et réseau. Ce score de Dice est le score moyen obtenu en calculant séparément le score de chaque structure et en réalisant la moyenne. Le score de Dice avant la fusion est également exposé, ainsi que la moyenne et l'écart type pour l'ensemble des

Patient	Méthode DL-OVÀSSION	Méthode DL-RàPC v1	Méthode DL-RàPC v2	Réseau DL (cf. Image 5.1)	Réseau par Tang et al. [Tang2016b]	Réseau par Xiao et al. [Xiao2018]	Réseau par Qu et al. [Qu2017]
P1	<b>0.7002</b> (7861/11226)	0.6895 (7740/11226)	0.6913 (7760/11226)	0.6056 (6799/11226)	0.6041 (6782/11226)	0.6162 (6918/11226)	0.6400 (7185/11226)
P2	<b>0.7655</b> (1942/2537)	0.7469 (1895/2537)	0.7418 (1882/2537)	0.6173 (1566/2537)	0.6795 (1724/2537)	0.6831 (1733/2537)	0.6926 (1757/2537)
P3	<b>0.9336</b> (225/241)	0.8257 (199/241)	0.8589 (207/241)	0.3817 (92/241)	0.7054 (170/241)	0.0249 (6/241)	0.7012 (169/241)
P4	<b>0.8075</b> (3117/3860)	0.7461 (2880/3860)	0.7311 (2822/3860)	0.6609 (2551/3860)	0.7189 (2775/3860)	0.7915 (3055/3860)	0.2650 (1023/3860)
P5	<b>0.6527</b> (3264/5001)	0.6185 (3093/5001)	0.6139 (3070/5001)	0.5367 (2684/5001)	0.5655 (2828/5001)	0.6003 (3002/5001)	0.5655 (2928/5001)
P6	0.5834 (3128/5362)	0.4385 (2351/5362)	0.4717 (2529/5362)	<b>0.6591</b> (3534/5362)	0.4717 (3118/5362)	0.6087 (3264/5362)	0.5692 (3052/5362)
P7	0.6891 (1924/2792)	0.4789 (1337/2792)	0.5605 (1565/2792)	<b>0.7339</b> (2049/2792)	0.6819 (1904/2792)	0.6143 (1715/2792)	0.6658 (1859/2792)
P8	<b>0.7840</b> (1212/1546)	0.7180 (1110/1546)	0.7516 (1162/1546)	0.3862 (597/1546)	0.3758 (581/1546)	0.3461 (535/1546)	0.3396 (525/1546)
P9	0.6242 (1945/3116)	0.5270 (1642/3116)	0.6104 (1902/3116)	0.6120 (1907/3116)	0.6399 (1994/3116)	<b>0.6553</b> (2042/3116)	0.6287 (1959/3116)
P10	0.6490 (1466/2259)	0.6184 (1397/2259)	0.6494 (1467/2259)	0.5919 (1337/2259)	0.6175 (1395/2259)	0.5321 (1202/2259)	<b>0.6658</b> (1504/2259)
P11	0.3152 (255/809)	0.6007 (486/809)	<b>0.6539</b> (529/809)	0.4969 (402/809)	0.4574 (370/809)	0.6378 (516/809)	0.4586 (371/809)
P12	0.5708 (1197/2097)	0.5551 (1164/2097)	0.3948 (828/2097)	<b>0.6924</b> (1452/2097)	0.5656 (1186/2097)	0.5622 (1179/2097)	0.4683 (982/2097)
P13	0.5965 (8474/14207)	0.4905 (6969/14207)	0.5486 (7794/14207)	0.4793 (6810/14207)	0.6222 (8840/14207)	0.4777 (6786/14207)	<b>0.6255</b> (8887/14207)
P14	<b>0.6539</b> (5425/8297)	0.6001 (4979/8297)	0.6064 (5031/8297)	0.5948 (4935/8297)	0.5788 (4802/8297)	0.6482 (5378/8297)	0.5226 (4336/8297)
<b>Moy.</b>	<b>0.6541</b>	0.5879	0.6085	0.5796	0.6072	0.5893	0.5767
$\sigma$	0.1423	0.1150	0.1200	0.1067	<b>0.0938</b>	0.1842	0.1335

TABLE 5.2 – Pourcentage des pixels en conflit correctement résolu par les méthodes DL-OVÀSSION et DL-RàPC, en comparaison avec d'autres réseaux utilisés dans la littérature ainsi que la méthode par agrégation de multiples méthodes classiques (deuxième modèle, présenté dans la Figure 4.4, avec fusion par vote majoritaire).

patients. Les valeurs en rouge sont les résultats en dessous des résultats des segmentations avant la fusion, c'est-à-dire que les qualités des segmentations ont baissé. Cette fois-ci, la méthode obtenant les meilleures segmentations après la fusion est l'utilisation de notre réseau avec un entraînement classique (score de Dice de 88,18%), et augmente en moyenne le score de Dice de plus de 2%. Ce score de Dice se rapproche même du score de Dice moyen inter-opérateurs de 90,64%, à 2% près (cf. Table 4.1). Les segmentations obtenues après cette méthode peuvent être potentiellement médicalement exploitables (sous couvert de validation par les personnels compétents en l'absence actuelle de données suffisantes pour une validation clinique).

La méthode obtenant le deuxième meilleur score de Dice de 88,06% est notre méthode DL-OV<sup>2</sup>ASSION, score proche du meilleur score. Le score de Dice obtenu de la méthode DL-RàPC v2 est dans la moyenne, avec néanmoins des segmentations de moins bonnes qualités pour certains patients. Cependant le score de Dice de la méthode DL-RàPC v1 est le plus faible, à 85,98%. Pour en savoir plus sur la difficulté de la méthode DL-RàPC à fournir de meilleures segmentations, nous nous sommes intéressés aux résultats par structure individuelle.

La Table 5.4 montre également les scores de Dice obtenus pour chaque méthode et réseau, mais en fonction de la structure de la tumeur et du rein pathologique. On peut observer que bien que les méthodes DL-RàPC v1 et v2 réussissent à retranscrire la segmentation de la tumeur, elles ont finalement du mal à transcrire la segmentation du rein pathologique et obtiennent un score de Dice inférieur par rapport au score de Dice initial, ainsi qu'un écart type plus important. Nous observons également que les résultats moyens de la première version sont inférieurs aux résultats de la deuxième version (de même pour les résultats des pixels en conflit correctement résolus), bien que sa base de cas soit plus importante. Pour ce qui est de la méthode DL-OV<sup>2</sup>ASSION, elle obtient le score de Dice le plus élevé pour la structure de la tumeur, avec 92,66%. Pour la segmentation du rein pathologique, elle obtient le second meilleur score avec un score Dice de 83,45% (Il s'agit du réseau DL qui obtient les meilleures segmentations et le meilleur écart type). Pour rappel, le score de Dice inter-opérateur de la tumeur et du rein pathologique est de 94,75% et 86,53%. Nous ne sommes donc plus qu'à 2% et 3% d'atteindre les valeurs inter-opérateur.

#### 5.4.2.2/ COMPARAISON AVEC DES MÉTHODES N'UTILISANT PAS L'IA

Les performances de nos méthodes d'IA peuvent également être comparées à celle des méthodes plus classiques, dont notre méthode présentée dans le Chapitre 4. La Table 5.5 présente les résultats, en pourcentage des pixels en conflit qui ont été correctement classifiés ainsi que les scores de Dice moyens obtenus, avec l'écart type noté  $\sigma$ . Nos trois méthodes sont comparées avec le modèle de fusion v2 (cf. Figure 4.4), avec l'utilisation de la métrique *VoI* dans le cadre de la fusion (cf. Section 2.3.2) ainsi qu'avec l'utilisation de la méthode NandI (cf. Section 2.4.1). Pour plus de clarté, seuls les résultats globaux sur l'ensemble des 14 patients sont montrés. Nous utilisons le modèle v2 Fusion VM (Vote Majoritaire) comme comparaison car il s'agit d'un modèle de fusion efficace permettant de ne pas avoir de problèmes d'égalité comme c'est le cas pour les modèles v1, et qui se trouve être plus performant avec une fusion des résultats des méthodes par vote majoritaire que par la fusion basée sur les CRFs.

Nos méthodes de fusion par IA présentent des résultats de gestion de conflits cohérents,

Patient	Dice avant Fusion	Méthode DL-OVÀSSION	Méthode DL-RàPC v1	Méthode DL-RàPC v2	Réseau DL (cf. Image 5.1)	Réseau par Tang et al. [Tang2016b]	Réseau par Xiao et al. [Xiao2018]	Réseau par Qu et al. [Qu2017]
P1	0.8961	<b>0.9189</b>	0.9157	0.9144	0.9100	0.9013	0.9007	0.9100
P2	0.8904	<b>0.9057</b>	0.8964	0.8977	0.9056	0.8913	<b>0.8813</b>	0.8951
P3	0.8706	<b>0.8939</b>	0.8803	0.8743	<b>0.8607</b>	0.8714	<b>0.8676</b>	0.8841
P4	0.9357	<b>0.9496</b>	0.9461	0.9460	0.9430	0.9377	<b>0.9343</b>	0.9408
P5	0.8862	<b>0.9105</b>	0.9010	0.9030	0.9042	0.8955	<b>0.8687</b>	0.8969
P6	0.8748	0.8796	<b>0.8668</b>	<b>0.8684</b>	<b>0.9017</b>	<b>0.8709</b>	0.8754	0.8860
P7	0.8377	0.8560	0.8470	0.8465	<b>0.8674</b>	0.8399	0.8416	0.8493
P8	0.9145	<b>0.9339</b>	0.9177	0.9176	0.9237	0.9163	0.9182	0.9231
P9	0.8069	<b>0.8599</b>	<b>0.7642</b>	<b>0.7913</b>	0.8488	0.8228	0.8279	0.8338
P10	0.7731	<b>0.7720</b>	<b>0.7471</b>	<b>0.7458</b>	0.7799	0.7850	0.7868	<b>0.7965</b>
P11	0.8471	0.8617	0.8500	<b>0.8444</b>	<b>0.8923</b>	<b>0.8291</b>	<b>0.8356</b>	0.8590
P12	0.7881	0.8354	0.7962	<b>0.7878</b>	<b>0.8361</b>	0.7971	0.8071	0.8096
P13	0.8438	<b>0.8433</b>	<b>0.8049</b>	<b>0.8242</b>	<b>0.8694</b>	<b>0.8437</b>	0.8508	0.8555
P14	0.8835	<b>0.9074</b>	0.9035	0.8997	0.9030	0.8874	<b>0.8775</b>	0.8957
Moy.	0.8606	0.8806	<b>0.8598</b>	0.8615	<b>0.8818</b>	0.8635	0.8624	0.8737
$\sigma$	0.0472	0.0465	<b>0.0611</b>	<b>0.0578</b>	0.0418	0.0452	<b>0.0410</b>	0.0416

TABLE 5.3 – Résultat du score de Dice moyen (moyenne du score de Dice entre les deux structures) avant et après la fusion par les méthodes DL-OVÀSSION et DL-RàPC, et en comparaison avec d'autres réseaux utilisés dans la littérature.

	Dice avant Fusion		Méthode DL-OVÀSSION		Méthode DL-RàPC v1		Méthode DL-RàPC v2		Réseau DL (cf. Image 5.1)		Réseau par Tang et al. [Tang2016b]		Réseau par Xiao et al. [Xiao2018]		Réseau par Qu et al. [Qu2017]	
Dice Tumeur	0.9154		<b>0.9266</b>	0.9175	0.9186	0.9199	0.9163						<b>0.9129</b>			0.9218
$\sigma$ Tumeur	0.0449		<b>0.0373</b>	0.0397	0.0416	<b>0.0496</b>	0.0449						0.0466			0.0384
Dice Rein P.	0.8058		0.8345	<b>0.8021</b>	<b>0.8044</b>	<b>0.8438</b>	0.8108						0.8118			0.8256
$\sigma$ Rein P.	0.1127		0.1048	<b>0.1323</b>	<b>0.1287</b>	<b>0.0963</b>	0.1101						0.1009			0.0987

TABLE 5.4 – Résultats du score Dice moyen obtenu pour la structure de la tumeur et du rein pathologique par les méthodes DL-OVÀSSION et DL-RàPC, et en comparaison avec d'autres réseaux utilisés dans la littérature.

similaires voire meilleurs en comparaison avec d'autres méthodes non-IA dans la littérature. Finalement, notre méthode DL-OV<sup>2</sup>ASSION est supérieure à la méthode basée sur la métrique *Vol* et à la méthode Nandl. Le modèle de fusion v2 avec vote majoritaire est la méthode obtenant les meilleures gestions des conflits, avec 71,95% de bonnes résolutions. Cependant, pour ce qui est de la qualité des segmentations finalement obtenues, DL-OV<sup>2</sup>ASSION permet une amélioration significative et supérieure aux méthodes n'utilisant pas l'IA. Ce résultat pouvant paraître paradoxal s'explique par le fait que les méthodes qui n'utilisent pas l'IA ne peuvent modifier que les étiquettes des pixels en conflit, contrairement aux méthodes par réseau de neurones qui modifient les étiquettes de l'ensemble des pixels des segmentations à fusionner.

		Méthode DL-OVASSION	Méthode DL-RàPC v1	Méthode DL-RàPC v2	Modèle V2 Fusion VM	Métrique Vol	Méthode Nandl
% de pixels en conflit	Moy.	0.6541	0.5879	0.6085	<b>0.7195</b>	0.6523	0.6538
	$\sigma$	0.1423	0.1150	0.1200	0.0900	0.1132	<b>0.0698</b>
Dice	Moy.	<b>0.8806</b>	0.8598	0.8615	0.8611	0.8596	0.8605
	$\sigma$	<b>0.0465</b>	0.0611	0.0578	0.0481	0.0494	0.0481

TABLE 5.5 – Comparaison des résultats obtenus (Pourcentage des pixels en conflit correctement résolus et Dice) entre les méthodes par IA et les méthodes non-IA.

### 5.4.3/ LA REPRÉSENTATION VISUELLE DES RÉSULTATS

Les scores de Dice obtenus par nos différentes méthodes nous permettent finalement, après fusion, d'obtenir des segmentations relativement proches des vérités terrain. La Figure 5.5 montre les segmentations obtenues par nos deux méthodes par IA. Les colonnes de gauche à droite permettent de visualiser : les images scanner des coupes des patients, les vérités terrain réalisées par les experts qui sont la référence (en rouge le rein pathologique et en vert la tumeur), le résultat des segmentations calculées avec les conflits à résoudre en bleu clair, le résultat de la fusion de segmentations avec la méthode DL-OV<sup>2</sup>ASSION, et enfin le résultat de la fusion de segmentations avec la méthode DL-RàPC v1. Pour chaque résultat, le score de Dice de la tumeur (marqué *T*) et du rein pathologique (marqué *K*) est présent ainsi que le pourcentage de bonne résolution des conflits.

Les résultats obtenus par la deuxième version de la méthode DL-RàPC ne sont pas exposés car les segmentations sont très similaires aux résultats de la première version, et il serait difficile de trouver les différences visuellement. De manière générale, le processus de fusion a permis de lisser les contours, ce qui améliore donc le score de Dice obtenu avant fusion. La segmentation de la tumeur entre les deux méthodes reste visuellement et globalement similaire, sauf pour le patient P13 pour lequel la tumeur possède une forme plus arrondie avec la méthode DL-RàPC v1. Au niveau de la segmentation du rein pathologique, des différences plus importantes sont visuellement perceptibles entre les deux méthodes, comme c'est le cas pour les coupes présentées des patients P8, P9 et P13.

La Figure 5.6 nous montre les résultats de trois coupes de patients différents, pour différentes méthodes utilisées (par IA ou non IA). Pour chaque segmentation, un zoom est affiché en haut à droite pour permettre de mieux visualiser la zone de conflits. Cette zone de conflits est entourée en bleu. Nous avons également, comme pour la figure

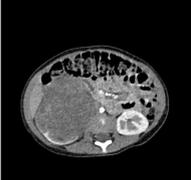
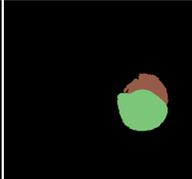
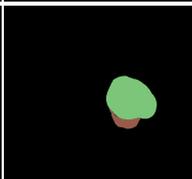
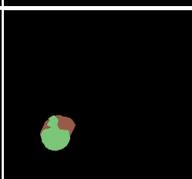
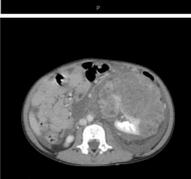
Patient	Images scanner	Vérités terrain	Segmentations avec conflits	Résultats par la méthode DL-OV <sup>2</sup> ASSION	Résultats par la méthode DL-RàPC v1
P1 Coupe 72			T 0.9715 K 0.8225	T 0.9781 K 0.8968 80.92%	T 0.9767 K 0.8992 72.54%
P4 Coupe 84			T 0.9594 K 0.9236	T 0.9737 K 0.9447 90.24%	T 0.9631 K 0.9361 46.83%
P5 Coupe 75			T 0.9542 K 0.9337	T 0.9671 K 0.9531 86.29%	T 0.9659 K 0.9427 88.79%
P6 Coupe 34			T 0.8974 K 0.8782	T 0.9082 K 0.8363 81.68%	T 0.9026 K 0.9229 85.88%
P8 Coupe 71			T 0.8539 K 0.8275	T 0.9655 K 0.9196 98.96%	T 0.8947 K 0.7938 83.16%
P9 Coupe 12			T 0.9636 K 0.5285	T 0.9710 K 0.6606 62.27%	T 0.9639 K 0.5072 28.07%
P13 Coupe 80			T 0.9316 K 0.8006	T 0.9248 K 0.8983 86.85%	T 0.9380 K 0.8941 70.74%
P14 Coupe 83			T 0.8265 K 0.9439	T 0.8822 K 0.9645 70.13%	T 0.8809 K 0.9586 62.58%

FIGURE 5.5 – Résultats de la fusion de segmentations par la méthode DL-OV<sup>2</sup>ASSION et DL-RàPC v1 sur quelques coupes de différents patients.

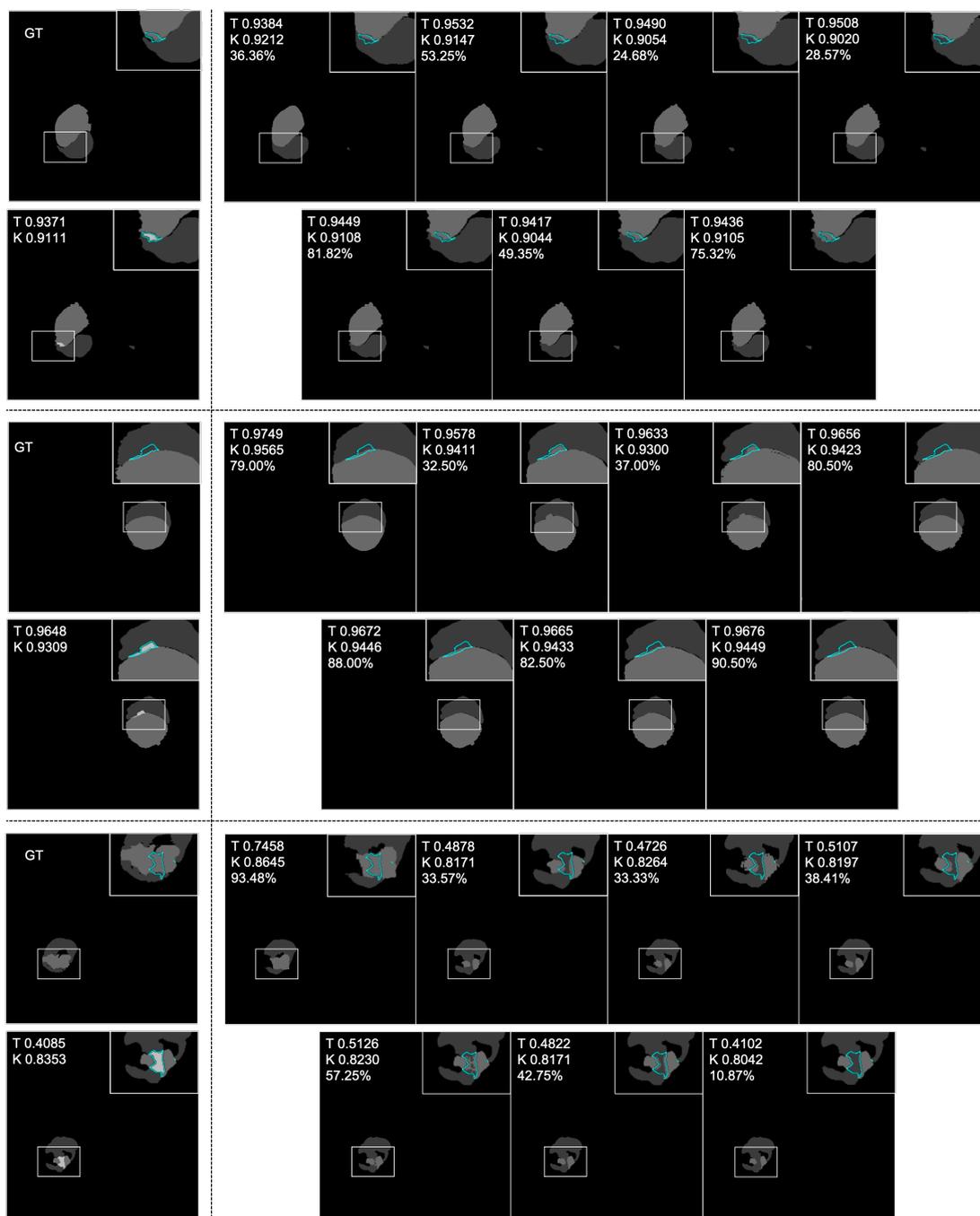


FIGURE 5.6 – Comparaison des différents résultats en fonction des méthodes utilisées, sur trois coupes de trois patients différents. Sur la colonne de gauche nous avons les vérités terrain ainsi que la segmentation avec conflits, et sur la colonne de droite nous avons de gauche à droite : les résultats des méthodes DL-OV<sup>2</sup>ASSION, de Qu et al. [Qu2017], de Tang et al. [Tang2016b] et de Xiao et al. [Xiao2018] sur la ligne du haut, ainsi que les résultats du modèle V2 avec fusion VM, de Nandl ainsi que de la métrique Vol sur la ligne du bas.

précédente, la visualisation du score de Dice de la tumeur et du rein pathologique ainsi que le pourcentage de bonne résolution des conflits.

Dans le cas de nos trois exemples, la méthode DL-OVASSION obtient en moyenne une meilleure amélioration de la qualité des segmentations. En revanche, dans le cas de la gestion des pixels conflictuels, les résultats sont plus hétérogènes. Dans le premier exemple, les méthodes non basées sur l'IA sont les plus performantes en matière de fusion. C'est également le cas dans le second exemple, mais certaines méthodes basées sur l'IA y parviennent également. Dans le troisième exemple par contre, aucune des méthodes n'y parvient, à l'exception de la méthode DL-OVASSION avec un bon pourcentage de résolution de 93,48%. En moyenne, par rapport à toutes les coupes de tous nos patients, nous avons déjà montré que notre résolution des conflits était bonne (cf. Table 5.2), mais il s'agit du deuxième modèle de fusion par vote majoritaire (cf. Figure 4.4) qui obtient les meilleurs résultats (cf. Table 5.5).

## 5.5/ DISCUSSIONS

Cette section analyse les performances obtenues par les réseaux de neurones dans le cadre de la fusion de segmentations. Elle analyse et compare également leurs performances à celles des méthodes plus classiques de notre précédente contribution présentée dans le Chapitre 4. La discussion sur les différentes difficultés induites par la gestion de conflits sur les pixels, présentée dans la Section 4.4.3 de la précédente contribution, est également prise en considération ici pour cette seconde contribution.

### 5.5.1/ LES PERFORMANCES DES DIFFÉRENTS RÉSEAUX DE NEURONES

Dans le cadre d'un entraînement classique pour les réseaux de neurones, d'après les résultats obtenus, un réseau possédant moins de couches obtient des meilleurs résultats de gestion de conflits qu'un réseau plus profond, comme le montrent les résultats des réseaux de Tang et al. et Xiao et al. avec seulement trois couches de convolution, comparés aux résultats du réseau de Qu et al. et de notre réseau DL. Cependant, en considérant les segmentations finalement obtenues, un réseau de neurones profond améliore de manière plus significative la qualité des segmentations qu'un réseau plus petit. Ceci est dû au fait qu'un réseau plus profond permet de mieux retranscrire la segmentation finale dans sa globalité, comme c'est le cas pour les réseaux prévus spécifiquement pour la segmentation.

### 5.5.2/ LA DIFFÉRENCE DE PERFORMANCES ENTRE LES DEUX STRUCTURES ANATOMIQUES

Il est normal que la fusion des segmentations ait un impact plus important sur la segmentation du rein pathologique que la segmentation de la tumeur, avec une variation plus élevée du score de Dice avant et après fusion. En effet, de manière générale, le rein pathologique est une structure plus petite et moins imposante que la tumeur, avec des formes plus aléatoires car il est poussé et déformé par la tumeur. De ce fait, même si le nombre de pixels en conflit se trouvant dans l'image est faible par rapport au nombre total de pixels, la résolution de ces pixels étiquetés rein ou tumeur aura un impact plus significatif sur la structure plus petite du rein (contenant moins de pixels) que sur celle de la tumeur (contenant plus de pixels).

Il en est de même lors de la modification légère de la forme de la structure, car les réseaux de neurones peuvent éventuellement modifier l'étiquette de tous les pixels de l'image lors de la fusion. Changer l'étiquette d'une dizaine de pixels sur une structure comportant des milliers de pixels aura un faible impact sur son score de Dice alors que sur une structure ne comportant qu'une petite centaine de pixels, la variabilité potentielle de la structure sera plus importante. Ce sont les raisons pour lesquelles l'utilisation de la méthode DL-OV<sup>2</sup>ASSION gérant bien la structure de la tumeur permet une amélioration de 1,12% de son score de Dice pour cette structure, et l'utilisation de notre réseau avec entraînement classique, gérant le mieux la structure du rein pathologique, permet une amélioration de 3,8% de son score de Dice. Et ce sont les raisons pour lesquelles, au contraire, l'utilisation de la méthode DL-RàPC, qui présente des difficultés à retranscrire la segmentation du rein, obtient un score de Dice légèrement inférieur au score de Dice de cette structure avant la fusion.

### 5.5.3/ DISCUSSION SUR LES PERFORMANCES DE LA MÉTHODE DL-RÀPC

Finalement, les deux méthodes DL-OV<sup>2</sup>ASSION et DL-RàPC proposent un sur-apprentissage des coupes pour la fusion très similaires : celles du patient quand cela est possible (DL-OV<sup>2</sup>ASSION et DL-RàPC v2), celles d'autres patients lorsqu'elles ne suffisent pas (DL-RàPC v1). Il semble donc que les cas remémorés par DL-RàPC v1 ne permettent pas toujours un sur-entraînement optimum. Il est toutefois logique que DL-RàPC v1 et v2 permette d'obtenir de moins bons résultats que DL-OV<sup>2</sup>ASSION si nous tenons compte du nombre de fusions apprises provenant du patient considéré. En effet, finalement, un CNN entraîné selon OV<sup>2</sup>ASSION et un écart  $g = 4$  compte plus d'exemples proches du patient considéré (entre 8 et 15) qu'un CNN entraîné avec seulement les 3 fusions les plus similaires (DL-RàPC). Nous pouvons donc conclure de cette étude comparative que c'est la méthode basée sur un CNN dédié à la fusion et sur-entraîné selon OV<sup>2</sup>ASSION qui a permis de fournir les meilleures segmentations finales. Nous pouvons également déduire de ces observations qu'un certain nombre-seuil d'exemples de fusions similaires doivent être apprises préalablement par le réseau avant de pouvoir fournir des segmentations finales plus proches des vérités terrain. Pour les dossiers des patients que nous avons utilisés pour nos tests, ce nombre-seuil se situe donc entre 3 et 15.

La méthode de fusion DL couplée au RàPC a été initialement testée sur un nombre de patients plus limité (en effet seules les données de 9 patients étaient disponibles au début pour l'évaluation de cette méthode), et les premiers résultats préliminaires étaient convaincants. Cependant, en enrichissant la base de cas avec tous nos patients, cette méthode a finalement obtenu un pourcentage de gestion des conflits dans la moyenne des autres réseaux comparés, mais nous avons également remarqué qu'elle a obtenu un score de Dice moyen inférieur au score de Dice des segmentations avant la fusion. En particulier, elle présente des difficultés à fournir des segmentations correctes sur le rein pathologique, et fait baisser le Dice pour cette structure de 0,37%.

C'est pourquoi, nous avons comparé cette méthode à sa deuxième version, ne comprenant dans sa base que les segmentations du patient considéré utilisées précédemment dans le processus de segmentation avec la méthode OV<sup>2</sup>ASSION. Ces nouvelles données dans la base de la deuxième version sont les données les plus proches du cas que l'on veut résoudre, puisqu'ils proviennent du même patient. Cette version 2 est d'ailleurs plus

performante, pour la gestion des pixels en conflit comme pour l'amélioration des segmentations, même si elle présente toujours de légères difficultés à optimiser la segmentation du rein pathologique.

Au regard des cas remémorés par DL-RàPC v1 et v2, il serait donc utile d'améliorer le calcul de similarité afin que les cas remémorés soient en priorité ceux du patients considérés quand ils existent dans la base de cas. À ce titre, il serait intéressant de pondérer dans le calcul de similarité les informations venant des segmentations (avec le calcul de score de Dice) et les informations venant des images scanner (avec le calcul de MSSIM). Le nombre de fenêtres utilisées dans le calcul de MSSIM, actuellement de 16, pourrait également être modifié, mais un nombre de fenêtres plus important engendre une grande augmentation du temps de calcul. Enfin, il est également possible de modifier le nombre de cas choisis pour l'entraînement, ce qui aura également un impact sur le temps de calcul. À terme, avec une amélioration de la sélection des cas les plus similaires, une augmentation du nombre de cas pris en compte ainsi qu'une base de cas de plus en plus étoffée, cette méthode pourrait concurrencer la méthode de fusion DL-OV<sup>2</sup> ASSION.

#### 5.5.4/ LE TEMPS DE CALCUL

Actuellement, le processus de segmentation des patients par la méthode OV<sup>2</sup>ASSION est réalisé en 8 heures pour chaque patient, et toutes les méthodes de fusion présentées sont également réalisées sur plusieurs heures, avec 4 heures et 7 heures environ pour les méthodes de fusion DL-OV<sup>2</sup>ASSION et DL-RàPC. Les temps d'exécution finaux (segmentation et fusion) peuvent sembler longs, mais il faut tenir compte du fait que les experts (chirurgiens et radiologues) ne sont pas monopolisés pendant ce temps de calcul : un calcul déporté sur un centre de calcul dédié accessible 24h/24 est donc envisageable. Les praticiens peuvent en effet se consacrer à d'autres tâches et laisser une machine ou un supercalculateur effectuer la segmentation et la fusion pendant le temps nécessaire. De plus, nous ne sommes pas en état d'urgence absolue et les experts estiment qu'il est raisonnable de produire une représentation numérique de l'abdomen d'un patient en quelques jours, voire en une semaine : il s'agit d'une période de planification d'une intervention chirurgicale.

#### 5.5.5/ COMPARAISON DES PERFORMANCES DES MÉTHODES UTILISANT OU NON L'INTELLIGENCE ARTIFICIELLE

C'est donc une approche basée sur les expériences les plus similaires avec la méthode DL-OV<sup>2</sup>ASSION qui nous a permis d'optimiser les segmentations finalement obtenues, que ce soit pour la gestion des conflits que pour la qualité globale des segmentations. Notre réseau avec entraînement classique obtient également des segmentations de meilleure qualité mais possède des difficultés pour gérer les conflits. Toutefois, il est intéressant de remarquer que, même si la méthode DL-OV<sup>2</sup>ASSION obtient la meilleure gestion des pixels en conflit par rapport aux autres méthodes IA et réseaux de neurones, comparée à d'autres méthodes n'utilisant pas d'IA, elle n'est pas la plus efficace. En effet, la gestion des conflits par le modèle v2 avec fusion des résultats des différentes méthodes par vote majoritaire (VM) (cf. Figure 4.4) obtient les meilleurs résultats pour la gestion des conflits.

De manière générale, les méthodes utilisant les réseaux de neurones obtiennent en moyenne des scores de Dice légèrement supérieurs à ceux obtenus par les méthodes n'utilisant pas l'IA. Cela s'explique par le fait que ces méthodes peuvent modifier légèrement les segmentations (lisser et affiner les bords des structures segmentées) en tant que post-traitement car elles réalisent une nouvelle segmentation à partir de la segmentation actuelle. Cela explique également pourquoi, dans certains cas, le pourcentage de pixels correctement étiquetés de certains réseaux est plus faible que la moyenne, mais que les scores de Dice sont meilleurs car les segmentations ont été correctement améliorées. C'est le cas pour notre réseau Deep Learning qui n'obtient que 57,96% de bonnes résolutions de pixels en conflit, mais qui obtient le score de Dice de 88,18%, en première position en termes de performances devant la méthode DL-OV<sup>2</sup>ASSION (qui suit de peu avec un score de Dice de 88,06%). Il serait alors intéressant, afin d'optimiser la performance de gestion des pixels en conflit et de la qualité des segmentations, de combiner notre réseau avec entraînement classique, s'occupant uniquement d'ajouter un post-traitement aux segmentations, avec le modèle v2 Fusion VM (Vote Majoritaire) réalisant l'étiquetage final des pixels en conflit. Il est également possible de combiner la méthode DL-OV<sup>2</sup>ASSION au modèle v2 Fusion VM, en ajoutant ses bons résultats de gestion de pixels en conflit aux autres résultats des méthodes déjà présentes, et d'utiliser son amélioration de la qualité des segmentations en tant que post-traitement. Nous aurions alors un processus de fusion des segmentations combinant des méthodes d'IA pour l'amélioration des segmentations et des méthodes utilisant ou non l'Intelligence Artificielle pour gérer les pixels en conflit.

Enfin, il convient de noter que notre nouvelle méthode de fusion DL-OV<sup>2</sup>ASSION est, contrairement aux autres méthodes, une méthode semi-automatique, dans le sens où plusieurs coupes d'un patient doivent déjà être segmentées manuellement pour être utilisées pendant la formation du réseau. Dans le cadre du projet SAIAD, cela ne nous pose pas de problème, car nous disposons déjà de ces coupes, utilisées dans le précédent processus de segmentation, qui est donc également semi-automatique. Nous ne pouvons pas, pour l'instant, effectuer une segmentation automatique en raison d'une part, de l'imprévisibilité de la tumeur et du rein déformé et d'autre part, du fait que le nombre de patients est limité. Mais, l'utilisation de notre architecture réseau avec un entraînement classique nous permet d'obtenir une méthode de fusion et d'amélioration des segmentations complètement automatique.

**Synthèse du chapitre****SYNTHÈSE**

Dans ce chapitre, nous avons présenté notre deuxième contribution pour la fusion de segmentations complémentaires, avec l'utilisation de techniques d'Intelligence Artificielle (IA). Deux méthodes ont été implémentées et testées, les deux utilisant un réseau de neurones pour la fusion de segmentations spécialement conçu pour cette contribution. Ce réseau, de type convolutif (CNN), est composé de sept couches de convolution, deux couches de pooling et une couche de sur-échantillonnage.

La première méthode combine ce réseau de neurones à un Raisonnement à Partir de Cas (RàPC). Le RàPC est alors présent afin de donner au réseau les cas à utiliser pour son entraînement, présents dans sa base de cas. Les cas récupérés étant les plus similaires au nouveau cas à fusionner. L'entraînement général du réseau est également spécifique, car il s'agit d'entraînements successifs en fonction de chaque cas, effectué dans l'ordre croissant des similarités obtenues, et le nombre d'itérations d'entraînement pour chacun de ces cas est calculé automatiquement en fonction de leur similarité.

La deuxième méthode utilise également ce même réseau mais avec un entraînement particulier par patient, appelé OV<sup>2</sup>ASSION. Cette méthode d'entraînement est basée sur le sur-apprentissage de certaines coupes voisines d'un même patient, coupes ayant été préalablement segmentées manuellement ou semi-manuellement par un expert. Ces coupes sont également choisies de manière homogène car elles sont espacées par un écart (*gap*) fixé et unique.

Ces deux méthodes, en plus de réaliser la fusion, permettent indirectement de gérer les éventuels pixels en conflit, en rendant une nouvelle segmentation fusionnée dépourvue d'erreur. Contrairement aux méthodes présentées dans le chapitre précédent, ces méthodes utilisant l'IA travaillent également sur l'ensemble des pixels des segmentations, permettant d'y ajouter un post-traitement et rendant donc des segmentations potentiellement améliorées.

Ces méthodes ont été testées pour la fusion de segmentations du rein pathologique et de la tumeur rénale sur un ensemble de 14 patients atteints de néphroblastome, dans le cadre du projet SAIAD. Ces segmentations ayant été précédemment segmentées par Deep Learning avec la méthode d'apprentissage particulière OV<sup>2</sup>ASSION. Les résultats démontrent une gestion des pixels en conflit cohérente avec les résultats des réseaux de fusion présents dans la littérature, mais surtout une grande amélioration de la qualité des segmentations.

# CONCLUSION ET PERSPECTIVES

## CONCLUSION GÉNÉRALE

Les travaux présentés dans ce mémoire ont consisté à proposer des solutions pour l'agrégation de segmentations complémentaires, intégrée dans la dernière couche de la plate-forme intitulée COLISEUM-3D du projet SAIAD : plate-forme dédiée à la segmentation d'images médicales et plus particulièrement du néphroblastome chez l'enfant. Cette étape de fusion permet alors d'agrèger les segmentations de différentes structures de l'abdomen calculées en parallèle par des techniques d'Intelligence Artificielle (IA), et permet d'obtenir la représentation définitive de l'ensemble de ces structures nécessaires pour une prise en charge optimale des jeunes patients atteints.

Le cœur de ce travail a permis de proposer différentes solutions pour la fusion de segmentations de deux structures : la fusion de la représentation de la tumeur rénale et de la représentation du rein pathologique. La première contribution est consacrée à la gestion des pixels dont l'étiquette est en conflit, suite aux processus de segmentation séparés des différentes structures anatomiques en présence, avec des techniques ne provenant pas d'IA. Ainsi, nous avons imaginé deux modèles exploitant un ensemble de 6 différentes méthodes indépendantes permettant de résoudre ces conflits. Certaines de ces méthodes ont été spécialement conçues pour cette contribution et d'autres ont été adaptées de la littérature, et l'ensemble de leurs résultats est alors combiné selon deux méthodes différentes possibles. Nous avons pu mettre en évidence l'efficacité de ces modèles en comparaison avec les méthodes qu'ils utilisent individuellement, ainsi que l'efficacité de la résolution des pixels en conflit de 71,96% pour le meilleur modèle.

Les contributions suivantes utilisent l'IA et plus particulièrement un réseau de neurones que nous avons spécialement conçu pour mener à bien la fusion de nos segmentations complémentaires. Notre première approche par IA, nommée DL-RàPC, combine notre réseau de neurones à un Règle à PC lui permettant d'apporter de la connaissance grâce à sa base de cas. Nous avons également conçu un processus d'entraînements spécifiques pour notre réseau, en fonction des données issues de cette base. Le réseau est ainsi appelé à être sur-entraîné sur les cas de fusion les plus similaires. Cependant, même si les premiers tests effectués avec les premiers dossiers médicaux anonymisés à notre disposition ont été concluants, les performances de notre implémentation de cette méthode se sont légèrement dégradées lorsque nous avons inclus la totalité des dossiers anonymisés des patients finalement (en fin de projet SAIAD) à notre disposition dans la base. Nous pensons toutefois être sur une voie prometteuse si nous améliorons la phase de sélection de cas similaires.

Ce sentiment est conforté par les bons résultats obtenus par notre seconde approche basée sur la même philosophie et l'utilisation du même réseau de neurones pour la fusion des segmentations. L'idée est également de préalablement sur-entraîner ce réseau sur des fusions similaires, via la méthode d'apprentissage appelée DL-OV<sup>2</sup>ASSION.

Ainsi, notre réseau de fusion entraîné avec la méthode d'apprentissage spécifique  $OV^2$  ASSION, permet de s'affranchir des grandes quantités d'images nécessaires pour l'apprentissage, en focalisant l'entraînement sur un unique patient. L'entraînement est alors réalisé sur quelques coupes d'un patient, ces coupes devant être les plus fidèles possible à la réalité (segmentations effectuées manuellement au préalable par un expert) et espacées régulièrement. Les résultats obtenus attestent de l'efficacité de la gestion des pixels en conflit, par rapport aux méthodes de fusion par réseau de neurones présents dans la littérature, avec 65,41% de bonnes résolutions. De plus, cette méthode permet d'améliorer significativement la qualité des segmentations, passant d'un score de Dice moyen de 86,06% au total pour nos 14 patients avant la fusion à un score de Dice de 88,06%, et est donc utile en tant que post-traitement. Cette méthode permet aux segmentations calculées de chaque patient de se rapprocher au mieux de la variabilité inter-opérateur calculées dans le cadre du projet, qui possède un score de Dice moyen de 90,64% pour la segmentation du rein pathologique et de la tumeur, ce qui permet de rendre les segmentations encore plus pertinentes pour être médicalement exploitables (sous couvert de validation par les personnels compétents). Toutefois, il est à noter qu'il s'agit d'une méthode semi-automatique, car son entraînement impose que certaines segmentations soient réalisées manuellement. Dans le cas où nous souhaiterions une méthode automatique pour l'amélioration de la qualité des segmentations, le réseau que nous avons conçu, utilisé avec un entraînement classique permet d'obtenir en moyenne des segmentations tout aussi justes que par la méthode DL- $OV^2$ ASSION.

De même, même si notre méthode DL- $OV^2$ ASSION surpasse d'autres méthodes de fusion par réseau de neurones, les modèles introduits en première contribution, et n'utilisant pas de réseau, sont plus efficaces dans la gestion des conflits. Cependant, il serait alors intéressant d'utiliser ces deux nouvelles contributions conjointement afin d'obtenir d'une part une meilleure gestion des pixels en conflit que par les modèles utilisant les méthodes classiques, et d'autre part afin d'obtenir une amélioration de la qualité des segmentations par la méthode DL- $OV^2$ ASSION. Les pistes suivantes offrent des perspectives d'amélioration des résultats de ces travaux de thèse.

## PERSPECTIVES

Ces travaux offrent différentes perspectives qui nous paraissent très intéressantes, autant sur l'amélioration des solutions que nous proposons que sur leur adaptation à d'autres segmentations futures qui seront prochainement réalisées dans le cadre du projet SAIAD.

### A - AMÉLIORATION DE LA MÉTHODE DL-RÀPC

Comme nous l'avons montré, notre méthode de fusion par Deep Learning couplée au RàPC a montré ses limites, aussi bien sur la partie de gestion des conflits que sur la qualité des segmentations rendues. Deux versions étaient proposées, la première (DL-RàPC v1) contenant l'ensemble des coupes de tous nos patients dans la base de cas, et la seconde version (DL-RàPC v2) ne contenant que les coupes du patient actuel précédemment utilisées pour la segmentation.

Logiquement, avec une base de cas plus conséquente, la première version de notre

méthode devrait obtenir de meilleurs résultats que la seconde version, ou au minimum équivalent. Cependant, la seconde a été plus performante. DL-RàPC v1, avec plus de possibilité de cas à remémorer, ne récupère alors pas forcément les cas les plus intéressants pour l'entraînement du réseau. En revanche DL-RàPC v2 n'a le choix qu'entre certaines coupes du patient actuel, et donc récupère des cas plus similaires. Le calcul de similarité pour l'extraction des cas est alors à revoir, car le RàPC de la première version peut sélectionner des cas de fusion d'autres patients, potentiellement moins similaires à d'autres cas. Le calcul de similarité est composé de deux parties, la première compare les informations des segmentations avec le score de Dice, et la seconde compare les informations des images scanner associées avec la métrique MSSIM.

Dans un premier temps, il serait judicieux d'ajouter une pondération à ces parties, afin de privilégier l'une ou l'autre de ces informations. De plus, le calcul de MSSIM est réalisé en fonction d'un nombre de fenêtres à définir. Plus le nombre de fenêtres est important, plus leur taille est petite et plus les images peuvent être comparées sur des détails plus précis et plus fins. En revanche, cela augmente également considérablement le temps de calcul. Actuellement, nous avons 16 fenêtres pour le calcul du MSSIM pour la similarité. Ce nombre pourrait être agrandi afin d'affiner la comparaison entre deux images scanner. Cependant, il pourrait également être intéressant d'obtenir des informations sur la similitude globale entre les deux images. Nous pourrions alors calculer deux MSSIM, un contenant un grand nombre de fenêtres et un autre contenant un très faible nombre. Si qu'une seule fenêtre est choisie, cela reviendrait à calculer le SSIM : pour rappel, le MSSIM est simplement la moyenne entre chaque SSIM calculé avec les fenêtres définies. Là encore, ces informations, à différents niveaux, peuvent être pondérées. L'équation 60 présente le potentiel nouveau calcul de similarité en fonction des segmentations  $S$  et  $S'$ , des images scanner  $I$  et  $I'$  et du nombre de fenêtres pour le calcul du premier MSSIM ( $M$ ) et du second ( $M'$ ), sachant que l'un doit être largement plus petit que l'autre :

$$s(S, S', I, I', M, M')_{M' < M} = \frac{1}{2} \left( \alpha \text{Dice}(S, S') + \beta \left( \frac{1}{2} (\gamma \text{MSSIM}(I, I', M) + \delta \text{MSSIM}(I, I', M')) \right) \right) \quad (60)$$

$\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  étant des poids à définir.

Même si la méthode DL-RàPC v2 obtient de meilleurs résultats, elle présente, comme DL-RàPC v1, des difficultés à retranscrire la segmentation du rein pathologique. En effet, pour les deux versions nous obtenons un score de Dice du rein inférieur au score avant la fusion. Ceci s'explique par le manque d'entraînement. Actuellement, nous avons défini un nombre maximal d'itérations lors de l'entraînement à 3000 par cas. Ce nombre devrait être augmenté afin que la segmentation du rein, structure plus petite que la tumeur, puisse être correctement retranscrite. De plus, nous utilisons 3 cas pour l'entraînement du réseau de neurones. Il serait intéressant de pousser l'entraînement avec 4 ou 5 cas.

La principale problématique de ces changements de valeurs (nombre de fenêtres pour le calcul de MSSIM, augmentation du nombre d'itérations pour l'entraînement du réseau, augmentation du nombre de cas à prendre en compte, ...) est l'augmentation considérable du temps de calcul. Il faudra alors certainement trouver un moyen de les maîtriser afin qu'ils ne deviennent pas problématiques.

## B - AMÉLIORATION DE LA MÉTHODE DL-OV<sup>2</sup>ASSION ET COMBINAISON AVEC LE MODÈLE NON-IA

Nous avons pu montrer que la méthode de fusion DL-OV<sup>2</sup>ASSION obtenait de bons résultats pour gérer les pixels en conflit et permet une amélioration de la qualité des segmentations. Cependant, le score de Dice obtenu pour l'ensemble des patients est similaire au score de Dice de son réseau avec l'utilisation d'un entraînement classique, et elle obtient un écart type élevé par rapport aux autres méthodes. En effet, en se focalisant sur les résultats de gestion de conflits par patient, la méthode présente des difficultés à gérer les conflits de certains patients. Il serait alors intéressant de pouvoir adapter l'entraînement en fonction de chaque patient à fusionner. Actuellement, nous avons un nombre d'itérations fixe de 30000 qui pourrait être modulable afin de permettre au réseau de s'entraîner plus efficacement sur des cas plus problématiques. Nous pourrions, par exemple en amont, calculer la similarité moyenne du nouveau patient à fusionner en fonction des patients existants, et adapter le nombre d'itérations en conséquence.

La méthode DL-OV<sup>2</sup>ASSION étant semi-automatique, nous pouvons remplacer cette méthode d'IA par l'utilisation de son réseau de neurones avec un entraînement classique, afin de la rendre complètement automatique. En effet, l'utilisation de ce réseau pour la fusion des segmentations d'un patient avec entraînement sur l'ensemble des segmentations des autres patients a également permis d'obtenir une bonne amélioration du score de Dice des segmentations.

Nous avons montré que notre première contribution sur l'agrégation de différentes méthodes dites classiques dans un unique modèle permet de mieux gérer les pixels en conflit que des méthodes de fusion par réseau de neurones. Cependant, notre méthode DL-OV<sup>2</sup>ASSION permet d'augmenter considérablement la qualité des segmentations. Nous pouvons alors aisément définir une nouvelle méthode de fusion combinant ces deux contributions : permettant ainsi d'obtenir un résultat avec une amélioration du score de Dice ainsi qu'une bonne gestion des pixels en conflit. La méthode DL-OV<sup>2</sup>ASSION peut être utilisée en tant que post-traitement des segmentations, en ne prenant pas en compte les pixels en conflit qui ont été gérés préalablement. Par ailleurs, comme représenté sur la Figure 5.7, il serait envisageable que cette méthode soit incluse dans le modèle de fusion, en tant que méthode concurrente des six autres, et ainsi intégrer son résultat dans le vote majoritaire effectué pour arbitrer les conflits.

## C - ADAPTATION DES MÉTHODES POUR LA FUSION DES SEGMENTATIONS D'AUTRES STRUCTURES

L'objectif du projet SAIAD est d'arriver à terme à obtenir une représentation par segmentation du rein pathologique, de la tumeur, mais également des cavités rénales et des vaisseaux sanguins avec l'artère et la veine rénale. Au total, ce sont cinq structures qui devront être gérées lors de l'agrégation des segmentations. Un cas particulier est la segmentation des cavités rénales qui doivent se superposer à la segmentation du rein pathologique, il ne doit alors pas y avoir de conflit. Pour les autres structures, des conflits peuvent être présents, et nos systèmes doivent être en mesure de tous les résoudre.

Dans le cas de l'utilisation des modèles de gestion des conflits par des méthodes classiques, des modifications mineures sont à apporter afin de gérer les conflits provenant de toutes ces structures. La méthode ADS2 peut gérer l'ensemble de ces structures

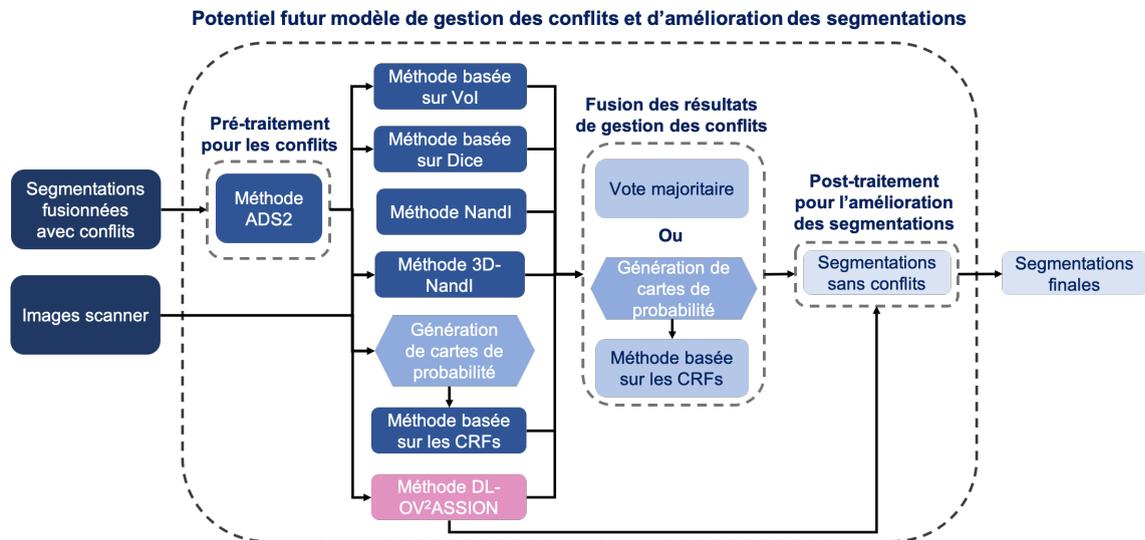


FIGURE 5.7 – Nouvelle version du modèle de gestion des conflits avec l'ajout d'une méthode d'IA permettant également l'amélioration de la qualité des segmentations.

et leurs étiquettes, car elle se contente de considérer les segmentations de la coupe supérieure et inférieure et résoudre les conflits ayant les mêmes étiquettes dans ces segmentations voisines. Il en est de même pour les méthodes basées sur le *Vol* et le score de Dice : prenant en compte les différentes étiquettes possibles dans leur algorithme ICM. La méthode basée sur les CRFs peut également prendre en considération toutes ces segmentations. Il y aura alors plus de cartes de probabilités à prendre en charge à l'entrée de son système : en effet, il faut une carte de probabilité par structure possible. Cependant, des modifications sont à apporter aux méthodes NandI et 3D-NandI. En effet la Formule 37 est à modifier car elle ne prend en compte que deux étiquettes. Il doit être possible de comparer les résultats de similarité de l'ensemble des étiquettes possibles qui sont présentes dans le voisinage effectif du pixel considéré. Le vote majoritaire pourra également être problématique et devra être capable de traiter les cas d'égalité. En effet, il se pourra que deux ou plusieurs étiquettes obtiennent le même nombre de voix. Une première approche pourrait consister à ordonner les méthodes de résolution de conflit en fonction de leurs performances en termes de qualité d'arbitrage.

Dans le cas de l'utilisation des méthodes d'IA pour la fusion de toutes les structures, l'entrée du réseau de neurones sera alors plus conséquente, avec une image à  $n + 1$  canaux,  $n$  étant le nombre de structures. Si la base de cas du RàPC est utilisée, elle sera également plus conséquente car un cas représentera alors l'ensemble des cinq segmentations calculées ainsi que les cinq vérités terrain correspondant, en plus de l'image scanner.



# MA BIBLIOGRAPHIE PERSONNELLE

## ARTICLES EN REVUES RÉFÉRENCÉES DONT JE SUIS L'AUTEURE PRINCIPALE

- [Corbat20a] Corbat Lisa, Henriët Julien and Lapayre Jean-Christophe. Conflict management in the fusion of complementary segmentations of deformed kidneys and nephroblastoma. *Journal of Medical Image Analysis* - IF 8.88 - Q1. Elsevier - 2020, Volume : 60
- [Corbat20b] Corbat Lisa, Nauval Mohammad, Henriët Julien and Lapayre Jean-Christophe. A fusion method based on deep learning and case-based reasoning which improves the resulting medical image segmentations. *Journal of Expert Systems with Applications* - IF 4.29 - Q1. Elsevier - 2020, Volume : 147
- [Corbat20c] Corbat Lisa, Henriët Julien, Chaussy Yann and Lapayre Jean-Christophe. Fusion of multiple segmentations of medical images using OVASSION and Deep Learning methods : Application to CT-Scans for tumoral kidney. *Journal of Computers in Biology and Medicine* - IF 2.29 - Q2. Elsevier - 2020

## ARTICLE EN REVUE RÉFÉRENCÉE COMME DEUXIÈME AUTEURE

- [Marie19] Marie Florent, Corbat Lisa, Chaussy Yann, Henriët Julien and Lapayre Jean-Christophe. Segmentation of deformed kidneys and nephroblastoma using case-based reasoning and convolutional neural network. *Journal of Expert Systems with Applications* - IF 4.29 - Q1. Elsevier - 2019, Volume : 127, Pages : 282–294

## PUBLICATION EN CONFÉRENCE COMME DEUXIÈME AUTEURE

- [Marie18] Marie Florent, Corbat Lisa, Delavelle Thibault, Chaussy Yann, Henriët Julien and Lapayre Jean-Christophe. Segmentation of Kidneys Deformed by Nephroblastoma

using Case-Based Reasoning. ICCBR 2018, 26th International conference on Case-based reasoning - Core B. 2018, Pages : 233–248 - Best Student Paper Award

### AUTRE ARTICLE EN REVUE RÉFÉRENCÉE

[Chaussy20] Chaussy Yann, Vieille Lorédane, Lacroix Elise, Lenoir Marion, Marie Florent, Corbat Lisa, Henriet Julien and Auber Frédéric. 3D Reconstruction of Wilms' Tumor and Kidneys in Children : Variability, Usefulness and Constraints. Journal of Pediatric Urology - IF 1.58 - Q2. Elsevier - 2020

# BIBLIOGRAPHIE

- [Aamodt1994] Agnar Aamodt and Enric Plaza. Case-based reasoning : Foundational issues, methodological variations, and system approaches. *AI communications*, 7(1) :39–59, 1994.
- [Aljabar2009] Paul et al. Aljabar. Multi-atlas based segmentation of brain images : atlas selection and its effect on accuracy. *Neuroimage*, 46(3) :726–738, 2009.
- [Alkama2015] Sadia Alkama, Youssef Chahir, and Daoud Berkani. Label maps fusion for the marginal segmentation of multi-component images. *Neural Network World*, 25(4) :405, 2015.
- [Artaechevarria2009] Xabier Artaechevarria, Arrate Munoz-Barrutia, and Carlos Ortiz-de Solórzano. Combination strategies in multi-atlas image segmentation : application to brain mr data. *IEEE transactions on medical imaging*, 28(8) :1266–1277, 2009.
- [Attig2012] Anja Attig and Petra Perner. Incremental learning of the model for watershed-based image segmentation. In *International Workshop on Combinatorial Image Analysis*, pages 209–222. Springer, 2012.
- [Badrinarayanan2017] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet : A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12) :2481–2495, 2017.
- [Bandhyopadhyay2013] Samir Kumar Bandhyopadhyay and Tuhin Utsab Paul. Automatic segmentation of brain tumour from multiple images of brain mri. *Int J Appl Innovat Eng Manage (IJAEM)*, 2(1) :240–8, 2013.
- [Besag1986] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society : Series B (Methodological)*, 48(3) :259–279, 1986.
- [Beucher1993] Serge Beucher and Fernand Meyer. The morphological approach to segmentation : the watershed transformation. *Mathematical morphology in image processing*, 34 :433–481, 1993.
- [Brown1992] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4) :325–376, 1992.
- [Cabria2017] Iván Cabria and Iker Gondra. Mri segmentation fusion for brain tumor detection. *Information Fusion*, 36 :1–9, 2017.
- [Chen2017] Liang-Chieh et al. Chen. Deeplab : Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4) :834–848, 2017.

- [CHRU2019]** Yann Chaussy, Lorédane Salvi, Élise Lacroix, Marion Lenoir, and Frédéric Auber. Rapport technique CHRU, SAIAD. Technical report, Centre Hospitalier Régional Universitaire de Besançon, 10 2019.
- [Everingham2015]** Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge : A retrospective. *International journal of computer vision*, 111(1) :98–136, 2015.
- [Feng2017]** Yuncong Feng, Xuanjing Shen, Haipeng Chen, and Xiaoli Zhang. Segmentation fusion based on neighboring information for mr brain images. *Multimedia Tools and Applications*, 76(22) :23139–23161, 2017.
- [Frucci2008]** Maria Frucci, Petra Perner, and Gabriella Sanniti Di Baja. Case-based-reasoning for image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 22(5) :829–842, 2008.
- [Gallo2011]** Luigi Gallo, Alessio Pierluigi Placitelli, and Mario Ciampi. Controller-free exploration of medical image data : Experiencing the kinect. In *2011 24th international symposium on computer-based medical systems (CBMS)*, pages 1–6. IEEE, 2011.
- [Geman1987]** Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *Readings in computer vision*, pages 564–584. Elsevier, 1987.
- [Glorot2010]** Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.
- [Glorot2011]** Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- [Gorthi2013]** Gorthi Subrahmanyam. *Multi Atlas Fusion Methods for Medical Image Segmentation*. PhD thesis, EPFL (5747), 2013.
- [Harrabi2012]** Rafika Harrabi and Ezzedine Ben Braiek. Color image segmentation using multi-level thresholding approach and data fusion techniques : application in the breast cancer cells images. *EURASIP Journal on Image and Video Processing*, 2012(1) :11, 2012.
- [Hazirbas2016]** Caner Hazirbas, Lingni Ma, Csaba Domokos, and Daniel Cremers. Fuseset : Incorporating depth into semantic segmentation via fusion-based cnn architecture. In *Asian conference on computer vision*, pages 213–228. Springer, 2016.
- [He2016]** Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [Hong2015]** Seunghoon Hong, Hyeonwoo Noh, and Bohyung Han. Decoupled deep neural network for semi-supervised semantic segmentation. In *Advances in neural information processing systems*, pages 1495–1503, 2015.

- [Hu2019a] Xinxin Hu, Kailun Yang, Lei Fei, and Kaiwei Wang. Acnet : Attention based network to exploit complementary features for rgb-d semantic segmentation. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1440–1444. IEEE, 2019.
- [Hu2019b] Yang Hu, Andrea Soltoggio, Russell Lock, and Steve Carter. A fully convolutional two-stream fusion network for interactive image segmentation. *Neural Networks*, 109 :31–42, 2019.
- [IncaNephro07] Fédération nationale des centres de lutte contre le cancer et La ligue nationale contre le cancer. Comprendre le néphroblastome, 2007. <http://onc-orient.org/wp-content/uploads/2015/04/INCANEPHRO.pdf>.
- [Islam2013] Saiful Islam and Majidul Ahmed. Implementation of image segmentation for natural images using clustering methods. *Int J Emerg Technol Adv Eng*, 3(3) :175–80, 2013.
- [Jenkinson2001] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2) :143–156, 2001.
- [Jia2014] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe : Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678, 2014.
- [Jiang2017] Jindong Jiang, Zhijun Zhang, Yongqian Huang, and Lunan Zheng. Incorporating depth into both cnn and crf for indoor semantic segmentation. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 525–530. IEEE, 2017.
- [Jiang2018] Jindong Jiang, Lunan Zheng, Fei Luo, and Zhijun Zhang. Rednet : Residual encoder-decoder network for indoor rgb-d semantic segmentation. *arXiv preprint arXiv :1806.01054*, 2018.
- [Kamnitsas2017] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36 :61–78, 2017.
- [Kaste2008] Sue C Kaste, Jeffrey S Dome, Paul S Babyn, Norbert M Graf, Paul Grundy, Jan Godzinski, Gill A Levitt, and Helen Jenkinson. Wilms tumour : prognostic factors, staging, therapy and late effects. *Pediatric radiology*, 38(1) :2–17, 2008.
- [Khan2015] Ahmad Khan and Muhammad Arfan Jaffar. Genetic algorithm and self organizing map based fuzzy hybrid intelligent method for color image segmentation. *Applied soft computing*, 32 :300–310, 2015.
- [Khelifi2017] Lazhar Khelifi and Max Mignotte. A Novel Fusion Approach Based on the Global Consistency Criterion to Fusing Multiple Segmentations. *IEEE Transactions on Systems, Man, and Cybernetics : Systems*, 47(9) :2489–2502, 2017.

- [**Khelifi2017a**] Lazhar Khelifi and Max Mignotte. Efa-bmfm : A multi-criteria framework for the fusion of colour image segmentation. *Information Fusion*, 38 :104–121, 2017.
- [**Khelifi2017b**] Lazhar Khelifi and Max Mignotte. A multi-objective decision making approach for solving the image segmentation fusion problem. *IEEE Transactions on Image Processing*, 26(8) :3831–3845, 2017.
- [**Kolodner1983**] Janet L Kolodner. Reconstructive memory : A computer model. *Cognitive science*, 7(4) :281–328, 1983.
- [**Kolodner1992**] Janet L Kolodner. An introduction to case-based reasoning. *Artificial intelligence review*, 6(1) :3–34, 1992.
- [**Krahenbuhl2011**] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011.
- [**Kybic2003**] Jan Kybic and Michael Unser. Fast parametric elastic image registration. *IEEE transactions on image processing*, 12(11) :1427–1442, 2003.
- [**Lafferty2001**] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In *ICML '01- Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289, 2001.
- [**Lecun1988**] Yann LeCun, D Touresky, G Hinton, and T Sejnowski. A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school*, volume 1, pages 21–28. CMU, Pittsburgh, Pa : Morgan Kaufmann, 1988.
- [**Lecun1989**] Yann LeCun et al. Generalization and network design strategies. In *Connectionism in perspective*, volume 19, 1989.
- [**Lee2015**] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570, 2015.
- [**Li2016**] Guanbin Li and Yizhou Yu. Deep contrast learning for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 478–487, 2016.
- [**Liang2015**] Ming Liang and Xiaolin Hu. Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3367–3375, 2015.
- [**Lin2016**] Guosheng Lin, Chunhua Shen, Anton Van Den Hengel, and Ian Reid. Efficient piecewise training of deep structured models for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2016.
- [**Litjens2017**] Geert et al. Litjens. A survey on deep learning in medical image analysis. *Medical image analysis*, 42 :60–88, 2017.

- [Long2015] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [Luo2016] Yuansheng Luo, Lu Yang, Ling Wang, and Hong Cheng. Efficient cnn-crf network for retinal image segmentation. In *International Conference on Cognitive Systems and Signal Processing*, pages 157–165. Springer, 2016.
- [Macqueen1967] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1(14), pages 281–297. Oakland, CA, USA, 1967.
- [Martin2001] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 416–423, 02 2001.
- [Martin2004] David R Martin, Charless C Fowlkes, and Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence*, 26(5) :530–549, 2004.
- [Marie2018] Florent Marie, Lisa Corbat, Thibault Delavelle, Yann Chaussy, Julien Henriet, and Jean-Christophe Lapayre. Segmentation of kidneys deformed by nephroblastoma using case-based reasoning. In *ICCBR 2018, 26th International conference on Case-based reasoning*, pages 351–365, Stockholm, Sweden, July 2018.
- [Marie2019] Florent Marie. *COLISEUM-3D, Une plate-forme innovante pour la segmentation d’images médicales par Raisonnement à Partir de Cas (RàPC) et méthodes d’apprentissage de type Deep Learning*. PhD thesis, Université de Bourgogne Franche-Comté, 12 2019.
- [Marie2019] Florent Marie, Lisa Corbat, Yann Chaussy, Thibault Delavelle, Julien Henriet, and Jean-Christophe Lapayre. Segmentation of deformed kidneys and nephroblastoma using case-based reasoning and convolutional neural network. *Expert Systems with Applications*, 2019.
- [Meilua2003] Marina Meilă. Comparing clusterings by the variation of information. In *Learning theory and kernel machines*, pages 173–187. Springer, 2003.
- [Meilua2007] Marina Meilă. Comparing clusterings—an information based distance. *Journal of multivariate analysis*, 98(5) :873–895, 2007.
- [Mignotte2008] Max Mignotte. Segmentation by fusion of histogram-based  $k$ -means clusters in different color spaces. *IEEE Transactions on image processing*, 17(5) :780–787, 2008.
- [Mignotte2010] Max Mignotte. A label field fusion bayesian model and its penalized maximum rand estimator for image segmentation. *IEEE Transactions on Image Processing*, 19(6) :1610–1624, 2010.

- [Mignotte2014a]** Max Mignotte. A label field fusion model with a variation of information estimator for image segmentation. *Information Fusion*, 20 :7–20, 2014.
- [Mignotte2014b]** Max Mignotte and Charles H elou. A precision–recall criterion based consensus model for fusing multiple segmentations. *Int J Signal Process Image Process Pattern Recognit*, 7(3) :61–82, 2014.
- [Mitchell2010a]** Harvey B Mitchell. *Image fusion : theories, techniques and applications*. Springer Science & Business Media, 2010.
- [Mitchell2010b]** HB Mitchell. Staple : Simultaneous truth and performance level estimation. In *Image Fusion*, pages 233–236. Springer, 2010.
- [Nguyen2018]** Dac Cong Tai Nguyen, Said Benameur, Max Mignotte, and Fr ed eric Lavoie. Superpixel and multi-atlas based fusion entropic model for the segmentation of x-ray images. *Medical image analysis*, 48 :58–74, 2018.
- [Noh2015]** Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015.
- [Perner1999]** Petra Perner. An architecture for a cbr image segmentation system. *Engineering Applications of Artificial Intelligence*, 12(6) :749–759, 1999.
- [Plantaz15]** Dominique Plantaz. N ephroblastome ou tumeur de wilms, corpus m edical de la facult e de m edecine de grenoble. <http://www-sante.ujf-grenoble.fr/SANTE/corpus/disciplines/pedia/canyped/144c/lecon144c.htm>, 2015. Accessed on 10.01.2020.
- [Provost1998]** Foster Provost and R Kohavi. Glossary of terms. *Journal of Machine Learning*, 30(2-3) :271–274, 1998.
- [Qu2017]** Liangqiong Qu, Shengfeng He, Jiawei Zhang, Jiandong Tian, Yandong Tang, and Qingxiong Yang. Rgb-d salient object detection via deep fusion. *IEEE Transactions on Image Processing*, 26(5) :2274–2285, 2017.
- [Rand1971]** William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336) :846–850, 1971.
- [Reisbeck1989]** Christopher K Reisbeck and Roger C Schank. Inside case-based reasoning. *Hillsdale NJ*, 1989.
- [Rohlfing2004a]** Torsten Rohlfing, Robert Brandt, Randolph Menzel, and Calvin R Maurer Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4) :1428–1442, 2004.
- [Rohlfing2007]** Torsten Rohlfing and Calvin R Maurer. Shape-based averaging. *IEEE Transactions on Image Processing*, 16(1) :153–161, 2007.
- [Ronneberger2015]** Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

- [**Rosenblatt1958**] Frank Rosenblatt. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6) :386, 1958.
- [**Rumelhart1988**] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Neurocomputing : foundations of research*, page 696–699, 1988.
- [**Senthilkumaran2009**] N Senthilkumaran and R Rajesh. Edge detection techniques for image segmentation-a survey of soft computing approaches. *International journal of recent trends in engineering*, 1(2) :250—254, 2009.
- [**Shen2002**] Dinggang Shen and Christos Davatzikos. Hammer : hierarchical attribute matching mechanism for elastic registration. *IEEE transactions on medical imaging*, 21(11) :1421–1439, 2002.
- [**Simonyan2014**] *Very deep convolutional networks for large-scale image recognition*, 2014.
- [**SiteWebCFI2020**] Cfi web site. <https://www.cfi.ch>. Accessed on 10.01.2020.
- [**SiteWebCFI2020**] Cfi web site. <https://www.cfi.ch>. Accessed on 10.01.2020.
- [**SiteWebIDOin2020**] Ido-in web site. <https://www.ido-in.com>. Accessed on 10.01.2020.
- [**SiteWebITK2020**] Itk web site. <http://www.itksnap.org>. Accessed on 10.01.2020.
- [**SiteWebKidney2020**] Kidney structures web site. [https://commons.wikimedia.org/wiki/File:Kidney\\_PioM.png?uselang=fr](https://commons.wikimedia.org/wiki/File:Kidney_PioM.png?uselang=fr). Accessed on 10.01.2020.
- [**Soesanti2011**] Indah et al. Soesanti. Optimized fuzzy logic application for mri brain images segmentation. *International Journal of Computer Science & Information Technology*, 3(5) :137–146, October 2011.
- [**Sun2019**] Yuxiang Sun, Weixun Zuo, and Ming Liu. Rtfnet : Rgb-thermal fusion network for semantic segmentation of urban scenes. *IEEE Robotics and Automation Letters*, 4(3) :2576–2583, 2019.
- [**Szegedy2015**] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabino-vich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [**Tang2016b**] Youbao Tang and Xiangqian Wu. Saliency detection via combining region-level and pixel-level predictions with cnns. In *European Conference on Computer Vision*, pages 809–825. Springer, 2016.
- [**Tang2016a**] Youbao Tang, Xiangqian Wu, and Wei Bu. Deeply-supervised recurrent convolutional neural network for saliency detection. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 397–401. ACM, 2016.
- [**Thong2018**] William Thong, Samuel Kadoury, Nicolas Piché, and Christopher J Pal. Convolutional networks for kidney segmentation in contrast-enhanced ct scans. *Computer Methods in Biomechanics and Biomedical Engineering : Imaging & Visualization*, 6(3) :277–282, 2018.

- [Torres2018] Helena R Torres, Sandro Queiros, Pedro Morais, Bruno Oliveira, Jaime C Fonseca, and Joao L Vilaca. Kidney segmentation in ultrasound, magnetic resonance and computed tomography images : A systematic review. *Computer methods and programs in biomedicine*, 157 :49–67, 2018.
- [Tosta2017] Thaína A Azevedo Tosta, Paulo Rogério Faria, Leandro Alves Neves, and Marcelo Zanchetta do Nascimento. Computational method for unsupervised segmentation of lymphoma histological images based on fuzzy 3-partition entropy and genetic algorithm. *Expert Systems with Applications*, 81 :223–243, 2017.
- [Unnikrishnan2007] Ranjith Unnikrishnan, Caroline Pantofaru, and Martial Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 6 :929–944, 2007.
- [Wang2004] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment : from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4) :600–612, 2004.
- [Weichert2013] Frank et al. Weichert. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5) :6380–6393, 2013.
- [Xiao2018] Fen et al. Xiao. Msdsn : Multi-scale deep neural network for salient object detection. *IET Image Processing*, 12(11) :2036–2041, 2018.
- [Zaitoun2015] Nida M Zaitoun and Musbah J Aqel. Survey on image segmentation techniques. *Procedia Computer Science*, 65 :797–806, 2015.
- [Zhao2017] Bo Zhao, Jiashi Feng, Xiao Wu, and Shuicheng Yan. A survey on deep learning-based fine-grained object classification and semantic segmentation. *International Journal of Automation and Computing*, 14(2) :119–135, 2017.
- [Zhang2018a] Wenkai Zhang, Hai Huang, Matthias Schmitz, Xian Sun, Hongqi Wang, and Helmut Mayer. Effective fusion of multi-modal remote sensing data in a fully convolutional network for semantic labeling. *Remote Sensing*, 10(1) :52, 2018.
- [Zheng2015] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 1529–1537, 2015.
- [Zhou1988] Yi-Tong Zhou and Rama Chellappa. Computation of optical flow using a neural network. In *IEEE International Conference on Neural Networks*, volume 1998, pages 71–78, 1988.

# NOMENCLATURE

ADS2	ADjacent Segmented Slices
CBR	Case-based Reasoning
CFI	Centre de Formation Informatique
CHRU	Centre Hospitalier Régional Universitaire
CNN	Convolutional Neural Network
COLLISEUM-3D	COLaborative platform with artificial Intelligence for SEgmentation of tUmoral kidney in Medical images in 3D
CRF	Conditional Random Field
CT-scan	Computerized-Tomography scan
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
EFA-BMFM	Multi-criteria Fusion Model Based on the Entropy-Weighted Formula Approach
EPFL	École Polytechnique Fédérale de Lausanne
FC	Fully Connected
FCM	Fusion Convolution Module
FCN	Fully Convolutional Networks
GCE	Global Consistency Error
ICM	Iterated Conditional Modes
ILSVRC	ImageNet Large Scale Visual Recognition Competition
LS	Learning Set
MOBFM	Multi-objective Optimization Based-Fusion Model
MSSIM	Mean Structural SIMilarity
NandI	Neighbours and Intensity
OVASSION	Overlearning Vector for Valid Sparse SegmentatIOns
PRI	Probabilistic Rand Index

ReLU	Rectified Linear Unit
RI	Rand Index
RàPC	Raisonnement à Partir de Cas
SAIAD	Segmentation Automatique de reins tumoraux chez l'enfant par Intelligence Artificielle Distribuée
SBA	Shape based Averaging
SIOP	Société Internationale d'Oncologie Pédiatrique
SSIM	Structural Similarity
STAPLE	Simultaneous Truth And Performance Level Estimation
VGG	Visual Geometry Group
Vol	Variation of Information
WFA	Weighted Formula Approach

# TABLE DES FIGURES

1.1	Schéma du rein et de sa vascularisation (issue de <i>KidneyStructures</i> by P.M. Jaworski [SiteWebKidney2020]). . . . .	16
1.2	Illustration de la représentation d'images de taille $6 \times 6$ en noir et blanc et en couleurs. . . . .	19
1.3	Exemple d'une image scanner d'une coupe d'un patient atteint de néphroblastome (a) en temps vasculaire et (b) en temps tardif. . . . .	20
1.4	Représentation de la segmentation sémantique du néphroblastome à partir d'une image scanner en temps vasculaire. . . . .	21
1.5	Exemple d'un cycle basique d'un système RàPC. . . . .	23
1.6	Représentation 3D de la segmentation sémantique du néphroblastome, obtenue à partir des segmentations successives des coupes 2D des images scanner. La couleur marron représente les reins, le rouge l'artère, le bleu la veine, le vert la tumeur, et en jaune les cavités rénales. . . . .	25
1.7	Représentation en pyramide holographique de la segmentation sémantique 3D du néphroblastome - réalisée par l'entreprise CFI dans la cadre du projet SAIAD. . . . .	25
1.8	Organisation de la plate-forme COLISEUM-3D. . . . .	26
1.9	La matrice de confusion. . . . .	29
2.1	Exemple de la fusion de 3 segmentations de taille $5 \times 5$ par vote majoritaire. . . . .	38
2.2	Exemple de la fusion de 3 segmentations de taille $5 \times 5$ par vote majoritaire pondéré globalement et localement. . . . .	38
3.1	Le perceptron. . . . .	56
3.2	Le perceptron multicouche. . . . .	58
3.3	Le principe général de l'apprentissage. . . . .	58
3.4	Le principe de rétropropagation du gradient. . . . .	59
3.5	Le principe de la convolution. . . . .	60
3.6	Le principe du <i>max pooling</i> . . . . .	61
3.7	L'architecture VGG16. . . . .	62
3.8	L'architecture du réseau de segmentations FCN-32s, FCN-16s et FCN-8s. . . . .	63
3.9	Le réseau pour la détection de saillances avec une couche de fusion pour la concaténation des résultats des deux sorties [Li2016]. . . . .	65

3.10	Le réseau de fusion à trois couches pour la détection de saillances [Tang2016b]. . . . .	67
3.11	Le module de fusion FCM pour la détection de saillances [Xiao2018]. . . . .	67
3.12	Le réseau de fusion inclus dans le réseau pour la détection de saillances [Qu2017]. . . . .	68
3.13	Le réseau pour la segmentation interactive incluant un réseau de fusion profond [Hu2019b]. . . . .	68
3.14	L'architecture de FuseNet, réseau de type encodeur-decodeur avec deux flux encodeur qui sont fusionnés [Hazirbas2016]. . . . .	69
4.1	Fusion de segmentations calculées du rein pathologique et de la tumeur, appartenant à une coupe de l'abdomen d'un patient. . . . .	76
4.2	La fusion de segmentations et gestion des conflits avec la méthode basée sur les CRFs. . . . .	82
4.3	Première version du modèle de gestion des conflits pour la fusion de segmentations rénales tumorales. . . . .	83
4.4	Deuxième version du modèle de gestion des conflits, avec un pré-traitement, pour la fusion de segmentations rénales tumorales. . . . .	83
4.5	Exemple de création de cartes de probabilités pour la fusion par CRF dans le cas d'un pixel en conflit dans une segmentation $3 \times 3$ . . . . .	85
4.6	Évolution de la performance des méthodes basées sur le Vol et le score de Dice par rapport au nombre de coupes adjacentes considérées. . . . .	89
4.7	Évolution de la performance des méthodes NandI et 3D-NandI par rapport à la taille du rayon choisie. . . . .	90
4.8	Exemple d'un résultat de fusion de segmentations d'une coupe d'un patient, avec un avant/après de la coupe ainsi qu'un zoom sur une zone conflictuelle. . . . .	95
5.1	L'architecture générale du réseau de neurones proposée pour la fusion de segmentations complémentaires. . . . .	102
5.2	Présentation générale du système DL-RàPC pour la fusion de segmentations. . . . .	103
5.3	Représentation des données définissant un cas de la base du RàPC. . . . .	104
5.4	La représentation 3D du néphroblastome : chaque ligne noire représente les coupes sélectionnées pour l'entraînement du CNN, séparées par un <i>gap</i> similaire entre chaque coupe choisie. . . . .	108
5.5	Résultats de la fusion de segmentations par la méthode DL-OVASSION et DL-RàPC v1 sur quelques coupes de différents patients. . . . .	116

5.6	Comparaison des différents résultats en fonction des méthodes utilisées, sur trois coupes de trois patients différents. Sur la colonne de gauche nous avons les vérités terrain ainsi que la segmentation avec conflits, et sur la colonne de droite nous avons de gauche à droite : les résultats des méthodes DL-OV <sup>2</sup> ASSION, de Qu et al. [Qu2017], de Tang et al. [Tang2016b] et de Xiao et al. [Xiao2018] sur la ligne du haut, ainsi que les résultats du modèle V2 avec fusion VM, de Nandl ainsi que de la métrique Vol sur la ligne du bas. . . . .	117
5.7	Nouvelle version du modèle de gestion des conflits avec l'ajout d'une méthode d'IA permettant également l'amélioration de la qualité des segmentations. . . . .	127



# LISTE DES TABLES

4.1	Calcul de la variabilité inter-opérateurs entre deux experts en score de Dice, pour la segmentation de la tumeur et du rein pathologique en fonction des différents patients [CHRU2019]. . . . .	86
4.2	Résultats de la segmentation (avec le score de Dice) de la tumeur et du rein pathologique, par le réseau FCN-8s avec la méthode d'apprentissage OV <sup>2</sup> ASSION. . . . .	88
4.3	Pourcentage de pixels en conflit pouvant être résolu par la méthode ADS2 sur l'ensemble de nos patients. . . . .	90
4.4	Performance moyenne des différentes méthodes sur plusieurs patients avec la première version du modèle de gestion des conflits. . . . .	91
4.5	Performances finales pour le premier modèle sur différents patients avec les méthodes de fusion par vote majoritaire et basée sur le CRF. . . . .	92
4.6	Performance moyenne des différentes méthodes sur plusieurs patients avec la deuxième version du modèle de gestion des conflits. . . . .	93
4.7	Performances finales pour le deuxième modèle sur différents patients avec les méthodes de fusion par vote majoritaire et basée sur le CRF. . . . .	94
4.8	Présentation du Dice moyen (moyenne sur les 14 patients) des segmentations avant et après la gestion des conflits. . . . .	94
4.9	Comparaison des différentes images prises en compte dans le processus de gestion des conflits pour chaque méthode. . . . .	96
5.1	Exemple des vecteurs possibles pour un <i>gap</i> de 2 ( $V_2$ ) <sub><i>i</i></sub> et pour 60 coupes d'un patient ( $n = 60$ ). . . . .	109
5.2	Pourcentage des pixels en conflit correctement résolus par les méthodes DL-OV <sup>2</sup> ASSION et DL-RàPC, en comparaison avec d'autres réseaux utilisés dans la littérature ainsi que la méthode par agrégation de multiples méthodes classiques (deuxième modèle, présenté dans la Figure 4.4, avec fusion par vote majoritaire). . . . .	112
5.3	Résultat du score de Dice moyen (moyenne du score de Dice entre les deux structures) avant et après la fusion par les méthodes DL-OV <sup>2</sup> ASSION et DL-RàPC, et en comparaison avec d'autres réseaux utilisés dans la littérature. . . . .	114
5.4	Résultats du score Dice moyen obtenu pour la structure de la tumeur et du rein pathologique par les méthodes DL-OV <sup>2</sup> ASSION et DL-RàPC, et en comparaison avec d'autres réseaux utilisés dans la littérature. . . . .	114

- 5.5 Comparaison des résultats obtenus (Pourcentage des pixels en conflit correctement résolus et Dice) entre les méthodes par IA et les méthodes non-IA. 115

