Segmentation automatique de la colonne vertébrale lombaire à partir d'images à résonance magnétique par combinaison de réseau de neurones convolutifs et coupe de graphe

par

Firas BEN DHAOU

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE
AVEC MÉMOIRE EN GÉNIE DE LA PRODUCTION AUTOMATISÉE
M. Sc. A.

MONTRÉAL, LE 13 AOÛT 2019

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Ismail Ben ayed, Directeur de Mémoire
Département de génie de la production automatisée à l'École de technologie supérieure

M. Jose Dolz, Co-directeur
Département de génie de la production automatisée à l'École de technologie supérieure

M. Carlos Vazquez, Président du Jury
Département de génie logiciel et des TI à l'École de technologie supérieure

M. Vincent Duchaine, Examinateur Externe
Département de génie des systèmes à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 1 AOÛT 2019

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

C'est avec grand plaisir que je réserve ces quelques lignes en reconnaissance de tous ceux qui ont contribué à la réalisation de ce mémoire.

Tout d'abord, je voudrais exprimer ma gratitude à mon directeur de recherche, le professeur Ismail BEN AYED, pour sa patience, sa disponibilité et surtout ses conseils judicieux, qui ont contribué à ma réflexion. Je le remercie pour son attention aux détails lors de son encadrement et la vérification de mes progrès ainsi que pour ses encouragements tout au long de la période de mes études.

J'aimerais également remercier mon co-directeur, M. Jose Dolz, qui m'a fourni les outils nécessaires à la réussite de mes études universitaires. M. Jose Dolz a été le premier à m'expliquer la nouvelle technologie des réseaux de neurones profonds, leurs développements et la manière de les exploiter dans mon étude.

Je voudrais exprimer ma gratitude aux collègues de LIVIA, Laboratoire d'imagerie, de vision et d'intelligence artificielle, qui m'ont donné leur soutien moral et intellectuel tout au long de mon processus. Je voudrais surtout remercier Imtiaz Ziko pour ses conseils qui ont grandement facilité mon travail ainsi que son aide précieuse dans la réalisation des logiciels d'expérimentation.

Je voudrais également exprimer ma reconnaissance à mes chers parents qui m'ont permis d'arriver à ce niveau, mes deux sœurs et mon beau-frère pour leur patience et leurs encouragements.

Je remercie tous mes amis pour m'avoir encouragé et supporté tout le long de ce parcours.

À tous les membres du jury, j'offre mes remerciements, mon respect et ma gratitude.

SEGMENTATION AUTOMATIQUE DE LA COLONNE VERTÉBRALE LOMBAIRE À PARTIR D'IMAGES À RÉSONANCE MAGNÉTIQUE PAR COMBINAISON DE RÉSEAU DE NEURONES CONVOLUTIFS ET COUPE DE GRAPHE

Firas BEN DHAOU

RÉSUMÉ

Le mal de dos, le mal du siècle comme beaucoup de gens le décrivent, est un terme général pour une maladie potentiellement grave et l'un des problèmes médicaux les plus courants dans le monde. Il peut se produire à n'importe quel endroit au niveau de la colonne vertébrale. Pour identifier l'origine d'une douleur et déterminer si un traitement est nécessaire, les experts dans ce domaine se basent sur l'analyse des images médicales telles que l'IRM et CT-scan pour identifier les zones endommagées ou les anomalies.

Un examen de radiologie classique est une tâche compliquée et coûteuse en temps précieux pour le malade et le médecin. De plus, dans certaines situations, l'identification de ces anomalies à l'oeil nu n'est pas toujours évidente, ce qui nécessite l'application de certaines techniques de traitement d'image afin de guider l'expert à réaliser un bon diagnostic. Parmi les techniques les plus employées dans ce domaine nous citons la segmentation d'images qui permet de délimiter et d'identifier les zones d'intérêt. Une segmentation précise et robuste des structures est une condition préalable au diagnostic assisté par ordinateur et à l'identification des anomalies. Elle peut également être utilisée pour la planification assistée par ordinateur et la simulation d'une chirurgie. Cependant, malgré les inventions technologiques dans ce domaine, les approches utilisées pour la segmentation des images médicales restent limitées de point de vue performance et nécessitent l'intervention d'un expert humain. Récemment, les réseaux de neurones convolutifs (RNC) ont montré des performances exceptionnelles surtout dans le domaine de traitement d'images médicales surpassant les approches de segmentation existantes dans la littérature.

C'est dans ce contexte que s'inscrit ce travail, qui vise à proposer une nouvelle approche pour la segmentation des vertèbres et des disques intervertébraux de la partie lombaire de la colonne vertébrale basée sur la combinaison des réseaux de neurones convolutifs avec la segmentation par coupe de graphe appliquée sur des images IRM 3D. Au lieu d'appliquer directement les RNC pour obtenir une segmentation finale, la technique proposée utilise les cartes de probabilités générées par le réseau de neurones comme initialisation pour la méthode de coupe de graphe afin de raffiner la segmentation initiale. Afin d'améliorer les résultats dans le cas de la segmentation multi-classes, nous avons utilisé l'algorithme $\alpha - expansion$ qui constitue une extension de la coupe de graphe appliquée sur des images multi-classes. L'approche a été évaluée quantitativement sur deux bases de données différentes utilisées dans la compétition annuelle MICCAI pour la segmentation des vertèbres et des disques. Nous avons aussi évalué qualitativement notre méthode sur une nouvelle base de données de dix sujets qui contient des annotations manuelles multi-classes des deux structures ; vertèbres et disques. L'évaluation ex-

périmentale, basée sur le coefficient de similarité de Dice et la distance de Hausdorff, montre que notre approche réalise de bonnes performances sur les trois bases de données.

Mots-clés: Colonne vertébrale, image résonance magnétique, segmentation de l'image, réseau de neurones convolutif (RNC), coupe de graphe, α-expansion

AUTOMATIC SEGMENTATION OF THE LUMBAR SPINE FROM MAGNETIC RESONANCE IMAGES BY COMBINING CONVOLUTIONAL NEURAL NETWORKS AND GRAPH CUT

Firas BEN DHAOU

ABSTRACT

Back pain, the sickness of the century as many people describe it, is a general term for a potentially serious illness and one of the most common medical problems in the world. It can occur anywhere in the spine. To identify the origin of pain and determine if treatment is needed, experts in this area rely on the analysis of medical images such as MRI and CT-scan to identify damaged areas or anomalies.

The conventional radiological examination is a complicated and expensive task in precious time for the patient and the doctor. Moreover, in certain situations, identifying these anomalies with the naked eye is not always obvious, which requires the application of certain image processing techniques to guide the expert to make a good diagnosis. Among the most used techniques in this area we mention the image segmentation that delimits and identifies areas of interest. Precise and robust structural segmentation is a prerequisite for computer-assisted diagnosis and anomaly identification. It can also be used for computer-aided planning and simulated surgery. However, despite the technological inventions in this field, the approaches used for the segmentation of medical images are limited in terms of performance and require the intervention of a human expert. Recently, convolutional neural networks (CNN) have shown outstanding performance especially in the field of medical image processing surpassing existing segmentation approaches in the literature.

It is in this context that this work aims to propose a new approach for the joint segmentation of the vertebrae and intervertebral discs of the lumbar spinal column based on the combination of convolutional neural networks with graph cut segmentation applied on 3D MRI images. Instead of directly applying the CNNs to obtain a final segmentation, the proposed technique uses the probability maps generated by the neural network as initialization for the graph cut method in order to refine the initial segmentation. To improve the results in the case of multi-label segmentation, we used the $\alpha - expansion$ algorithm which is an extension of the graph cut applied to multi-label images. The approach was quantitatively evaluated on two different databases used in the annual MICCAI competition for segmentation of vertebrae and discs. We also qualitatively evaluated our method on a new database of ten subjects that contains manual multi-label annotations of both structures ; vertebrae and discs. The experimental evaluation, based on Hausdorff distance and Dice similarity coefficient, shows that our approach performs well on all three databases.

Keywords: Spine, Image segmentation, Image magnetic resonance, Convolutional neural network (CNN), Graph cut, α-expansion

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

LISTE DES ALGORITHMES

Page

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

CRF	Conditional Random Fields
CS	Coefficient de similarité
HD	Distance de Hausdorff
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IRM	Imagerie par résonance magnétique
MICCAI	Medical Image Computing and Computer Assisted Intervention
MNIST	Modified National Institute of Standards and Technology
ReLU	Unité Rectifiée Linéaire
RNC	Réseau de neurones convolutif
TDM	Tomodensitométrie
VGG	Groupe de géométrie visuelle
VT	Vérité terrain

INTRODUCTION

0.1 Contexte

0.1.1 Le mal de dos

Le dos est une structure composée d'os, d'articulations, de ligaments et de muscles. Les entorses, les muscles contractés, les disques de rupture et les irritations des articulations peuvent entraîner des douleurs au dos. Le mal de dos peut aussi être accompagné d'une gamme de symptômes, y compris la fatigue, la tension ou la raideur musculaire, et la faiblesse.

De nombreuses études ont montré que les maux de dos sont très fréquents dans la société moderne (Guzmán *et al.* (2001); Pengel *et al.* (2003); Panjabi (2003); Schneider *et al.* (2005); Katz (2006); Chou *et al.* (2007); Dagenais *et al.* (2008); Donelson *et al.* (2012); Hoy *et al.* (2010, 2012, 2014); Knecht *et al.* (2017)). Plus de 80% des gens vont éprouver un mal de dos au courant de leur vie (Frymoyer (1988); Andersson (1999); Gross *et al.* (2006); Katz (2006); Rubin (2007); Lacasse *et al.* (2017)).

Le mal de dos peut se produire n'importe où dans la colonne vertébrale. Il est soit lombaire ou dorsal, selon la zone de douleur. La lombalgie est une douleur localisée au bas du dos tandis que la dorsalgie est une douleur entre les épaules et la taille. De nos jours, le mal de dos reste l'une des raisons les plus fréquentes pour consulter un médecin, un chiropraticien ou un physiothérapeute. C'est aussi considéré comme la principale cause de diminution de mobilité physique et d'absentéisme au travail (Lidgren (2003)). Ceci crée un énorme fardeau économique pour les individus, leurs familles, leurs communautés, les employeurs et la société (Hoy *et al.* (2010)). En raison de la baisse de productivité des travailleurs, des arrêts et des prestations d'invalidité associées qui en découlent, les maux de dos et leurs affectations connexes ont un impact économique important sur la société. Bien que la récurrence de la douleur au dos ne

soit généralement pas associée à un problème physique accru, elle entraîne généralement une augmentation de nombres de visites chez le médecin et de tests médicaux. Ces facteurs sont responsables, en majeure partie, de la hausse du coût de système de santé.

Au Canada, les dépenses médicales liées à la lombalgie sont estimées entre 6 et 12 milliards de dollars par année et continuent d'augmenter (Brown *et al.* (2005)). Selon un article publié dans Le Devoir [1] en 2006, l'indemnisation pour les jours de congé de travail liés aux maux de dos coûte environ 650 millions de dollars chaque année.

Aux États-Unis, 149 millions de journées de travail sont perdues chaque année à cause des maux de dos. Les coûts combinés d'absences du travail, de la perte de productivité et des soins médicaux sont estimés entre 100 et 200 milliards de dollars chaque année (Freburger *et al.* (2009)). La région du dos la plus affectée est le bas du dos, car c'est la partie la plus mobile et la plus stressée de la colonne vertébrale. Cette région supporte le plus de poids et subit le plus de contraintes mécaniques.

0.1.2 Anatomie de la colonne vertébrale

La colonne vertébrale, aussi nommée rachis, constitue l'axe central du corps humain. Elle se compose de 7 vertèbres cervicales au cou, 12 vertèbres dorsales dans la partie supérieure et la partie médiane du tronc et 5 vertèbres lombaires dans la partie inférieure (figure 0.1). Ces 24 vertèbres reposent sur le sacrum et le bassin. Ils sont séparés par des disques qui servent d'amortisseurs et de joints souples. Ces disques absorbent les chocs et protègent la colonne vertébrale contre les traumatismes. Les courbures des segments cervicaux, thoraciques et lombaires rendent la colonne vertébrale plus élastique et plus résistante à la compression due à la posture verticale. Un système ligamentaire entoure la colonne vertébrale. Combiné avec les tendons et les muscles, ce système fournit un support naturel pour aider à protéger la colonne

1. https ://www.ledevoir.com/societe/117349/metro-boulot-et-mal-de-dos

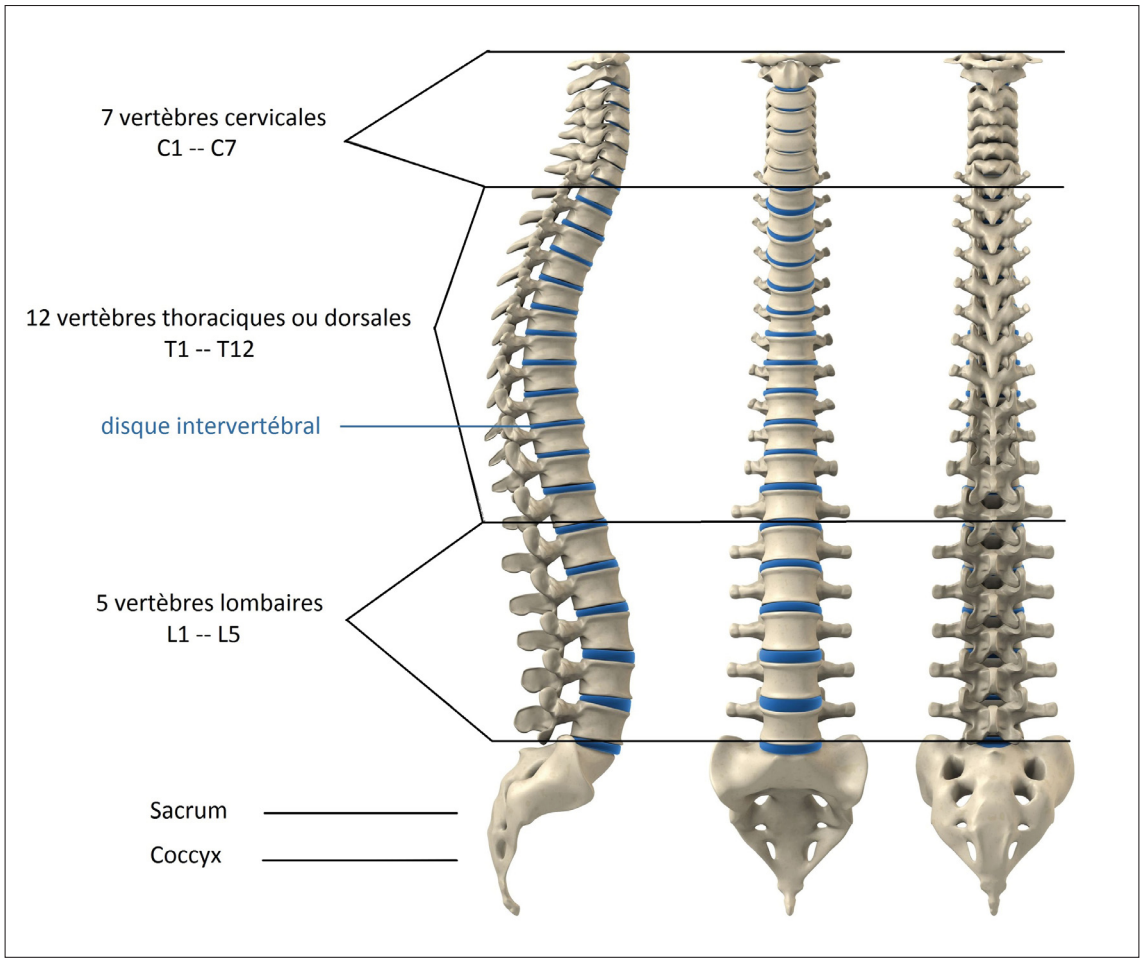

Figure 0.1 Anatomie de la colonne vertébrale, adaptée de (Themis Medica, 2019).

contre les blessures. Les ligaments aident à maintenir la stabilité des articulations tant durant le repos que durant le mouvement et aident à prévenir les blessures d'hyperextension et d'hyperflexion suite à des mouvement excessifs.

0.1.3 Anomalies de la colonne vertébrale

La colonne vertébrale peut être affectée par plusieurs anomalies résultant des maladies, d'infections ou de malformations (figure 0.2). Certaines de ces anomalies sont dues à des malformations congénitales (Lonstein (1999); Kaplan *et al.* (2005)), mais la plupart sont la conséquence d'une mauvaise posture, de la sédentarité, de l'obésité, du soulèvement d'objets lourds, ou du

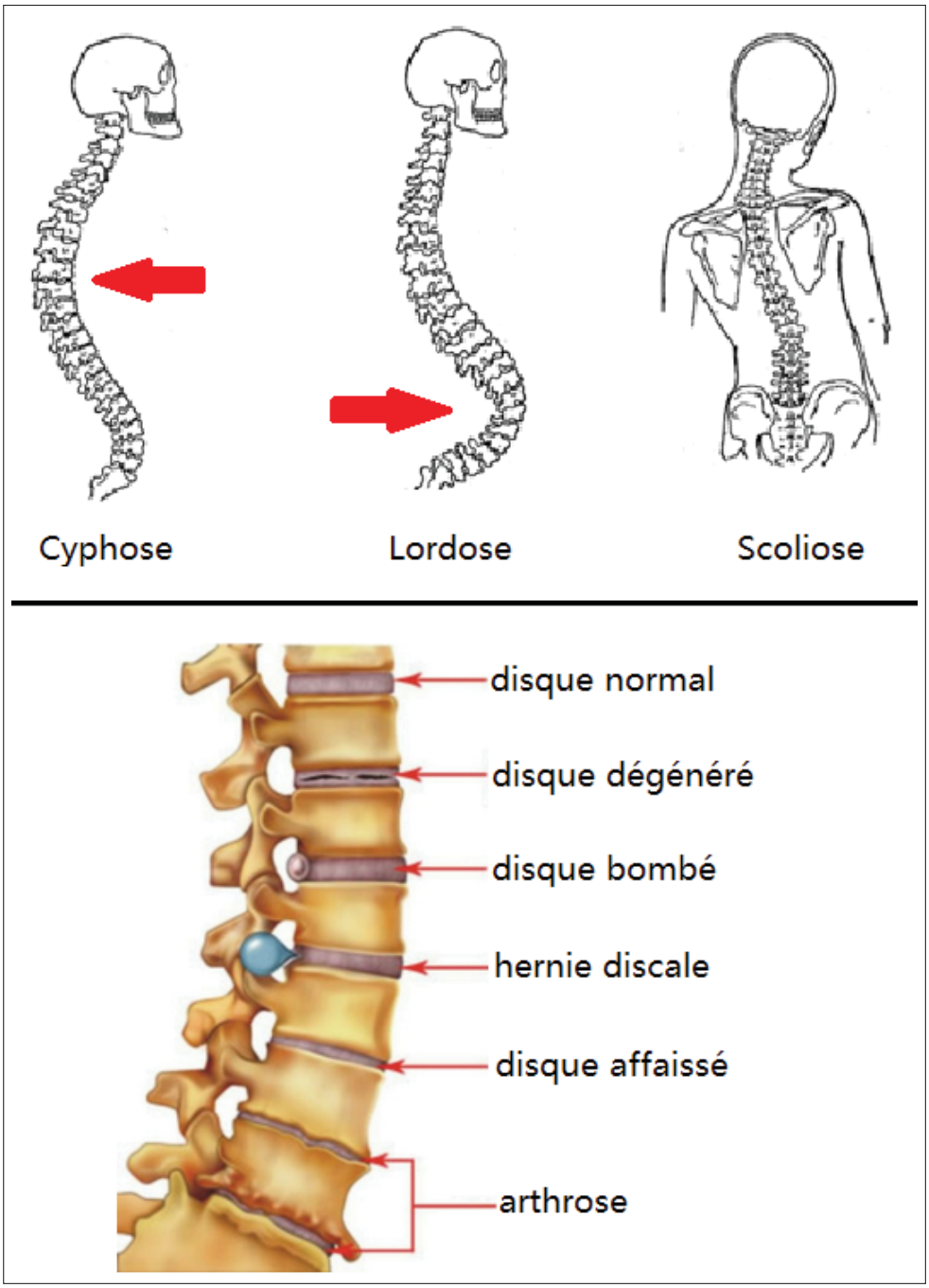

Figure 0.2 Anomalies de la colonne vertébrale, adaptée de (Georges Dolisi, 2019; Pinterest, 2019).

vieillissement, entre autres (Lonstein & Carlson (1984); Wiggins *et al.* (2003)). L'habitude de tenir le dos trop courbé, par exemple, provoque une tension sur les muscles de la colonne ver-

tébrale ce qui modifie le centre de gravité du corps et provoque une déformation de la courbure de la colonne vertébrale. La faiblesse ou la contraction excessive des muscles peuvent entraîner également un déséquilibre des vertèbres. Les anomalies de la colonne vertébrale peuvent être globales ou locales en fonction des parties endommagées. Les anomalies globales sont celles qui affectent un segment ou la colonne entière, tandis que les anomalies locales sont celles qui affectent une structure individuelle comme une vertèbre ou un disque.

Les anomalies du rachis globales peuvent être aussi classées en trois catégories en fonction de la déformation qui modifie la courbure normale de la colonne vertébrale. La première est la lordose (Been & Kalichman (2014)). C'est une courbe concave (vers l'intérieur) de la colonne vertébrale au bas du dos. Cette déformation peut être héréditaire, mais elle est généralement causée par une mauvaise posture ou un excès de poids. La deuxième est la cyphose, qui est une exagération convexe de la région dorsale donnant au dos une position trop arrondie et penchée. Ce déséquilibre peut être génétique, dû à la mauvaise posture ou dû à des changements dégénératifs liés au vieillissement. La troisième est la scoliose (Coillard & Rivard (1996)), qui est une courbure latérale de la colonne vertébrale d'un côté du corps. Elle provoque également la rotation des vertèbres individuelles autour de l'axe longitudinal de la colonne vertébrale. Cette déformation peut être liée à une malformation congénitale à la naissance ou une maladie neuromusculaire ou osseuse.

Ces anomalies globales peuvent être liées à des maladies locales affectant les vertèbres ou les disques de la colonne vertébrale. Parmi les maladies qui affectent les disques, on peut citer le disque bombé qui se produit lorsqu'un disque affaibli ou détérioré augmente de volume et sort de sa position normale. Si le patient ne reçoit aucun traitement, la situation peut s'aggraver et le disque bombé peut se transformer en disque hernié. La hernie est la saillie d'une partie du noyau gélatineux mou (l'intérieur du disque) de sa position normale en déchirant l'anneau fibreux. La dégénérescence discale est une autre maladie susceptible de provoquer des douleurs

dorsales (Modic *et al.* (1988); Luoma *et al.* (2000)). Les disques perdent leur teneur en eau et, par conséquent, leur hauteur diminue et se rapprochent des vertèbres. Les disques se frottent et perdent la capacité d'absorber les chocs, provoquant une douleur massive au patient lorsqu'il fait des efforts.

Les vertèbres lombaires sont les plus grandes de la colonne vertébrale. Elles supportent tout le poids et la pression du haut du corps, ce qui en fait la partie la plus vulnérable aux dommages. Les maladies affectant cette partie touchent tout le monde, indépendamment de l'âge ou de la condition physique, et sont la principale cause des douleurs au bas du dos.

0.1.4 Les différentes modalités de l'imagerie médicale

Pour diagnostiquer des maladies et déterminer si un traitement est nécessaire, les professionnels de la santé doivent souvent examiner l'intérieur du corps humain, ce qui est impossible à faire à l'œil nu. Les os, les tissus, les organes, les cellules et presque toute l'anatomie peuvent être visualisés sur des différentes modalités d'images médicales. L'imagerie médicale est généralement conçue pour la visualisation scientifique. Elle est également utilisée pour les études cliniques et la planification du traitement. Elle fournit des informations plus détaillées qui aident à diagnostiquer rapidement les maladies et à sauver de nombreux cas urgents.

L'imagerie médicale est un outil de diagnostic et de dépistage qui permet de préparer une intervention ou de suivre l'évolution d'une pathologie. La radiographie, la résonance magnétique et la tomodensitométrie sont les techniques radiographiques les plus utilisées en diagnostic (figure 0.3). Cependant, ces différentes modalités n'utilisent pas les mêmes techniques et ne sont pas utilisées aux mêmes fins. Toutes ces techniques offrent des images tridimensionnelles du corps humain et la différence entre elles est liée aux phénomènes physiques avec lesquels elles sont acquises.

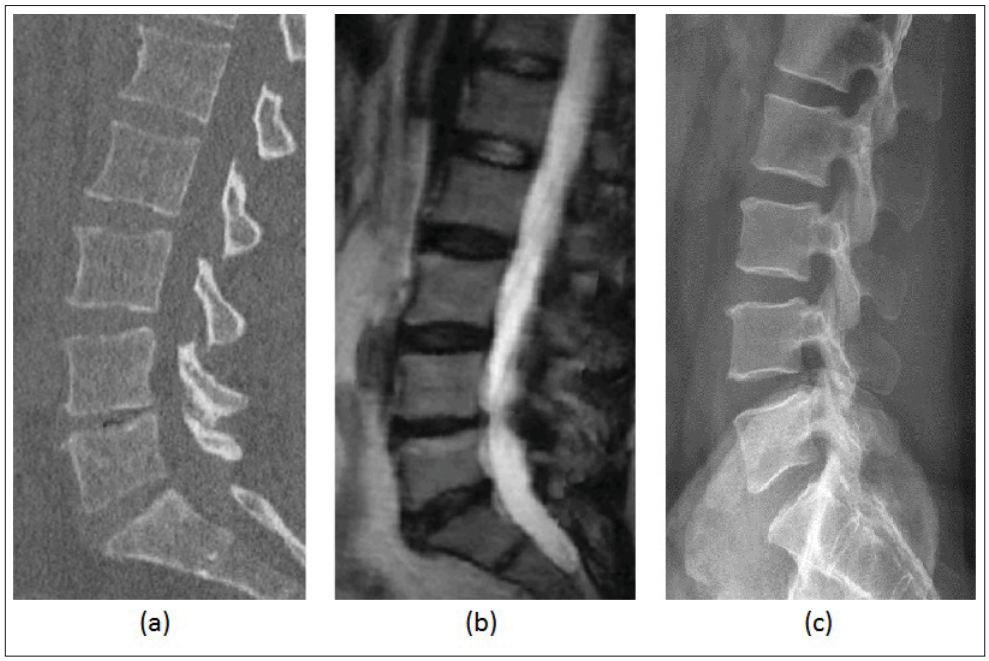

Figure 0.3 Images médicales de la partie lombaire de la colonne vertébrale acquises avec différentes modalités. (a) image TDM (b) image IRM (c) image de rayon X.

La radiographie utilise des rayons X qui produisent et envoient un certain nombre d'électrons à une certaine vitesse. En traversant les différentes couches du corps humain, ces rayons seront absorbés différemment par les tissus en fonction de leur densité.

Les os, les calcifications, certaines tumeurs et d'autres structures denses apparaissent blancs ou légers parce qu'ils arrêtent très bien les électrons. Les tissus mous et les fissures dans les os laissent passer les rayons ce qui rend ces parties plus sombres dans l'image.

Ces différents niveaux d'absorption des rayons X par les particules corporelles permettent de former une image qui révèle l'architecture des organes. Ces images seront recueillies par un film radio-graphique qui est souvent remplacé par un détecteur électronique permettant la numérisation de l'image.

Les radiographies permettent de détecter les déplacements, les fractures et les fissures osseuses. Elles sont également utilisées pour vérifier les courbures vertébrales et pour contrôler la progression de l'arthrose n'importe où sur la squelette. Les rayons X peuvent également détecter l'eau et les tumeurs dans les poumons, déceler une hypertrophie cardiaque, et également dans la mammographie pour la détection de la structure mammaire anormale. Les rayons X peuvent ne pas montrer autant de détails qu'une image produite avec des techniques plus sophistiquées, mais elles restent l'outil d'imagerie le plus rapide et le plus commun utilisé pour évaluer un problème orthopédique.

La tomodensitométrie (TDM ou CT scan) est un outil d'imagerie qui combine les rayons X avec la technologie informatique. Une succession d'images du corps sont prises à 360° grâce à des capteurs qui mesurent les degrés restants d'intensité des rayons X après leurs absorption partielle. Ces images, coronales et sagittales, seront utilisées pour reconstruire des images 2D ou 3D ce qui ajoutera des informations anatomiques et des éléments diagnostiques. Les différentes intensités d'absorption permettent d'identifier la matière grasse, l'eau, l'air ou tout autre élément important. L'image numérisée sera représentée par différentes nuances de gris selon les différents degrés de densité. La TDM aide à visualiser la taille, la forme et la position des structures internes du corps, tels que les organes, les os, les tissus mous et les vaisseaux sanguins. Elle peut être utilisée pour diagnostiquer plus facilement le cancer, les maladies cardiaques, les troubles musculo-squelettiques, les traumatismes et même les maladies infectieuses.

Les maladies vertébrales sont généralement diagnostiquées par la radiographie ou la tomodensitométrie, en raison de leur capacité à montrer clairement les structures osseuses. Dans la littérature, la majorité des approches de localisation et de segmentation vertébrales sont centrées sur ces types d'images. D'autre part, ces modalités ne peuvent pas être utiles pour visualiser d'autres structures telles que les ligaments ou les disques intervertébraux, ce qui nécessite l'utilisation d'un autre type d'image médicale.

L'imagerie par résonance magnétique (IRM) est une technologie d'imagerie médicale qui utilise des ondes radio et un champ magnétique pour créer des images détaillées d'organes et de tissus. Des ondes radio sont pulsées à travers le patient ce qui sert à varier les mini-champs magnétiques du corps. Les électrons sont stimulés, et sortent de l'équilibre, en s'opposant à la traction du champ magnétique. Lorsque le champ radio est désactivé, les capteurs IRM sont capables de détecter l'énergie libérée lorsque les électrons se réalignent avec le champ magnétique. Le temps nécessaire pour que les électrons se réalignent avec le champ magnétique, ainsi que la quantité d'énergie libérée, changent en fonction de l'environnement et de la nature chimique des molécules.

L'IRM s'est révélée très efficace pour diagnostiquer un certain nombre d'anomalies en montrant la différence entre les tissus des organes normaux et des organes affectés. Les IRM sont particulièrement bien adaptés à l'imagerie des parties non osseuses et des tissus mous du corps. Contrairement à la radiographie et à la tomodensitométrie, cette technique n'utilise pas les rayons X. Le cerveau, la moelle épinière et les nerfs, ainsi que les muscles, les ligaments et les tendons, sont nettement plus visibles avec l'IRM qu'avec la radiographie et la TDM. En outre, la plupart des anomalies qui affectent la colonne vertébrale et qui sont la principale cause des maux de dos chez les patients sont liées aux structures discales. Des malformations telles que l'hernie, la dégénérescence et le glissement des disques ne sont visibles que dans ce type d'image médicale. C'est pourquoi, dans la littérature, presque toutes les techniques et les études de recherche visant à identifier ou à segmenter des disques intervertébraux, utilisent cette modalité.

0.1.5 Segmentation des images médicales

Habituellement, les experts analysent directement l'image obtenue à partir d'un organe ou d'un tissu en regardant l'image sur un écran ou en visualisant une succession de coupes bidimen-

sionnelles. La décision dans ce cas est purement qualitative et ne dépend que de la connaissance de l'expert.

L'utilisation de logiciels dédiés à l'analyse d'images médicales est nécessaire pour l'extraction de paramètres quantitatifs des formes des organes, la comparaison et la détection de changements entre différentes images d'un même patient ou en comparant avec un patient en bonne santé. Ces outils permettent également de fusionner des informations acquises par différentes modalités qui peuvent aider à clarifier des informations cachées. Ils permettent également d'étudier l'aspect dynamique des organes par une visualisation d'une succession d'images dans une séquence temporelle. Cela peut aider à mesurer le mouvement et identifier les dysfonctionnements des organes.

Ces outils sont utiles en imagerie médicale puisqu'ils peuvent réduire le temps d'analyse et aider à visualiser plus de détails. Souvent, la détection d'organes est difficile en raison de la complexité de leur structure et la manque de netteté entre les tissus voisins. Ceci peut causer un manque de détection d'une anomalie, ce qui peut entraîner une paralysie ou la mort du patient. Le traitement de ces images nécessite aussi une forte concentration et une longue durée de la part du radiologue. Tout ça démontre l'importance de développer et d'améliorer des logiciels de diagnostic pour mieux identifier et localiser les anomalies.

Un problème fondamental dans l'analyse des images médicaux est la segmentation. La tâche principale de la segmentation d'une image est d'extraire les zones d'intérêts, à travers un processus automatique ou semi-automatique. Elle sert à identifier les limites, les formes, les contours, les mouvements des organes, et ainsi déceler les régions anormales. Il existe plusieurs techniques de segmentation qui diffèrent les unes des autres de façon significative.

De nombreux journaux et articles ont catégorisé les techniques les plus répandues pour la segmentation d'images (Pham *et al.* (2000); Withey & Koles (2008); Dey *et al.* (2010); Kaur & Kaur

(2014); Zaitoun & Aqel (2015); Norouzi *et al.* (2014); Kamdi & Krishna (2012)). Ces techniques peuvent être classées en trois catégories selon deux critères : les régions et les contours.

Dans la segmentation basée sur les régions, les différentes zones sont construites en associant ou en dissociant les pixels voisins. Elle fonctionne sur le principe de l'homogénéité, en considérant que les pixels voisins dans une région ont des caractéristiques similaires et différentes de ceux des autres régions. Chaque pixel est comparé aux pixels voisins pour vérifier leur similarité selon la couleur, la texture, le niveau de gris et la forme. En cas d'une similarité, un pixel particulier est ajouté aux pixels de la région d'intérêt.

Dans la segmentation basée sur les contours, les différentes zones sont construites en analysant les variations d'intensité des pixels voisins. Les changements soudains et significatifs des niveaux d'intensité entre les pixels voisins dans une certaine direction sont appelés arêtes. Ils provoquent une discontinuité entre les régions. Les pixels sont classés comme contour ou non en fonction du résultat de la convolution entre un filtre de bord et l'image. Les pixels qui ne sont pas séparés par un contour sont affectés aux mêmes régions.

La troisième catégorie combine les deux techniques utilisées dans la segmentation par régions et la segmentation par contours pour tirer profit des avantages des deux techniques. Elle est la catégorie la plus répandue dans la littérature et la recherche.

Au fil des années, diverses techniques semi-automatiques et automatiques ont été mises au point pour la localisation et la segmentation des vertèbres et des disques intervertébraux. Malgré leur efficacité dans divers domaines, les techniques semi-automatiques nécessitent toujours une intervention manuelle et une vérification des résultats. Les approches automatiques, de leur côté, ont surmonté ces limitations et ont prouvé leur capacité à fournir des résultats sans aucune intervention humaine. Cependant, dans la littérature, presque toutes ces méthodes sont basées sur des paramètres spécifiques liés à la forme ou au contexte de la cible étudiée. En d'autres

termes, elles ont utilisé des fonctions mathématiques prenant en compte des valeurs telles que l'intensité des pixels ou les caractéristiques dimensionnelles de la cible. Ces méthodes ont également fourni des résultats acceptables dans de nombreuses tâches médicales. Cependant, elles ne sont pas fiables dans des situations où elles doivent traiter des images présentant des organes présentant diverses anomalies. Elles sont également faibles dans les situations où les images sont affectées par le bruit.

Les dernières années ont été marquées par des avancées remarquables dans le domaine de l'apprentissage automatique, notamment avec l'utilisation de modèles d'apprentissage profond. Différents modèles ont été proposés dans plusieurs domaines et ont obtenu des performances exceptionnelles dans de nombreuses applications telles que la reconnaissance de formes ou la reconnaissance vocale. Parmi les différents types d'approches d'apprentissage profond, les réseaux de neurones convolutifs (RNC) ont montré un potentiel exceptionnel pour résoudre divers problèmes, notamment dans les tâches de segmentation d'images médicales. Les RNC sont capables d'extraire des informations complexes et des détails les plus petits de la zone d'intérêt sous étude, ce qui permet d'obtenir les résultats les plus performants dans des tâches les plus compliquées.

Étant donné que les codes de diverses architectures de RNC sont maintenant accessibles pour la recherche, l'objectif des récentes études et pas seulement la localisation ou la segmentation des zones d'intérêt, mais aussi l'obtention des résultats les plus performants avec une précision qui dépasse la compétence des radiologues mêmes. Pour faire cela, le courant de pensée est de combiner les réseaux de neurones convolutifs avec des algorithmes d'optimisation, qui est l'objet principal de ce mémoire.

0.2 Problématique

L'analyse rapide des images médicales et l'identification précise des maladies peut aider de nombreux cas urgents et même sauver des vies. Le temps étant un facteur critique, il est important de pouvoir diagnostiquer et analyser des images médicales le plus rapidement et efficacement possible. Habituellement, les experts analysent directement les différents organes du corps humain en regardant l'image médicale sur un écran ou en visualisant une succession de coupes bidimensionnelles.

Ce diagnostic manuel est une tâche fastidieuse et la décision dans ce cas est purement qualitative et ne dépend que de la connaissance de l'expert. Dans d'autres cas, l'expert doit analyser et identifier les régions d'intérêt en parcourant l'image 3D, tranche par tranche, ce qui nécessite beaucoup de précision, de concentration et d'interaction humaine intensive. Le défi est alors de développer un modèle qui puisse rendre cette tâche facile, tout en préservant la même qualité d'interprétation et d'analyse d'image.

La demande pour l'expertise en imagerie médicale est en croissance, notamment dû au vieillissement et la croissance de la population. De l'autre côté, l'offre est une ressource limitée puisqu'il y a un nombre restreint d'experts qui ne peuvent répondre à toute la demande et qui ont un cap sur leur temps. Alors, il est clair que les ressources actuelles ne suffiront pas à répondre à la demande future dans de délais raisonnables. En d'autre termes, les patients vont attendre de plus en plus longtemps pour recevoir leur analyses de radiologie, ce qui peut faire une différence dans leur état de santé. Vu cela il est évident qu'il faut trouver des solutions innovantes qui pourront résoudre ce problème.

Pour répondre à ce défi, un nombre de méthodes de segmentation d'images médicales ont été développées en fonction des propriétés structurelles et spatiales de l'image, telles que les bords et les régions. Elles ont été appliquées pour identifier les frontières entre les différents organes.

L'intensité des pixels est également l'une des caractéristiques des plus importantes utilisées pour développer des algorithmes de segmentation. Cependant, les frontières entre les différents tissus d'organes et les valeurs d'intensité des pixels sont parfois altérées par des obstacles tels que le bruit de l'image. Ce dernier remplace une partie des pixels de l'image par de nouveaux pixels dont les valeurs d'intensité sont plus proches ou plus éloignées de la plage d'intensité élémentaire raisonnable. Ces variations de pixels donnent à l'image un aspect généralement indésirable qui peut réduire la visibilité de certains éléments de l'image.

Les images médicales telles que l'imagerie par résonance magnétique, la tomodensitométrie par ordinateur et les images radiographiques sont collectées par différents types de capteurs et contaminées par différents types de bruit. Le bruit peut être la cause de divers facteurs liés au système ou à la technique utilisée pour la capture d'images. Le mouvement du patient lors de l'acquisition d'une image peut également affecter la qualité de l'image. Lors du stockage de données, des modifications peuvent aussi affecter la qualité de l'image lors de la compression, de la transmission et de la reproduction des images. Ainsi, développer des algorithmes de segmentation basés sur les caractéristiques, tels que l'intensité et les propriétés structurelles, conduira dans la plupart des cas à des résultats erronés.

La complexité de la structure analysée est aussi l'une des obstacles souvent rencontrée lors de la segmentation d'images médicales. Dans la plupart des cas, le développement d'un algorithme de segmentation se base sur des connaissances spécifiques de la structure cible. Vu cela, l'application d'un tel algorithme, pour segmenter d'autres cibles, ne garantit pas de résultats acceptables. Plusieurs facteurs peuvent complexifier sa structure, comme la forme ou la texture des tissus de l'organe. Aussi, la similarité des tissus de l'organe cible et des structures qui l'entourent rend difficile à identifier les régions d'intérêt. Dans notre situation, les vertèbres et les disques intervertébraux partagent une gamme d'intensité similaire. En outre, la gamme d'intensité des disques chevauche celle des structures adjacentes, telles que les ligaments, la

moelle épinière et les vaisseaux abdominaux. Ceci complique leur interprétation et mène dans la plupart des cas à de fausses segmentations. Tout cela prouve la nécessité de développer un algorithme qui ne se limite pas à de contraintes spécifiques, tels que l'intensité des pixels, et est efficace même dans des cas compliqués.

Dans la littérature, plusieurs algorithmes ont été développés pour répondre au besoin de la segmentation des images. Dans la plupart des études, un expert doit confirmer ou corriger les résultats obtenus, ce qui allonge le temps de traitement des images. Subséquemment, on doit développer un algorithme complètement automatique qui ne requiert aucune intervention humaine.

Pour évaluer et prouver l'efficacité d'une approche de segmentation, il est nécessaire de disposer de références manuelles d'images. Ces références sont limitées, en particulier dans le domaine médical. Les limitations sont dus au fait que la segmentation manuelle est une étape fastidieuse nécessitant du temps et de la concentration de la part des experts. Alors, pour résoudre ce dernier problème, il faut développer un algorithme fonctionnant même avec très peu d'échantillons.

0.3 Objectifs

L'objectif principal de ce mémoire est donc de développer une méthode complètement automatique pour la segmentation de la partie lombaire de la colonne vertébrale à partir d'images IRM. Pour atteindre cet objectif, la présente étude propose une méthode de segmentation entièrement automatique en combinant un réseau de neurones convolutif (RNC) profond et une méthode basée sur la coupe de graphe. Nous allons d'abord développer une architecture de réseau de neurones entièrement convolutif afin d'obtenir des cartes de probabilité des prédictions de chaque image. Au lieu d'utiliser directement les résultats du réseau comme une segmentation finale, nous allons utiliser les cartes de probabilité fournies par notre réseau comme

initialisation pour la méthode de coupe de graphe afin d'affiner la segmentation. Par la suite, nous allons tester et évaluer, quantitativement et qualitativement, les résultats de la segmentation sur trois différentes bases de données contenant les annotations manuelles des différentes structures.

0.4 Contributions de ce mémoire

Pour atteindre cet objectif, nous allons procéder comme suit :

- Motivés par la popularité des techniques d'apprentissage profonds, notamment du réseau de neurones convolutif, nous allons commencer par développer un réseau de neurones entièrement convolutif pour segmenter la partie lombaire de la colonne vertébrale à partir d'images IRM.

- Nous appliquons ensuite la régularisation par coupe de graphe, en prenant les cartes de probabilité obtenues par notre réseau comme initialisation, pour améliorer les résultats de segmentation.

- Dans la littérature, presque toutes les études portent sur la segmentation d'une seule structure à la fois, soit les vertèbres soit les disques. Dans notre étude, nous utilisons une nouvelle base de données combinant les annotations manuelles des deux structures afin de comparer les résultats de la segmentation multi-classes par rapport à la segmentation binaire. Pour améliorer les segmentations multi-classes primaires obtenues par notre RNC, nous utilisons l'algorithme α-expansion.

Plan du mémoire

Après avoir décrit le contexte, les objectifs et les contributions de cette recherche, la suite du présent mémoire est organisée comme suit :

Le chapitre 1 présente la littérature pertinente sur les méthodes de segmentation des vertèbres et des disques intervertébraux.

Le chapitre 2 définit les outils de base permettant de concevoir une architecture du réseau de neurones convolutif. Il présente aussi une revue des architectures les plus populaires utilisées pour la classification et la segmentation sémantique.

Le chapitre 3 présente les principales contributions de ce travail, ainsi que les réseaux de neurones convolutifs et les algorithmes développés pour segmenter les différentes structures de la colonne vertébrale.

Le chapitre 4 décrit les détails d'implémentation de chaque méthode de segmentation selon la structure visée tout en comparant les résultats obtenus avec d'autres travaux qui ont adopté les mêmes bases de données.

La conclusion, présentée dans le chapitre 5, explique les principales contributions de ce travail et suggère des améliorations aux futurs projets.

CHAPITRE 1

ÉTAT DE L'ART

1.1 Algorithmes de segmentation

La segmentation est une opération qui vise à séparer différentes zones homogènes d'une image, afin d'organiser des objets en classes selon différentes propriétés (intensité, texture, couleur, etc.). Les méthodes de segmentation peuvent être regroupées en deux catégories : méthodes non supervisées et méthodes supervisées. La segmentation non supervisée vise à segmenter l'image sans aucune connaissance préalable des classes. Ces méthodes peuvent également être classées en méthodes basées sur les limites ou sur les régions.

Les méthodes de segmentation basées sur les limites visent à détecter les changements brusques de la valeur d'intensité. Elles sont généralement utilisées pour rechercher les discontinuités dans les images en niveaux de gris qui représentent généralement les limites entre les différentes régions. La détection des discontinuités est effectuée à l'aide de différents opérateurs de détection de bord (Sobel, Prewitt, Canny, Laplacian, etc.).

Parmi les méthodes basées sur la région, on peut citer la segmentation par seuillage. C'est la méthode la plus simple et la plus basique. Elle permet la segmentation d'une image en comparant ses pixels un à un avec la valeur d'un seuil prédéfinie. Les pixels dont l'intensité est supérieure au seuil représentent une classe alors que les pixels dont l'intensité est inférieure au seuil représentent une autre classe. Le seuillage est souvent utilisé comme étape initiale dans une séquence d'opérations de traitement d'image. Il ne prend généralement en compte que les valeurs d'intensité et pas les caractéristiques spatiales d'une image. Cela le rend sensible au bruit d'image et non homogène en intensité. Ces deux obstacles majeurs se produisent dans les images à résonance magnétique, ce qui rend plus difficile la séparation des organes qui ont des pixels d'intensité proches.

La segmentation par la croissance de la région (Adams & Bischof (1994); Zhu & Yuille (1996)) est également l'une des méthodes les plus couramment utilisées. Elle sert à diviser l'image en différentes régions en fonction de la croissance des pixels d'initialisation dans l'image originale. Ces pixels peuvent être sélectionnés manuellement, ou automatiquement en fonction de critères de similarité tels que l'intensité des niveaux de gris ou de la couleur. La croissance de ces pixels d'initialisation est par la suite contrôlée par les connectivités entre les pixels. Cette procédure regroupe les pixels de l'image entière en sous-régions ou en régions plus grandes en fonction de critères prédéfinis. Une règle d'arrêt doit être définie pour arrêter le processus quand plus aucun pixel ne répond aux critères d'inclusion dans cette région.

Les techniques de regroupement (*clustering*) sont également des méthodes non supervisées qui ont été utilisées dans la segmentation d'images. Ces méthodes visent à partitionner une image en groupes d'objets. Les méthodes de regroupement, telles que les k-moyens et les c-moyens flous, sont l'une des méthodes populaires en raison de leur simplicité et de l'efficacité de leur calcul.

L'algorithme k-moyens génère une solution de regroupement optimale qui dépend des centres principaux des différents groupes. Une intensité moyenne pour chaque classe est d'abord calculée. L'algorithme fonctionne d'une manière itérative pour minimiser la somme des distances entre les pixels et les moyennes de chaque groupe. La segmentation de l'image est ensuite effectuée en classant chaque pixel dans la classe ayant la moyenne la plus proche.

L'algorithme c-moyens flou est basées sur la théorie des ensembles flous. Le regroupement dans ce cas est plus naturel car, dans la réalité, une division exacte n'est pas possible en raison de la présence de bruit. Les pixels sont partitionnés en groupes en fonction de l'appartenance partielle. Un pixel peut appartenir à plusieurs groupes et la segmentation sera basée sur son degré d'appartenance.

Le choix correct des paramètres initiaux pour les algorithmes de regroupement est très difficile. Une étude approfondie est nécessaire pour identifier les paramètres d'entrée optimaux afin d'obtenir des résultats de segmentation précis.

Récemment, les méthodes de segmentation supervisée utilisant les réseaux de neurones convolutifs sont devenues à la pointe de la technologie, en particulier dans le domaine médical. Ces méthodes sont basées sur l'apprentissage d'un réseau de neurones et nécessitent de grandes quantités de données de références manuelles. Lors de la phase d'apprentissage, la machine apprend automatiquement les représentations et extrait les caractéristiques utiles à partir d'une image brute.

1.2 Les modèles proposés pour la segmentation de la colonne vertébrale

La détection et la segmentation automatiques et semi-automatiques des structures rachidiennes à partir de différentes modalités d'images médicales est une tâche difficile en raison du degré relativement élevé de complexité anatomique de la cible et de la présence de limites peu claires entre vertèbres, disques et organes entourant la colonne. La résolution d'image insuffisante, les variations d'intensité et le bruit affectent également la qualité de l'image, ce qui rend difficile l'extraction de certaines caractéristiques.

Les limites des disques et des vertèbres pourraient être bien définies du point de vue d'un observateur humain dans les images médicales. La localisation, l'étiquetage et la segmentation sont difficiles pour un algorithme entièrement automatisé en raison de l'homogénéité des organes. Des structures adjacentes non pertinentes peuvent être connectées et avoir des intensités similaires. Récentes recherches sur la colonne vertébrale ont mis l'accent sur la localisation et la segmentation de la colonne vertébrale et discale de diverses modalités médicales telles que les images radiographiques, CT et IRM.

Au cours de la dernière décennie, plusieurs méthodes automatisées et semi-automatisées axées sur la segmentation des vertèbres et des disques intervertébraux ont été développées.

1.2.1 Segmentation des vertèbres

La segmentation vertébrale est nécessaire pour un certain nombre d'applications cliniques telles que la détection d'anomalies de la colonne vertébrale ou la chirurgie guidée par l'image.

Une segmentation robuste et automatique des vertèbres s'est révélée être une tâche difficile en raison de la complexité de l'anatomie vertébrale et de la grande variabilité entre les individus. De ce fait, la localisation exacte et la segmentation précise des vertèbres restent un sujet de recherche difficile en imagerie médicale.

Il existe un large éventail d'approches dans la littérature existante pour la localisation et la segmentation des vertèbres. Parmis ces approches, Aslan *et al.* (2010) ont proposé une méthode pour la segmentation des corps vertébraux à partir des images TDM tridimensionnelle. Tout d'abord, un filtre adapté est utilisé pour détecter automatiquement les régions des vertèbres. Cette procédure élimine l'interaction de l'utilisateur. Ensuite, une combinaison linéaire de la méthode gaussienne est utilisée pour approximer la distribution des niveaux de gris des vertèbres et des organes environnants. Cette distribution est utilisée comme segmentation initiale pour l'étape suivante. Enfin, la méthode de coupe de graphe est utilisée comme un algorithme d'optimisation globale pour trouver la segmentation qui minimise une certaine fonction d'énergie. La méthode a été testée sur des ensembles de données du rachis lombaire et thoracique. Elle a obtenu une erreur de segmentation moyenne de 4,7% sur 11 séries d'images cliniques 3D, ce qui correspond à coefficient de similarité moyen de 90,6%.

Ghosh *et al.* (2011) ont proposé une méthode entièrement automatique pour localiser et segmenter les vertèbres et diagnostiquer les anomalies vertébrales. Leur méthode consiste en cinq étapes principales. Ils localisent d'abord les disques intervertébraux en utilisant un modèle probabiliste à deux niveaux qui attribue des étiquettes de haut niveau à tous les disques. Les résultats sont optimisés en utilisant un algorithme de maximisation des attentes généralisées. Ils localisent ensuite le squelette lombaire en utilisant des opérations morphologiques sur une image seuillée qui donne une région rugueuse d'intérêt. L'étape de segmentation est basée sur l'intensité et la géométrie des vertèbres. Les résultats sont ensuite réalisés en utilisant des opérations morphologiques sur chacune des cinq vertèbres lombaires pour obtenir des contours lisses. Les deux dernières étapes consistent en la détection de l'axe de la colonne et le calcul des extrémités vertébrales. Ces deux étapes sont utilisées pour extraire des caractéristiques pour le diagnostic automatique des fractures par compression en coin ainsi que diverses anomalies

vertébrales. Cinq classificateurs issus des méthodes de modélisation existantes sont ensuite utilisés pour identifier les anomalies. L'étude a été testée sur 50 volumes d'images de tomographie assistée par ordinateur anonymisées qui contiennent un total de 250 vertèbres et seulement 15 cas sont utilisés pour le diagnostic. Ces cas contiennent différents types de fractures, la poursuite des os et plusieurs anomalies discales. Pour les résultats de la segmentation, ils ont obtenu une erreur de distance moyenne de Hausdorff de 1,5 mm et une précision de 97,33% pour le diagnostic.

Ayed *et al.* (2012) ont formulé la segmentation dans les images IRM comme un problème de correspondance de distribution. Pour un calcul efficace, ils ont divisé le problème en plusieurs sous-problèmes, où chacun pouvait être résolu par relaxation convexe et une méthode lagrangienne augmentée. Pour commencer, l'utilisateur doit marquer une seule vertèbre avec trois points dans une coupe centrale mi-sagittale. Cette initialisation permet de calculer une distribution de modèles multidimensionnels de caractéristiques qui codent des informations contextuelles sur les vertèbres et leurs structures voisines. La méthode a été évaluée sur 75 vertèbres lombaires acquises à partir de 15 tranches médio-sagittales 2D et a atteint un coefficient de similarité moyen de 85%.

Zukic *et al.* (2012) ont proposé une approche semi-automatique pour la détection et la segmentation des corps vertébraux à partir d'images IRM 3D. L'initialisation par l'utilisateur est nécessaire pour identifier les les corps vertébraux cibles. Leur méthode consiste à combiner les caractéristiques de bord et d'intensité pour estimer les limites de chaque corps vertébral. Chaque vertèbre est ensuite segmentée en utilisant un algorithme d'inflation itératif. L'algorithme commence par un petit maillage de surface triangulaire au centre approximatif du corps vertébral. Ce maillage est agrandi en utilisant l'algorithme de gonflage et orienté vers une géométrie en étoile. La méthode a été évaluée sur dix ensembles de données pathologiques et un ensemble de données provenant d'un patient en bonne santé. Elle a atteint un coefficient de similarité de 78%, ce qui est considéré comme un bon résultat compte tenu de la complexité de l'anatomie des sujets pathologiques.

Ils ont également proposé une extension de leur travail (Zukić *et al.* (2014)). Dans cette étude, ils ont utilisé la méthode Viola-Jones (Viola & Jones (2001)) pour détecter les centres et les tailles des corps vertébraux. Ensuite, les vertèbres sont segmentées en utilisant le même algorithme itératif d'inflation. Enfin, en se basant sur les résultats des étapes précédentes, des caractéristiques diagnostiques géométriques sont déduites afin de diagnostiquer trois maladies. La vérification de l'utilisateur est nécessaire avant la dernière étape pour corriger les résultats de la détection. L'utilisateur doit également choisir une étiquette pour la vertèbre inférieure à partir de laquelle les autres vertèbres sont estimées. Leur méthode a été évaluée sur 22 ensembles de données pathologiques et 4 ensembles de données de volontaires sains. Elle a atteint un coefficient de similarité de 79,3%.

Chu *et al.* (2015a) ont proposé une méthode entièrement automatique pour la localisation et la segmentation des corps vertébraux. Ils ont utilisé une classification aléatoire des forêts pour estimer les centres de chaque corps vertébral. Des régions d'intérêt sont définies autour de chaque centre détecté. La probabilité d'apparition et le pré-réglage spatial de chaque voxel sont ensuite calculés dans ces régions. Les résultats de tous les voxels sont combinés pour obtenir une carte de probabilité postérieure. La segmentation binaire finale de chaque corps vertébral est dérivée en seuillant la carte de probabilité avec une valeur de 0,5 et en ne retenant que la plus grande composante connexe. La méthode a été évaluée sur deux bases de données de modalités différentes. La première base de données est constitué de 23 images IRM 3D sur laquelle ils ont obtenu un coefficient de similarité de 88,7%. La deuxième est constitué de 10 images CT 3D sur laquelle ils ont obtenu un coefficient de similarité de 91%.

Korez *et al.* (2015b) ont présenté une nouvelle approche pour la détection et la segmentation automatisées des vertèbres lombaires et thoraco-lombaires à partir d'images du rachis TDM tridimensionnel. La méthode était basée sur la théorie de l'interpolation qui vise à détecter toute la colonne vertébrale dans l'image 3D. Les résultats de détection des vertèbres sont ensuite utilisés comme initialisation pour une approche améliorée du modèle déformable contraint par la forme. Leur travail a été évalué sur deux bases de données d'images TDM du rachis disponibles en ligne. Le premier ensemble de données contient un total de 50 vertèbres lombaires,

ils ont atteint un taux de détection réussi de 85,1% et un coefficient de similarité de 95,3% pour la segmentation. Le deuxième ensemble contient 170 vertèbres thoraco-lombaires, ils ont atteint 83,2% pour la détection et 94,4% pour la segmentation.

Korez *et al.* (2016) ont proposé une méthode automatisée pour la segmentation supervisée des corps vertébraux à partir d'images IRM 3D. Ils ont utilisé d'abord un réseau de neurones convolutif 3D. L'architecture du réseau proposée est similaire à celle de LeNet. Elle se compose de deux couches de convolution suivies chacune d'une couche de sous-échantillonnage. Les sorties de ces couches sont ensuite transmises via deux couches entièrement connectées. La sortie de la dernière couche est passée à travers une fonction softmax pour générer deux cartes de probabilité représentant les vertèbres et l'arrière-plan. Ces cartes sont ensuite utilisées pour guider un modèle déformable, basé sur la minimisation d'une fonction d'énergie, vers les limites des vertèbres pour obtenir une segmentation précise. La méthode proposée a été appliquée pour segmenter les cinq vertèbres lombaires et les deux dernières vertèbres thoraciques (T11-L5) à partir de 23 sujets différents. La base de données a été divisée en 11 images pour l'apprentissage et 12 images pour le test qui a été répété cinq fois. Ils ont atteint un coefficient de similarité moyen de 93.4%.

Hille *et al.* (2016) ont proposé un ensemble de niveaux hybrides pour la segmentation des corps vertébraux dans les images IRM. L'utilisateur doit d'abord marquer chaque vertèbre avec trois points dans la section sagittale médiane pour approximer la taille et le centre de chaque corps vertébral. Ensuite, une boîte de délimitation cylindrique est construite autour de chaque corps vertébral. A l'intérieur du cylindre, une pré-segmentation seuillée basée sur l'information d'intensité suivie d'un filtrage morphologique est utilisée pour estimer la forme du corps vertébral et pour définir le contour initial. La pré-segmentation est ensuite combinée à des statistiques d'intensité pour formuler une fonction d'ensemble hybride de niveaux. La minimisation de cette fonction encourage les contours actifs à entourer les régions avec une plage de niveaux de gris spécifique par vertèbre au sein de la segmentation primaire. L'étude a été évaluée sur six ensembles de données, contenant 34 vertèbres, de patients sains et pathologiques et a rapporté un coefficient de similarité de 84,8%.

Athertya & Kumar (2016) ont présenté une combinaison robuste de l'algorithme c-moyennes flou pour la segmentation et l'étiquetage des corps vertébraux lombaires à partir des images IRM. La méthode consistait en quatre étapes principales. Tout d'abord, les images sont lissées en utilisant un filtre de diffusion anisotrope préservant les bords pour éliminer l'inhomogénéité. Ensuite, un c-moyennes flou est utilisé pour regrouper les pixels d'images. Après cela, une série d'opérations morphologiques sont effectuées pour extraire les corps vertébraux de la sortie groupée. Les vertèbres segmentées, de L5 à L1, sont finalement étiquetées en utilisant l'entité de composant associée. Leur méthode a été testée sur des coupes sagittales d'images IRM en pondération T1 de 16 patients différents. Les résultats ont été comparés à la segmentation en utilisant le seuillage Otsu et le clustering K-means. Dans certaines images, la segmentation n'a pas permis d'obtenir des résultats de qualité. Le c-moyennes flou ne s'adapte pas à la topologie complexe de la colonne vertébrale en raison de la similarité de l'intensité des composants entourant les vertèbres. Au cours des dernières années, les méthodes d'apprentissage profonds ont abouti aux résultats les plus avancés dans de nombreuses applications telles que la reconnaissance de formes ou la parole. Parmi les différents types d'approches d'apprentissage en profondeur, les réseaux de neurones convolutifs (RNC) ont montré un potentiel exceptionnel pour résoudre divers problèmes tels que la segmentation d'images médicales.

Janssens *et al.* (2017) ont présenté une méthode de segmentation des vertèbres lombaires à partir d'images tomodensitométriques, basée sur deux étapes principales. Premièrement, ils ont utilisé un réseau neuronal convolutif complet qui vise à localiser la région lombaire. Deuxièmement, la région d'intérêt prédite est extraite et transmise à travers un réseau UNet tridimensionnel pour segmenter les vertèbres lombaires. La segmentation se fait en deux étapes. Tout d'abord, le réseau est formé pour la segmentation binaire du rachis lombaire afin de distinguer entre la colonne vertébrale et le fond. Après cela, les poids formés à partir du réseau de segmentation binaire sont utilisés pour initialiser une deuxième segmentation multi-classe où chaque classe correspond à une vertèbre. Des opérations morphologiques ont été menées pour éliminer les trous internes et isoler de petits volumes des résultats finaux. L'approche a été évaluée sur la base de données de la compétition MICCAI-2016 (xVertSeg). Les expériences finales ont été

réalisées en utilisant une validation croisée en prenant chaque fois 12 sujets comme données pour l'apprentissage et 3 sujets comme données de test. Le processus a été répété cinq fois, ce qui produit une similitude de coefficient moyenne de 95,77%.

Vania *et al.* (2017) ont développé une méthode de automatique pour segmenter la colonne vertébrale à partir d'images tomodensitométriques. Leur méthode consiste à utiliser une simple architecture de réseau de neurones à convolution composée de seulement deux couches de convolution et trois couches entièrement connectées. La fonction d'activation ReLU et la mise en pool maximale sont utilisées après chaque convolution. Un abandon de 50% est utilisé dans les couches entièrement connectées. La contribution principale de cette étude est la présentation d'une nouvelle approche pour préparer les données en utilisant une génération redondante d'étiquettes de classe. Quatre classes sont utilisées au total ce qui aide le modèle à distinguer avec précision entre la colonne vertébrale et son environnement. La méthode a été évaluée sur des images tomodensitométriques de 32 patients différents et a rapporté un coefficient de similarité de 94,29%.

Al Arif *et al.* (2018) ont proposé une version modifiée de l'UNet pour la segmentation des 5 vertèbres cervicales (C3-C7). Ils ont introduit un nouveau terme, sensible à la forme, dans la fonction de perte du réseau qui permet de préserver la forme de la vertèbre et d'apprendre à pénaliser les zones prédites à l'extérieur des frontières. Leur travail a été formé sur un ensemble de données augmenté de 26370 vertèbres et testé sur 792 vertèbres collectées à partir d'un total de 296 images radiographiques cervicales latérales et il a atteint un coefficient de similitude de 94.38%.

Lessmann *et al.* (2018) ont proposé une étude pour l'identification et la segmentation de la totalité de la partie thoracique et lombaire de la colonne vertébrale. Leur méthode consiste en deux étapes. Dans la première étape, un U-net tridimensionnel est utilisé itérativement pour localiser grossièrement et identifier chaque vertèbre à partir d'images à basse résolution. La vertèbre inférieure (L5) est identifiée en premier, puis utilisée comme référence pour l'étiquetage de toutes les vertèbres. La deuxième étape consiste à entraîner un autre réseau de la même

architecture en utilisant la sortie générée par le premier réseau comme masques pour segmenter les vertèbres à partir des images de résolution d'origine. Leur étude a été formée et évaluée sur les données publiquement disponibles du défi de la segmentation de la colonne vertébrale MICCAI 2014. L'ensemble de données se compose de 15 images tomodensitométriques de jeunes adultes en bonne santé. 10 images ont été utilisées pour la formation et 5 pour le test. Ils ont atteint un coefficient de similarité moyen de 94,7%. Ils ont testé leur approche sur cinq tomodensitogrammes de sujets malades, mais ils ont admis qu'elle ne donnait pas de bons résultats avec des vertèbres contenant de graves fractures par compression.

1.2.2 Segmentation des disques intervertébraux

Les disques intervertébraux sont des composants de la colonne vertébrale interposés entre chaque paire de vertèbres adjacentes. Ils servent d'amortisseurs et sont cruciaux pour le mouvement vertébral. La dégénérescence progressive des disques du rachis lombaire est l'une des causes les plus fréquentes de la douleur de dos. À cet égard, la localisation et la segmentation précises des disques intervertébraux dans les images médicales sont nécessaires pour réduire l'annotation manuelle et pour aider au diagnostic des pathologies discales. L'imagerie par résonance magnétique est considérée comme la meilleure modalité pour la visualisation des disques intervertébraux en raison de son excellent contraste des tissus mous. Ces dernières années, plusieurs méthodes automatiques et semi-automatiques de localisation et de segmentation des disques intervertébraux ont été développées.

Michopoulou *et al.* (2009) ont proposé une méthode basée sur un atlas probabiliste couplée à des classificateurs basés sur l'intensité pour la segmentation des disques intervertébraux lombaires normaux et dégénérés. Ils ont utilisé une méthode semi-automatique qui nécessitait la sélection des points de disques le plus à gauche et le plus à droite dans l'image par l'utilisateur qui vont servir comme points de repère pour un enregistrement rigide. Cette étude a utilisé un total de 170 disques intervertébraux lombaires à partir d'images 2D IRM médiosagittaux pondérées en T2 de 34 patients différents. 92 disques ont été considérés comme normaux et 78 comme dégénérés. 50 disques normaux aléatoires ont été utilisés pour la conception de l'atlas

probabiliste, tandis que les 42 disques normaux et 78 disques dégénérés restants ont été utilisés pour évaluer la méthode de segmentation. Trois méthodes ont été évaluées. La meilleure performance globale a été obtenue grâce à l'approche atlas-robuste-c-moyennes floue, compte tenu de la précision de la segmentation et de l'efficacité temporelle. cette méthode a atteint un coefficient de similarité de 91,6% pour les disques normaux et de 87,2% pour les disques dégénérés.

Ayed *et al.* (2011) ont proposé une méthode de segmentation par la coupe de graphe pour la délimitation des disques intervertébraux à partir d'images IRM médiosagittaux pondérées en T2 du rachis lombaire. Leur algorithme optimise une fonction de coût basée sur des mesures de similarité globales et des connaissances antérieures acquises sur l'interaction géométrique entre différentes structures dans l'image. L'utilisateur doit sélectionner le disque T12-T11 pour chaque sujet par une ellipse. Un seul sujet est requis pour la formation. L'évaluation a été réalisée sur un total de 60 disques lombaires acquis à partir de 10 images IRM sagittales pondérées en T2.

Law *et al.* (2013) ont proposé une étude pour la segmentation des disques intervertébraux non supervisé basé sur des scans de résonance magnétique du rachis sagittal moyen. Cette étude commence avec la détection de la région du corps vertébral qui sert a fournir les positions et les orientations des limites supérieure et inférieure des vertèbres. Sur la base de ces informations positionnelles et directionnelles des corps vertébraux adjacents, les positions, les tailles et les orientations des disques seront estimées. Ils ont utilisé un total de 69 séquences sagittales d'images IRM, provenant de 22 sujets différents, dans lesquelles il y a 455 disques à évaluer. Pour chaque séquence, deux positions ont été sélectionnées manuellement pour indiquer les vertèbres sous le premier et le dernier disques cibles. Ils ont atteint une précision de segmentation de 92,04%.

Castro-Mateos *et al.* (2014) ont proposé une méthode quasi-automatique pour la segmentation et la classification des disques intervertébraux à partir d'images IRM 2D. Une intervention manuelle, par sélection du centre du disque, permet d'initialiser un contour elliptique. La seg-

mentation initiale est d'abord réalisée en utilisant des modèles de contour actifs qui visent à contrôler la fuite de la bordure du disque vers les bords voisins en minimisant une fonction d'énergie. La segmentation est encore améliorée en utilisant des c-moyennes flous. Un classificateur Adaboost est finalement utilisé pour fournir les degrés Pfirrmann de dégénérescence des disques. La méthode a été évaluée sur un ensemble de 150 disques, sains et dégénérés. Elle a atteint un coefficient de similarité de 91,7% pour la segmentation et une spécificité de 93% pour la classification.

Dong & Zheng (2015) ont proposé une méthode pour la segmentation des disques intervertébraux lombaires basée sur des modèles graphiques. Partant de deux repères fournis par l'utilisateur pour indiquer les centres des corps vertébraux L1 et L5 sur une image sagittale médiane, les corps vertébraux et les disques lombaires sont identifiés en utilisant une stratégie basée sur un modèle graphique. Sur la base des résultats d'identification, des modèles 3D des vertèbres et de la colonne lombaire sont ensuite construits. En éliminant le modèle de vertèbres du modèle de la colonne, la région candidate pour chaque disque cible peut être localisée. La segmentation est ensuite effectuée sur les régions des disques extraites en utilisant un enregistrement difféomorphe multi-noyau entre un modèle de disque et les données observées. La méthode a été évaluée sur 15 images MR 3D, chacune contenant cinq disques lombaires, et elle a atteint un coefficient de similarité moyen de 85,12%.

Wang & Forsberg (2015) ont proposé une approche pour la localisation et la segmentation des disques intervertébraux dans les images IRM. Tout d'abord, des fonctions de canal intégrées couplées à un classificateur AdaBoost sont utilisées pour détecter et étiqueter les vertèbres à partir d'images médiosagittaux. La sortie de cette étape fournit un ensemble de points de repère indiquant les centres des vertèbres de T11 au sacrum. Ces points sont utilisés pour estimer les centres des disques cibles. La deuxième étape consiste en un enregistrement d'image, dans lequel un ensemble de volumes d'images avec des atlas de disques intervertébraux correspondants est enregistré en utilisant la sortie de la première étape en tant qu'initialisation. Dans la dernière étape, les atlas enregistrés sont combinés en utilisant la fusion d'étiquettes pour dériver la segmentation finale. La méthode a été évaluée sur l'ensemble de données du défi de

segmentation MICCAI 2015 qui consiste en 15 images IRM 3D fournies pour l'apprentissage et 10 sujets pour le test. Elle a atteint un coefficient de similarité moyen de 90%.

Zhu *et al.* (2016) ont proposé une méthode non supervisée pour la localisation et la segmentation des disques intervertébraux dans les images IRM basée sur l'utilisation d'un ensemble de filtres de Gabor (Daugman (1985)). Premièrement, l'ensemble de filtres est utilisé pour extraire les caractéristiques structurelles des disques intervertébraux. Deuxièmement, les caractéristiques de Gabor de la colonne vertébrale sont calculées et les courbes de la colonne vertébrale sont détectées. Troisièmement, les images des disques de Gabor sont calculées et ajustées en fonction des courbes de la colonne vertébrale. Quatrièmement, les disques sont localisés par une analyse de regroupement basée sur les images de caractéristiques Gabor traitées. La segmentation des disques est finalement réalisée sur la base des caractéristiques des filtres de Gabor, des résultats de localisation et d'un seuil auto-adaptatif amélioré. Des opérations morphologiques sont utilisées pour résoudre certains problèmes de segmentation comme les trous, les portions superflues et les zones non-lisses. Les régions petites et supplémentaires sont également enlevées du résultat de la segmentation et seules les plus grandes régions sont retenues. La méthode proposée est évaluée sur une base de données d'images IRM composée de 278 disques intervertébraux de 37 patients et atteint un coefficient de similarité de 92,37% pour la segmentation.

Le développement plus récent sur les réseaux de neurones profonds, et en particulier sur les réseaux de neurones convolutifs (RNC), a permis d'obtenir les méthodes les plus avancées pour résoudre le problème de la localisation et la segmentation des disques intervertébraux.

Chen *et al.* (2016) ont proposé un modèle de réseau neuronal à convolution complète 3D pour la localisation et la segmentation de disques intervertébraux à partir d'images IRM. Le point fort de leur méthode a été l'extension du réseau 2D en 3D en utilisant des filtres de convolution 3D. Le réseau proposé se compose de deux couches de convolution suivies chacune d'une couche de mutualisation maximale et de deux couches d'échantillonnage ascendant à la fin. Le réseau prend le volume entier en entrée et génère un masque de segmentation 3D de la même

taille que l'entrée. Les étapes de post-traitement sont utilisées pour générer des résultats de segmentation lisses locaux à l'aide de petits filtres. Les petites zones sont ensuite retirées et seuls les composants les plus grands sont conservés. La méthode a été évaluée sur l'ensemble de données du défi MICCAI 2015 qui comprend 25 images IRM 3D. Pour démontrer que l'exploration d'informations 3D flexibles permet d'obtenir des performances plus prometteuses, les tests dans ce travail ont été réalisés à l'aide de réseaux de convolution 2D et 3D. Le réseau 3D a atteint un coefficient de similarité moyen de 88,7% qui est meilleur que 82,7% atteint par le réseau 2D.

Ji *et al.* (2016a) ont proposé d'utiliser les réseaux de neurones convolutifs profonds pour la segmentation des disque intervertébraux. Ils ont effectué d'abord un pré-traitement pour définir la région d'intérêt dans les images. Ceci est fait pour chaque coupe sagittale dans le volume 3D par un seuillage prédéfini suivi par des opérations de fermeture pour éliminer les régions qui ne couvrent aucun tissu humain. Ils ont ensuite utilisé un échantillonnage de patch spécifique et construit deux types de patchs différents. Les données préparées sont ensuite utilisées comme entrées pour un réseau de neurones convolutionnels. L'architecture du réseau proposée est composée de 3 couches de convolution suivies chacune d'une couche de max-pooling, et de deux couches entièrement connectées à la fin. La dernière couche contient deux neurones qui représentent les deux classes (disque et arrière-plan). Le vecteur résultant est passé à travers une fonction softmax pour générer une distribution sur les 2 étiquettes de classe. Enfin, les cartes de probabilité générées par le RNC sont lissées en utilisant un filtre moyen et binarisées en utilisant un seuillage de 0,5 pour obtenir les masques de segmentation. Seuls les sept plus gros composants connectés à partir du bas de l'image de test sont conservés et toutes les zones contenant moins de 512 voxels sont supprimées. Evaluées sur les jeux de données de la segmentation MICCAI 2015, leur méthode a atteint un coefficient de similarité moyen de 89,2% et une distance moyenne de surface moyenne de 1,3 mm.

Ji *et al.* (2016b) ont proposé un réseau entièrement automatique pour localiser et segmenter les disques intervertébraux à partir d'IRM tridimensionnelles multimodales. Dans cette étude, un régresseur forestier aléatoire a été utilisé pour localiser grossièrement le disque interver-

tébral S1-L5. Ce centre est ensuite utilisé comme origine pour calculer la relation spatiale relative entre les sept disques intervertébraux. Un modèle de forme est ensuite construit pour localiser les sept disques intervertébraux en faisant simplement la moyenne des coordonnées de chaque centre des disques intervertébraux de l'ensemble de données d'apprentissage. Enfin, les disques intervertébraux sont segmentés séquentiellement en formant un classificateur de réseau neuronal convolutionnel pour chaque disque intervertébral. L'architecture du réseau proposé est simple. Elle se compose de deux couches de convolution suivies chacune d'une couche de pool maximum et de deux couches entièrement connectées à la fin du réseau. Les cartes de probabilité sont finalement obtenues en utilisant une fonction softmax pour deux classes. Quatre canaux sont utilisés pour représenter les images multimodales dans la couche d'entrée. Plusieurs régions peuvent être classées comme disque intervertébral, seule la région la plus proche du centre du modèle de forme est considérée comme disque. Leur méthode a été évaluée sur les données du challenge MICCAI 2016 qui consiste en 8 séries d'images 3D multimodales. Chaque ensemble d'images comprend quatre images 3D alignées de différentes modalités. Ils ont atteint une distance moyenne de 0,64 mm pour la locatlisation et un coefficient de similarité moyen de 90,8% pour la segmentation. La forme du modèle dans cette étude a été obtenue sur la base de cas sains, ce qui n'est pas le cas chez les sujets présentant des déformations de la colonne vertébrale. Les résultats de la segmentation ont également été comparés à la segmentation en utilisant uniquement des images monomodes et les résultats obtenus étaient assez similaires.

Korez *et al.* (2017) combinent les développements dans l'apprentissage automatique et l'analyse d'images, et proposent une nouvelle méthode entièrement automatisée en couplant des modèles déformables avec des réseaux convolutifs pour la segmentation de disques intervertébraux à partir d'images IRM 3D de rachis. Leur méthode consiste en trois étapes principales. Une approche basée sur les points de repère est d'abord utilisée pour identifier chaque disque avec 5 points de repère en considérant l'apparence de l'intensité globale et l'information de forme locale des disques. Selon les emplacements de points de repère optimaux obtenus, un modèle de forme moyenne est utilisé pour déterminer des régions d'intérêt pour chaque disque.

Ensuite, un simple réseau neuronal convolutif 3D à six couches est utilisé pour générer des cartes de probabilité des disques intervertébraux à partir des régions d'intérêt déterminées. L'architecture proposée est similaire à l'architecture Lenet. Il est composé de deux couches de convolution suivies chacune d'une couche de max-pooling, et de deux couches entièrement connectées à la fin du réseau. Les cartes de probabilité sont finalement obtenues en utilisant une fonction softmax pour deux classes. Une segmentation basée sur un modèle déformable est ensuite résolue en tant que minimisation d'énergie sur la base des résultats de la carte de probabilité spatiale obtenue du réseau de neurones. La méthode a été évaluée sur une base de données accessible au public utilisée dans le cadre du défi MICCAI-2015, qui a abouti à un coefficient de similarité moyen global de 92,8%.

L'utilisation d'images multi-modalités dans les réseaux de neurones convolutifs a montré un potentiel exceptionnel pour résoudre divers problèmes de segmentation d'images médicales. Ces images peuvent fournir des informations complémentaires pour identifier plus de pathologies, de tissus et d'anatomies.

Liu & Zhao (2018) ont proposé une méthode automatique basée sur un réseau multi-échelles entièrement convolutif pour la segmentation et la localisation de disques intervertébraux sur des images IRM 3D. Leur réseau proposé est une architecture UNet modifiée. Ils ont ajouté des connexions résiduelles entre les cartes de caractéristiques de même niveau et ils ont inséré des modules (SE) dans les connexions de saut entre la partie encodeur et la partie décodeur de l'architecture. L'entrée de leur réseau est une image 2.5D composée de quatre images IRM acquises avec quatre différentes modalités (en phase, en opposition de phase, d'eau, de graisse), tandis que la sortie du réseau est une image 2D. Leur méthode a obtenu les meilleures performances sur la base de données de test du défi MICCAI-2018 pour la localisation et la segmentation.

Dolz *et al.* (2018b) ont proposé aussi une architecture UNet modifiée pour la segmentation des disques intervertébraux à partir d'images multi-modalités. La partie encodeur de leur réseau est composée de quatre chemins, chacun traitant une modalité d'image différente. Plus que les

connexions de saut entre l'encodeur et le décodeur, des connectivités denses sont utilisées entre les différentes couches et les différents chemins de modalité. Chaque couche de convolution est un module *Inception* (Szegedy *et al.* (2016)) modifié qui comprend deux blocs de convolution dilatés supplémentaires. Leur méthode a été évaluée sur une base de données de 16 données IRM multimodales 3D, et a atteint un coefficient de similarité moyen de 91,91%.

Par la suite, nous analyserons quelques recherches qui ont été présentées à la *Conférence Internationale sur l'Informatique en Images Médicales et l'Intervention Assistée par Ordinateur MICCAI-2015* sur la localisation et la segmentation des disques intervertébraux. La base de données utilisée dans ce concours est composée de 25 images IRM 3D. Seulement 15 images avec leurs annotations manuelles respectives sont fournies en tant que base de données d'apprentissage alors que les 10 sujets restants sont conservés par les organisateurs du concours pour une évaluation juste et indépendante. Aux fins de précision, et pour tester et évaluer notre approche, nous allons nous baser sur la même base de données d'apprentissage que celle utilisée dans ce défi. Zheng *et al.* (2017) ont présenté une évaluation et une comparaison des résultats obtenus par les équipes participant à la compétition. Seulement, neuf équipes ont publié leur résultats sur les bases de données du test. En plus, six équipes ont évalué leurs méthodes sur la base de données d'apprentissage. Ils ont suivi la même validation croisée en prenant à chaque fois 14 sujets pour l'apprentissage et un sujet pour le test.

Dans l'étude présentée par (Wang & Forsberg (2015)), les corps vertébraux sont d'abord détectés et étiquetés à l'aide de caractéristiques de canal intégral couplées à un classificateur AdaBoost. Ensuite, sur la base des résultats de la détection, un ensemble de volumes d'images avec des atlas de disques intervertébraux correspondants est enregistré dans le volume cible. Enfin, les atlas enregistrés sont combinés en utilisant la fusion d'étiquettes pour obtenir la segmentation finale.

Chu *et al.* (2015c) ont utilisé une méthode de régression basée sur les données pour localiser les disques. Sur la base des sorties de localisation, ils ont ensuite défini les régions d'intérêt pour l'étape de segmentation. La probabilité de pixels est estimée dans la région d'intérêt en tant que

objet ou arrière-plan. Elle est ensuite combinée avec une probabilité antérieure, qui est tirée d'un ensemble de données d'apprentissage, pour obtenir la probabilité postérieure du pixel. La segmentation binaire des disques est ensuite dérivée par un seuillage sur les probabilités estimées.

Hutt *et al.* (2015) ont proposé une méthode de segmentation basée sur un champ aléatoire conditionnel (CRF) opérant sur des supervoxels (groupes de voxels similaires). Ils ont d'abord utilisé un clustering itératif linéaire simple pour générer des supervoxels 3D. Une approche d'apprentissage non supervisée des fonctionnalités est ensuite développée pour caractériser les régions des supervoxels. Pour estimer les étiquettes de classe des supervoxels, les entités sont utilisées pour former un SVM avec un noyau de fonction à base radiale (RBF) généralisé. Enfin, les prédictions du classificateur sont incorporées dans les fonctions potentielles d'un CRF avec un apprentissage métrique entre les supervoxels, ce qui permet une segmentation efficace en utilisant les coupes de graphes.

Urschler *et al.* (2015) ont présenté une méthode d'apprentissage automatique pour prédire les centres des corps vertébraux et des disques. Après la prédiction des points de repère, les corps vertébraux sont segmentés uniquement sur la base des informations relatives au gradient de l'image. Ensuite ils sont fusionnées en paires de corps vertébraux adjacents à des objets uniques pour initialiser la segmentation des disques. Enfin, la segmentation des disques est formulée comme une tâche d'optimisation de contour active géodésique convexe basée sur des arêtes ressemblant géométriquement aux cibles.

Dans l'étude proposée par Korez *et al.* (2015a), la détection des disques intervertébraux est effectuée par une approche basée sur des points de repère. Les caractéristiques pseudo-Haar sont ensuite utilisées pour décrire les informations d'apparence de chaque point de repère. Les relations spatiales entre les paires de repères de chaque disque sont utilisées pour décrire les informations de forme. Les positions optimales des repères correspondent au meilleur accord entre ces deux informations. Sur la base des résultats de la détection, la segmentation est effec-

tuée par une approche basée sur un modèle déformable utilisant le descripteur de contexte de similarité propre.

La méthode développée par Neubert *et al.* (2015) était une extension de leur étude antérieure (Neubert *et al.* (2012)) en utilisant le modèle de forme active. Tout d'abord, des atlas 2D d'apprentissage ont été extraits à partir de 35 images sagittales. Ces atlas sont ensuite utilisés pour créer un atlas moyen de la région L5-S1-S2. Un gabarit 3D déformable est utilisé pour la segmentation approximative des corps vertébraux S1 et L5. Enfin, les atlas et les gabarits déformables sont utilisés pour initialiser automatiquement la segmentation des disques intervertébraux à l'aide des modèles de forme active. Le processus s'est poursuivi de manière itérative jusqu'à ce que tous les disques sont segmentés.

1.2.3 Segmentation multi-classes (vertèbres et disques)

Dans la littérature, il existe un vaste choix de méthodes de segmentation des vertèbres et des disques intervertébraux, mais il y a un manque d'approches qui traitent les deux structures.

Neubert *et al.* (2012) ont proposé une méthode pour la détection et la segmentation des vertèbres et des disques de la région thoraco-lombaire. Les emplacements des vertèbres individuelles sont d'abord trouvés en utilisant une méthode de contour actif sous la forme de rectangles actifs interactifs. La segmentation est ensuite réalisée en utilisant l'analyse de forme statistique et l'enregistrement des profils d'intensité de niveau de gris. La méthode a été validée sur un sous-ensemble de 14 sujets (134 disques et 132 vertèbres) d'images MR 3D. Elle a obtenu un score de similarité de 89% pour la segmentation des disques et de 91% pour les vertèbres.

Ghosh & Chaudhary (2014) ont proposé une méthode supervisée pour la détection des disques intervertébraux et la segmentation des tissus dans les images IRM 2D lombaires cliniques. Ils localisent d'abord les disques inférieur et supérieur en calculant l'histogramme des caractéristiques des gradients orientés pour toutes les images d'apprentissage et modélisent des SVM binaires, où les deux classes sont les disques et l'arrière-plan. Les sorties sont des boîtes de

délimitation étroites pour chaque disque. Pour la tâche de segmentation, ils ont utilisé les informations de voisinage de chaque pixel dans un modèle de contexte automatique, où chaque pixel est simultanément marqué comme l'une des quatre étiquettes (vertèbre, disque intervertébral, sac dural ou arrière-plan). Les classificateurs forestiers aléatoires sont formés pour calculer les cartes de probabilité des pixels voisins en tant qu'information contextuelle à chaque itération. Leur méthode a été testée sur un ensemble de données cliniques comprenant 212 images MR mi-sagittales et a atteint une précision de localisation des disques de 98%. Pour la segmentation, leur méthode a atteint un coefficient de similarité de 87% pour le sac dural et de 84% pour les disques intervertébraux. Seuls des résultats qualitatifs ont été présentés pour la segmentation des vertèbres.

Les développements récents dans les réseaux de neurones convolutifs ont encouragé les chercheurs à tirer parti de l'efficacité de ces méthodes, notamment pour la segmentation des différentes structures de la colonne vertébrale. Moran *et al.* (2018) ont proposé une approche en deux étapes pour segmenter les vertèbres et les disques de la colonne vertébrale à partir d'images IRM 2D. Lors de la première étape, ils ont utilisé un réseau de neurones convolutif à 14 couches pour identifier les pixels appartenant à la colonne vertébrale. Lors de la deuxième étape, ils ont utilisé un autre réseau à 14 couches, qui prend comme entrée des images de taille (31×31) contenant au centre les pixels identifiés lors de la première étape. Ce réseau effectue une segmentation multi-classe simultanée et génère un masque à trois étiquettes qui classe les pixels en tant que corps vertébral, disque ou arrière-plan. Pour l'apprentissage des réseaux, ils ont utilisé 200 images IRM sagittales provenant de 42 patients, puis ils ont évalué leur approche sur 30 images IRM sagittales provenant de 20 patients. Ils ont atteint un coefficient de similarité de 82,8% pour la segmentation des vertèbres et de 78,6% pour la segmentation des disques.

Whitehead *et al.* (2018) ont développé un réseau multi-échelles pour la segmentation de vertèbres et des disques intervertébraux à partir d'images IRM 2D. L'architecture proposée est composée de quatre réseaux en cascade. Chaque réseau est composé de six blocs d'*Inception* et d'une couche de convolution finale avec des filtres de taille 1×1. Un bloc d'*Inception* est

constitué de deux convolutions en parallèle avec des filtres de taille 1×1 et 3×3. Les trois premiers réseaux commencent chacun avec une couche de sous-échantillonage afin de redimensionner l'image originale et de la traiter en trois échelles différentes. Les résultats du réseau de la plus petite échelle sont concaténés avec l'image originale et transmis au prochain réseau de la cascade. Pour mettre en œuvre et évaluer leur approche, les auteurs ont utilisé la même base de données de l'étude de Moran *et al.* (2018). Ils ont atteint un coefficient de similarité de 86,5% pour la segmentation des vertèbres et de 83,2% pour la segmentation des disques, qui sont plus précis que les résultats de l'étude mentionnée précédemment.

Tableau 1.1 Résumé des méthodes de pointe pour la localisation et la segmentation des vertèbres et des disques intervertébraux.

Cible	Auteurs	Dimensionnalité	Modalité	Catégorie	Mode	DICE	Base de données
Vertèbres							
	Aslan *et al.* (2010)	3D	TDM	coupe de graphe	automatique	90.6%	
	Ghosh *et al.* (2011)	3D	TDM	seuillage et opérations morphologiques	automatique	97.3%	250 vertèbres lombaires
	Ayed *et al.* (2012)	2D	IRM	coupe de graphe	semi-automatique	85%	15 images mid-sagittales
	Zukic *et al.* (2012)	2D	IRM	méthode basée sur l'inflation	semi-automatique	78%	92 vertèbres
	Zukić *et al.* (2014)	2D	IRM	méthode basée sur l'inflation	semi-automatique	79.3%	234 vertèbres
	Chu *et al.* (2015a)	3D	IRM \ TDM	apprentissage	automatique	88.7% \ 91%	161 vertèbres \ 50 vertèbres
	Korez *et al.* (2015b)	3D	TDM	modèle déformable à contrainte de forme	automatique	94.6%	SpineWeb
	Korez *et al.* (2016)	3D	IRM	RNC & modèles déformables	automatique	93.4%	161 vertèbres
	Hille *et al.* (2016)	3D	IRM	ensemble de niveaux hybrides	semi-automatique	84.8%	34 vertèbres
	Athertya & Kumar (2016)	2D	IRM	c-moyennes flou	semi-automatique	86.72%	16 images sagittales
	Janssens *et al.* (2017)	3D	TDM	RNC (UNet)	automatique	95.77%	MICCAI 2016 xVertSeg
	Vania *et al.* (2017)	2D	TDM	RNC	automatique	94.29%	32 sujets
	Al Arif *et al.* (2018)	2D	Rayon X	RNC (UNet)	automatique	94.38%	792 vertèbres
	Lessmann *et al.* (2018)	3D	TDM	RNC (UNet)	automatique	94.7%	MICCAI 2014
Disques							
	Michopoulou *et al.* (2009)	2D	IRM	Atlas	semi-automatique	89.4%	120 disques
	Ayed *et al.* (2011)	2D	IRM	coupe de graphe	semi-automatique	88%	60 disques
	Law *et al.* (2013)	2D	IRM	flux orienté anisotrope	semi-automatique	92.04%	455 disques
	Castro-Mateos *et al.* (2014)	2D	IRM	Modèles à contour actif & c-moyennes flou	semi-automatique	91.7%	150 disques
	Dong & Zheng (2015)	3D	IRM	modèles graphiques	semi-automatique	85.12%	75 disques
	Wang & Forsberg (2015)	3D	IRM	Atlas	automatique	90%	MICCAI 2015
	Zhu *et al.* (2016)	2D	IRM	filtres de Gabor	automatique	92.37%	278 disques
	Chen *et al.* (2016)	2D \ 3D	IRM	Réseau de neurones entièrement convolutif	automatique	82.7% \ 88.7%	MICCAI 2015
	Ji *et al.* (2016a)	3D	IRM	RNC	automatique	89.2%	MICCAI 2015
	Ji *et al.* (2016b)	3D	IRM	forêt de régression & RNC	automatique	90.8%	MICCAI 2016
	Korez *et al.* (2017)	3D	IRM	RNC & modèles déformables	automatique	92.8%	MICCAI 2015
	Liu & Zhao (2018)	2.5D	IRM	RNC	automatique	90.64%	MICCAI 2018 (IVDM3Seg)
	Dolz *et al.* (2018b)	2.5D	IRM	RNC	automatique	91.91%	16 sujets
Vertèbres & Disques							
	Neubert *et al.* (2012)	3D	IRM	modèles de forme active	automatique	91% \ 89%	132 vertèbres \ 134 disques
	Ghosh & Chaudhary (2014)	2D	IRM	contexte automatique	automatique	—-	212 sujets
	Moran *et al.* (2018)	2D	IRM	RNC	automatique	82.8% \ 78.6%	30 images
	Whitehead *et al.* (2018)	2D	IRM	RNC	automatique	86.5% \ 83.2%	30 images

CHAPITRE 2

RÉSEAUX DE NEURONES CONVOLUTIFS

2.1 Introduction

Les réseaux de neurones ont prouvé au fil des années leur efficacité dans de nombreuses tâches telles que la localisation, la classification, la détection et la segmentation d'objets. Dans ce chapitre, nous allons commencer par introduire le réseau de neurones convolutif et décrire les outils et les opérations de base utilisés pour construire une architecture d'un tel réseau. Ensuite, nous présenterons une revue de la littérature des réseaux de classification les plus populaires et expliquerons certains des composants les plus importants qui ont amélioré la précision de ces modèles au cours des dernières années. Nous terminerons par une revue littéraire des réseaux les plus connus utilisés pour les tâches de segmentation sémantique en mettant en évidence les clés de réussite de chaque modèle.

2.2 Réseaux de neurones convolutionnels

Ces dernières années, l'apprentissage machine a permis d'obtenir des résultats plus proches ou meilleurs que les humains dans de nombreux domaines tels que la reconnaissance vocale (Abdel-Hamid *et al.* (2014); Hannun *et al.* (2014); Graves & Jaitly (2014)), la reconnaissance faciale (Lawrence *et al.* (1997); Parkhi *et al.* (2015)), la classification d'images (Krizhevsky *et al.* (2012); He *et al.* (2015); Hu *et al.* (2017)), la détection d'objets (Girshick *et al.* (2014); Ren *et al.* (2015); Redmon *et al.* (2016)), etc.

L'apprentissage automatique consiste à utiliser un ordinateur pour apprendre à gérer et à résoudre des problèmes sans programmation. Le processus est de construire un modèle basé sur des entrées connus, puis utiliser ce modèle pour prédire sur des nouvelles données. L'idée est de permettre à la machine de créer un programme qui prend une entrée connue et de prédire une sortie intelligente similaire à une sortie connue d'une entrée différente.

Les réseaux de neurones artificiels sont des modèles d'apprentissage statistique, inspirés des réseaux neuronaux biologiques du système nerveux humain (figure). Ils sont utilisés dans l'apprentissage automatique et la reconnaissance de formes. Un neurone biologique a un ensemble de dendrites d'entrée qui reçoivent des signaux électriques. Lorsque le signal passe un seuil, le neurone est activé et le signal est transmis à l'axone connecté à d'autres neurones par des dendrites. Le fonctionnement d'un seul perceptron est limité. Un simple perceptron ne peut classer que des données qui pourraient être séparées par un hyperplan. Le problème, dans la plupart des cas, est qu'il n'est pas possible de séparer linéairement les données d'entrée. Même un simple XOR ne peut pas être adressé par un perceptron.

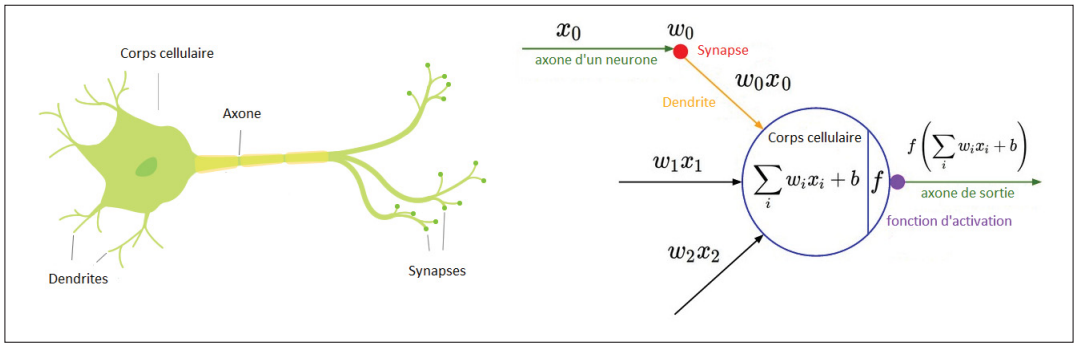

Figure 2.1 Le neurone biologique et son modèle mathématique, adaptée de (Andrej Karpathy, 2018).

Dans les années 1980, le problème a été résolu avec l'introduction du perceptron multicouches. Dans les réseaux de neurones artificiels, en tant que vrais neurones humains, les perceptrons sont reliés les uns aux autres et reproduisent un comportement similaire. Ces réseaux sont représentés comme des systèmes neuronaux interconnectés qui envoient des messages les uns aux autres. Les modèles qu'ils reconnaissent sont numériques, contenus dans des vecteurs et des matrices, dans lesquels toutes les données du monde réel, comme les images, les sons, les textes ou les vidéos, doivent être traduites en données caractéristiques. Les connexions au sein du réseau peuvent être systématiquement ajustées en fonction des entrées et des sorties, ce qui les rend idéales pour l'apprentissage supervisé. Les réseaux de neurones sont utilisés pour

regrouper et classer les données non étiquetées en fonction de similarités entre les nouvelles données et les échantillons déjà étudiés.

La combinaison des perceptrons permet d'approcher n'importe quelle fonction non linéaire en utilisant plus ou moins de couches et de neurones par couche. Cela permet de travailler sur des données plus complexes qui doivent être séparées par des hypersurfaces. Dans un réseau de neurones, les perceptrons sont organisés en couches. Il existe trois types de couches : couche d'entrée, couches cachées et couche de sortie. Les réseaux de neurones peuvent être classés en fonction de leur nombre de couches cachées et de leur connexion. Les réseaux neuronaux comportant plus de deux couches cachées peuvent être considérés comme un réseau neuronal profond. L'avantage d'utiliser des réseaux neuronaux plus profonds est que des modèles plus complexes peuvent être reconnus. Ces réseaux nécessitent plus de données pour éviter les sur-ajustements au cours d'apprentissage.

Les neurones lisent une entrée, la traitent et génèrent une sortie. Les neurones entre deux couches adjacentes sont entièrement connectés. Chaque connexion a un poids qui contrôle le signal entre les deux neurones. Chaque neurone artificiel calcule la somme des produits entre les poids et les entrées qui lui sont venues, puis ajoute un biais. Le résultat est ensuite passé à travers une fonction d'activation qui va ajouter de la non-linéarité (figure 2.1).

Un réseau de neurones est un système adaptatif, il peut changer sa structure interne en fonction de l'information qui lui est transmise en ajustant ses poids. Un ensemble de données appelé ensemble d'apprentissage doit être utilisé. Ces données sont formellement définies comme un ensemble de paires : entrées et cibles. L'objectif de l'apprentissage est d'optimiser les poids afin que le réseau neuronal puisse apprendre à mapper correctement les entrées inconnues aux sorties. Si le réseau génère une bonne sortie, il n'est pas nécessaire d'ajuster les poids. Cependant, si le réseau génère une mauvaise sortie, le système ajuste et modifie les poids pour améliorer les résultats ultérieurs.

Pour l'ensemble d'apprentissage, le résultat théorique est connu et l'optimisation consiste à minimiser l'erreur de prédiction. Cette erreur est la somme des carrés des différences entre les

sorties calculées et les sorties attendues. Il est donc nécessaire d'optimiser numériquement la fonction d'erreur pour trouver la fonction qui donne la meilleure approximation de l'entrée par rapport à la cible. La rétropropagation est un tel algorithme qui effectue une minimisation d'erreur en utilisant la descente de gradient.

L'utilisation de réseaux de couches entièrement connectés pour la segmentation d'image n'est pas une bonne idée. La raison en est qu'une telle architecture de réseau traite les pixels de l'image et ne prend pas en compte la structure spatiale des images. Pour surmonter ce problème, les couches de convolution ont été incluses dans les réseaux de neurones réguliers. Ces couches considèrent le contexte et les informations spatiales des pixels voisins, ce qui conduit à apprendre plus d'informations à partir de l'entrée. Les réseaux de neurones convolutionnels ont une architecture différente de celle des réseaux de neurones réguliers. Généralement, les réseaux de neurones convolutionnels sont composés de trois types principaux de couches ; les couches convolutives, les couches de sous-échantillonnage et les couches entièrement connectées. La configuration et le nombre de ces couches dans l'architecture du réseau dépendent du type et de la complexité du problème.

2.2.1 Les couches de convolution

Le but principal de la convolution est l'extraction d'informations à partir de l'image d'entrée. La convolution est effectuée en faisant glisser une matrice de poids, appelée filtre, sur l'image d'entrée et en multipliant à chaque position les valeurs de filtre avec les valeurs de l'entrée. Ce filtre est appliqué à chaque fois sur une zone spécifiée appelée champ réceptif. Les résultats de la multiplication sont ensuite additionnés en un seul nombre. La sortie de la convolution du filtre sur toute l'image est une matrice appelée la carte de caractéristiques (figure 2.2).

De nombreux filtres seront utilisés dans chaque couche de convolution, et chaque filtre vise à identifier un motif de base spécifique qui constitue les objets dans l'image tels que les formes et les bords. Les cartes de caractéristiques sont ensuite disposées les unes sur les autres pour former un volume de sortie en tant que sortie finale. La profondeur de la sortie finale correspond

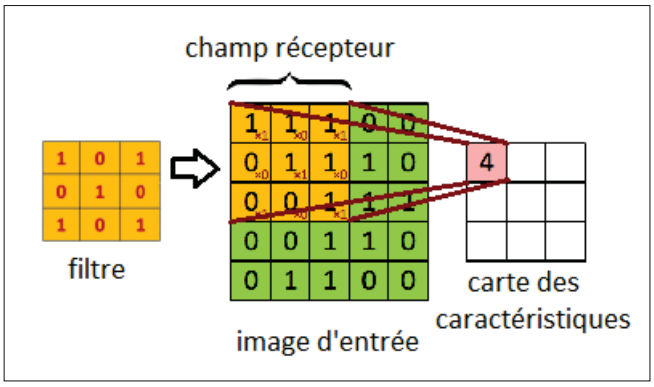

Figure 2.2 Convolution avec un filtre de taille 3×3. La région où le masque est appliqué est appelée champ réceptif. Figure adaptée de (NVIDIA, 2018).

au nombre de filtres utilisés dans la couche de convolution. La taille spatiale de ces résultats dépend de la taille des filtres et de certains paramètres supplémentaires tels que le pas et le remplissage à zéro.

Le pas est le nombre de pixels par lesquels le filtre de convolution se déplace à chaque fois sur la matrice d'entrée. Un pas de 1 est le plus utilisé, ce qui signifie que le filtre glisse sur l'entrée d'un pixel à la fois. Avec des valeurs de pas plus élevées, le filtre saute sur un grand nombre de pixels à la fois ce qui produit une sortie plus petite. Dans certains travaux, le concepteur de réseau convolutif préfère conserver la même taille de sortie que l'entrée. Dans ce cas, un processus connu sous le nom de remplissage avec zéro pourrait être utilisé. Cela nécessite seulement l'ajout de zéro pixel aux bordures de l'image d'entrée. Le mouvement du filtre avec un pas de 1 avec l'ajout d'un seul remplissage de zéro conservera la taille de l'entrée d'origine. La taille spatiale de chaque carte de caractéristiques dépend de tous ces paramètres, elle peut être calculée comme suit.

$$S = \frac{E - F + 2 \times Z}{P} + 1 \tag{2.1}$$

Où, S et E correspondent aux tailles d'entrée et de sortie, F est la taille du filtre, Z est le nombre de remplissage de zéro et P est la valeur de pas.

Comme les réseaux de neurones réguliers, la sortie de la convolution est passée à travers la fonction d'activation pour rendre la sortie non linéaire.

2.2.2 Fonction d'activation

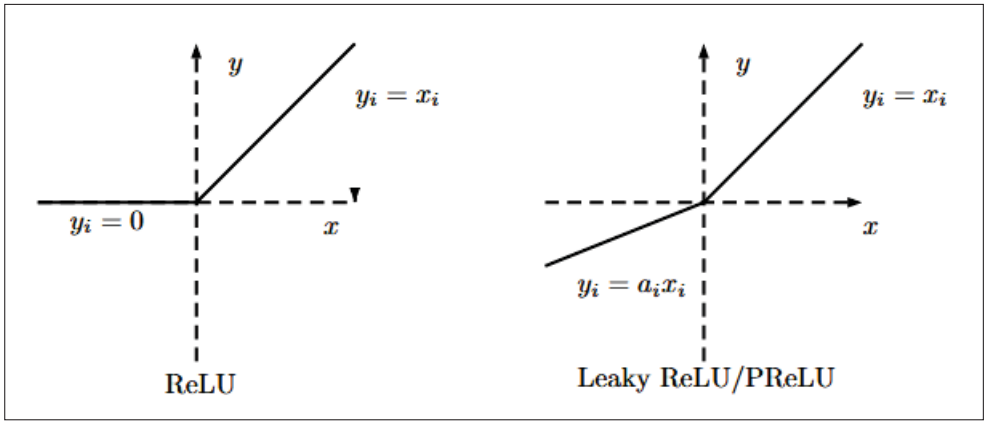

Figure 2.3 ReLU, Leaky ReLU et PReLU. Pour PReLU, a_i est appris dans l'entraînement par propagation arrière et pour Leaky ReLU a_i est fixe. Figure reproduite et adaptée avec l'autorisation de (He *et al.* (2015)).

La convolution est une opération linéaire. Le but de la fonction d'activation est d'introduire la non-linéarité. C'est une opération supplémentaire utilisée après chaque opération de convolution. La fonction tangente hyperbolique et la fonction sigmoïde ont été largement utilisées dans l'apprentissage automatique et dans certaines implémentations de réseaux neuronaux de base. L'unité linéaire rectifiée (ReLU), introduit par Nair & Hinton (2010), devient la non-linéarité la plus fréquemment utilisée dans les réseaux de neurones convolutionnels. Cette fonction met simplement les entrées négatives à zéro. Tous les éléments positifs, l'information spatiale et la profondeur restent inchangés. Toutes les valeurs négatives deviennent nulles, ce qui réduit la capacité du modèle à s'entraîner ou à s'ajuster correctement aux données. Le problème ici est que la sortie sera toujours nulle si l'entrée est négative aussi que le gradient. Cela peut essentiellement désactiver (tuer) les neurones et les empêcher d'apprendre. Pour surmonter ce

problème, Maas *et al.* (2013) ont proposé le Leaky ReLU qui est une variante de ReLU. Cette fonction n'affecte aucune sortie nulle pour une entrée négative. Elle compresse les entrées négatives avec un facteur prédéfinie. Cela permet de conserver la partie négative dans les cartes de caractéristiques. He *et al.* (2015) ont proposé l'unité linéaire paramétrique rectifiée (PReLU) qui généralise aussi le ReLU traditionnel. La fonction d'activation générique est définie comme suit :

$$f(x_i) = \max(0, x_i) + a_i \min(0, x_i) \tag{2.2}$$

Dans la fonction PReLU, contrairement au fonction LReLU, le paramètre a_i est appris lors de l'apprentissage avec les autres paramètres du réseau neuronal par propagation arrière. Xu *et al.* (2015) ont examiné les performances de divers types de fonctions d'activation rectifiées telles que ReLU, Leaky ReLU et PReLU (figure2.3). Basé sur des évaluations sur différentes tâches de classification d'image, ils ont prouvé que la fonction PReLU peut réduire l'erreur dans les étapes d'apprentissage et de test plus que le ReLU normal.

Après la convolution et la fonction de non-linéarité, la plupart des RNC ajoutent une couche de sous-échantillonnage (Pooling) entre les couches de convolution. Elle est utilisé en continu pour réduire la dimensionnalité et le nombre de paramètres. Cela raccourcit le calcul et le temps d'apprentissage dans le réseau et contrôle le sur-apprentissage.

2.2.3 Sous-échantillonnage

Figure 2.4 Différents types de pooling 2×2 avec un pas de 2. Pour le pool maximum, la sortie est la valeur maximale dans chaque fenêtre de taile 2×2. Pour la mise en pool moyenne, la sortie est la moyenne des valeurs dans chaque fenêtre. Chaque couleur représente la fenêtre utilisée et sa sortie correspondante.

Le sous-échantillonnage, également appelé pooling, réduit la taille des cartes d'activation en conservant les informations les plus importantes. Comme les couches convolutives, le pooling consiste à faire glisser une fenêtre sur les cartes d'activation. Il fonctionne indépendamment sur chaque tranche spatiale de l'entrée et redimensionne sa taille. Il existe plusieurs types de sous-échantillonnage, les plus populaires sont le pool maximum et le pool moyen (figure 2.4).

Le pool maximum consiste à faire glisser une fenêtre sur la carte d'activation et à prendre à chaque fois le maximum de valeurs. Pour le pool moyen, la sortie à chaque fois est la moyenne des valeurs dans la fenêtre. La taille de la fenêtre et la foulée utilisée doivent être prises en compte. Contrairement à l'opération de convolution, la mise en pool n'a pas de paramètres.

Une nouvelle version de sous-échantillonnage, connue sous le nom de pool moyens globales, a été proposée par Lin *et al.* (2013). L'objectif était d'éviter l'utilisation de couches traditionnelles entièrement connectées à la fin du réseau qui nécessitent beaucoup de temps en raison du grand nombre de paramètres qu'elles utilisent. L'idée était de générer une carte de caractéristiques pour chaque catégorie correspondante de la tâche de classification dans la dernière couche de convolution. Ensuite, au lieu d'utiliser les couches entièrement connectées, une moyenne de chaque carte de caractéristique est calculée puis passée directement par une couche softmax.

De manière similaire aux couches de regroupement moyennes, les couches de regroupement moyennes globales sont utilisées pour réduire les dimensions spatiales d'un volume tridimensionnel. Ils effectuent un type de réduction de dimensionnalité plus extrême, où le volume d'entrée est réduit à des dimensions de taille de $1 \times 1 \times n$ (n est le nombre des cartes de caractéristiques) en prenant simplement la moyenne de toutes les valeurs dans chaque carte.

2.2.4 Sur-échantillonnage

L'inverse exact du pooling (sous-échantillonnage) n'est pas possible dans le réseau de neurones convolutif. Le sur-échantillonnage, connu aussi par unpooling, est un inverse approximatif du pooling. Au cours de l'étape de sous-échantillonnage, les emplacements des pixels choisis à

chaque application d'un pooling sont enregistrés dans un ensemble de variables de commutateur. Ces variables sont ensuite utilisées pour la nouvelle reconstruction des régions en gardant la même structure avant l'application de unpooling. Par exemple, dans le pooling maximale, les pixels qui ont été maximum contiennent à nouveau les nouveaux maximums.

Après les couches de convolution et de sous-échantillonnage, les cartes d'activation doivent être préparées en tant qu'entrée pour les couches de classification. Cependant, ces couches finales ne peuvent accepter que des données unidimensionnelles. Le volume des cartes d'activation doit donc d'abord être aplati en un vecteur. Ensuite, le vecteur est utilisé comme entrée pour les couches de classification.

2.2.5 Couches entièrement connectées

Dans la plupart des architectures des RNC, le bloc convolutif du réseau est suivi par une ou plusieurs couches entièrement connectées. Ces couches prennent en entrée les cartes de caractéristiques aplaties et les transmettent à travers le réseau de neurones. Elles calculent le score de chaque classe à partir des entités extraites de haut niveau des couches de convolution. La dernière couche entièrement connectée est utilisée comme couche de classification. Elle contient un seul noeud pour chaque classe cible dans le modèle.

2.2.6 Fonction d'activation de la dernière couche

Il existe diverses fonctions d'activation qui pourraient être appliquées à la dernière couche entièrement connectée. Cette fonction diffère d'une tâche à l'autre et elle est liée au résultat de sortie souhaité. La fonction d'activation appropriée pour la tâche de classification des pixels multi-classes est la fonction *softmax*. Cette fonction est la plus répandue et la plus utilisée pour les tâches de segmentation d'images. Elle normalise les valeurs réelles de sortie de la dernière couche entièrement connectée en des probabilités des classes cibles, où chaque valeur est comprise entre 0 et 1 et la somme de toutes les valeurs est 1.

2.3 Architectures de classification populaires

Il existe plusieurs architectures dans le domaine des réseaux convolutifs. Les premières applications réussies de ces réseaux ont été développées par Yann LeCun dans les années 1990.

L'architecture la plus connue était LeNet-5 (LeCun *et al.* (1998)). Celle-ci a été conçue pour la reconnaissance des caractères manuscrits et imprimés à la machine. Elle a été entraînée et testée sur la fameuse base de données des chiffres manuscrits MNIST [1] pour classer l'entrée dans l'une des dix classes représentant les chiffres 0 à 9. LeNet-5 a été appliquée par plusieurs banques pour reconnaître les numéros manuscrits sur des chèques numérisés en images de 32x32 pixels.

La conception du LeNet-5 contient les éléments de base des réseaux de neurones convolutifs utilisés dans les modèles les plus récents. L'architecture LeNet-5 est composée de 7 couches au total ; trois couches convolutives, deux couches de sous-échantillonnage suivies d'un ensemble de couches entièrement connectées, et finie d'une couche sortie qui génère 10 sorties représentants les 10 chiffres de 0 à 9. Dans les couches de sous-échantillonnage, les pixels voisins dans une fenêtre de taille 4×4 sont additionnés, multipliés par un coefficient entraîné, ajoutés à un biais entraîné et transmis par la fonction *sigmoïde* pour la non-linéarité. Le sous-échantillonnage réduit le nombre de paramètres et de calculs dans le réseau. Il réduit aussi la dimension de chaque carte tout en conservant les informations les plus importantes.

Les limites du matériel informatique et la capacité de la mémoire pour l'apprentissage en réseau ont rendu difficile la mise en œuvre de tels algorithmes. Par conséquence, les réseaux de neurones ont été mis a l'écart par la recherche depuis 1998 jusqu'au 2010.

Au cours des dernières années, plus de données sont devenues disponibles en raison de l'augmentation du nombre de dispositifs d'image et de l'explosion des données sur Internet. La

1. La base de données MNIST (Modified National Institute of Standards and Technology), est une base de données de chiffres manuscrits. C'est un ensemble de données très utilisé dans la base de données d'apprentissage automatique des chiffres manuscrits. Elle comprend 60 000 exemples d'apprentissage et un ensemble d'essais de 10 000 exemples.

puissance des processeurs graphiques (GPU) a augmenté également, ce qui a encouragé à reprendre les recherches sur les réseaux de neurones.

En 2010, Dan Claudiu Ciresan et Jurgen Schmidhuber ont publié l'une des premières implémentations de réseaux de neurones utilisant un GPU (Cireşan *et al.* (2010)). Ils ont obtenu un taux d'erreur de 0,35% sur la base de données MNIST. Ils ont prouvé dans leur étude que l'utilisation de nombreuses couches cachées, de nombreux neurones par couche, de nombreuses images d'apprentissage et des GPU a accéléré considérablement l'apprentissage et a mené à de meilleurs résultats.

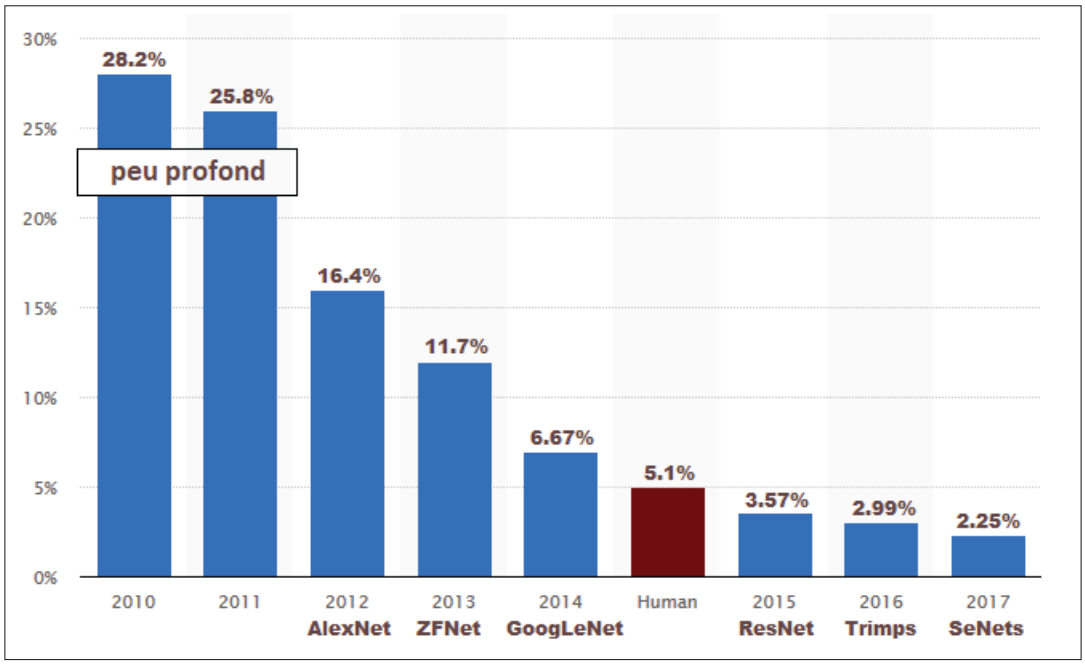

Figure 2.5 La progression de la performance de classement (erreur top-5%) des gagnants de la compétition ILSVRC au cours des dernières années.

En 2012, un modèle de réseau neuronal convolutionnel, connu sous le nom AlexNet (Krizhevsky *et al.* (2012)), a fait une percée dans la compétition ILSVRC [2]. Il a nettement surpassé les méthodes traditionnelles existantes et a remporté l'ILSVRC-2012 avec une erreur top-5 de 16,4% contre 26,2% atteint par le deuxième meilleur concurrent. C'était la première fois

2. ImageNet Large Scale Visual Recognition Challenge, ou "Compétition ImageNet de Reconnaissance Visuelle à Grande Échelle".

qu'une méthode d'apprentissage profond fut utilisée pour la classification d'images à grande échelle. AlexNet a une architecture similaire à LeNet mais plus profonde, avec plus de filtres par couche. Il se compose de cinq couches convolutives suivies de deux couches entièrement connectées et d'une couche Softmax pour la classification. Contrairement à LeNet, une opération de non-linéarité (ReLU) est utilisée après chaque convolution. Cette fonction est plus rapide que la fonction sigmoïde traditionnelle. À chaque étape d'apprentissage, la moitié des nœuds individuels de chaque couche entièrement connectée est retirée du réseau. Ce qui améliore la capacité du réseau à généraliser et à éviter le sur-ajustement des données d'apprentissage.

Depuis lors, une série de modèles de réseaux de neurones convolutionnels ont été proposés avec des architectures plus profondes et plus sophistiquées qui ont fait progresser l'état de l'art régulièrement sur la compétition ILSVRC comme le montre la figure 2.5.

En 2013, (Zeiler & Fergus (2014)) ont remporté la compétition ILSVRC-2013 avec une architecture connue sous le nom ZFNet. Leur modèle a été une amélioration de l'architecture AlexNet. Il a été conçu en ajustant le réseau en réduisant la taille du filtre de la première couche de (11×11) à (7×7), ainsi que le pas de convolution de 4 à 2. Cette amélioration conserve plus d'informations sur les pixels et réduit le nombre de paramètres du réseau. Les auteurs de ZFNet ont aussi introduit une nouvelle technique de visualisation des cartes de caractéristiques utilisant ce qu'on appelle la déconvolution. Cette opération fait l'opposé d'une couche convolutive permettant de visualiser les cartes caractéristiques dans le domaine des pixels et d'examiner différentes cartes d'activation et leurs relations avec l'espace d'entrée.

En 2014, Google (le gagnant de la compétition ILSVRC-2014) a développé une architecture de 22 couches connue sous le nom GoogLeNet (Szegedy *et al.* (2015)). La nouveauté de ce travail a été l'utilisation des couches appelées blocs d'*inception*. Ces blocs sont composés de couches de convolution de différentes tailles (1×1, 3×3 et 5×5) et d'une couche de pool maximum qui sont toutes calculées en parallèle. Des filtres de convolution de taille 1×1 sont utilisés avant les convolutions coûteuses de filtres plus grands ce qui réduit le coût total de calcul. Tous

les résultats sont concaténés les uns après les autres et transmis à la couche suivante. 9 blocs d'inception sont empilés les uns sur les autres avec deux couches de pool maximum au milieux qui servent à réduire les dimensions spatiales. La couche de classification effectue à la fin une opération de pool moyenne suivie d'une activation softmax.

La même année, (Simonyan & Zisserman (2014)) (deuxième meilleur taux d'erreur top-5 à la compétition ILSVRC-2014) ont introduit un nouveau modèle appelé VGGNet. Ce réseau de 19 couches a été une amélioration du réseau AlexNet. Il se compose de cinq blocs principaux d'opérations de convolution connectés via des couches de pool maximum suivies par trois couches entièrement connectées. Chaque bloc de convolution contient une série de couches de convolution par des filtres de taille 3×3. L'utilisation de filtres de petites tailles permet d'extraire des caractéristiques plus complexes. Cette étude a prouvé d'une part que l'utilisation de la convolution avec des filtres 3×3 en séquence peut émuler l'effet d'un filtre plus grand et d'autre part que l'utilisation de plusieurs couches non linéaires avec des filtres de petite taille permet au réseau d'apprendre des fonctionnalités plus complexes à moindre coût.

La révolution des réseaux de neurones convolutifs a eu lieu en décembre 2015. ResNet (He *et al.* (2016)), un réseau de 152 couches, a remporté la compétition ILSVRC-2015 avec un incroyable erreur top-5 de 3,57%. Ce résultat a dépassé la performance humaine (erreur top-5 de 5,1%) sur l'ensemble de données ImageNet. Dans ResNet, les couches de convolutions sont divisées en blocs résiduels. La première couche dans cette architecture est une couche de convolution avec un filtre de taille 7×7 suivie d'une couche de pool maximum, de blocs résiduels, d'une couche de pool moyenne et d'une couche de classification. Le bloc résiduel est composé de trois couches de convolution avec des filtres de tailles 1×1 et 3×3. Une connexion résiduelle dans chaque bloc est utilisée pour fusionner par sommation la sortie transmise par les deux couches de convolution avec l'entrée original du bloc.

En 2016, l'équipe Soushen de l'institut de recherche Trimps de Chine ont propose un réseau assemblant 5 versions différentes d'inception et de ResNet. Ce réseau a remporté la première

place pour la classification des objets dans la compétition ILSVRC-2016 avec une erreur top-5 de 2,99%.

En 2017, Hu *et al.* (2017) ont obtenu un taux d'erreur top-5 de 2,251% sur les données de test de la compétition ILSVRC-2017 avec un ensemble de réseaux appelé SENets. Leur étude combine des modules alternés de compression et d'excitation (SE : squeeze-and-excitation). Les blocs SE utilisent des informations globales pour pondérer sélectivement les caractéristiques informatives et supprimer les caractéristiques moins utiles. L'opération de compression effectue un pool maximum global de canaux. Elle collecte des informations et les conserve dans une dimension inférieure. L'opération d'excitation sert à décider lesquelles des cartes de caractéristiques sont vraiment importantes.

La plupart des nouveaux modèles sont des versions améliorées par rapport aux modèles novateurs précédents ou un assemblage de réseaux existants. Après avoir obtenu des résultats supérieurs aux performances humaines avec des taux d'erreur inférieurs à 3% sur la base de données ImageNet, les efforts de la recherche se sont plus focalisés à améliorer le coût de calcul et la minimisation du nombre de paramètres utilisés dans le modèle.

2.4 Architectures de segmentation

La sortie prévue dans les réseaux de neurones de classification est une étiquette de classe unique. Cependant, dans de nombreuses tâches visuelles, en particulier dans le traitement d'images médicales, le résultat souhaité doit inclure une segmentation de certaines structures ou en d'autres termes, attribuer une étiquette de classe pour chaque pixel de l'image d'entrée. Le succès des modèles d'apprentissage pour la classification a incité les chercheurs à explorer la capacité de tels exploits pour résoudre les problèmes d'étiquetage au niveau des pixels. Cet effort a donné naissance à de nouvelles techniques de segmentation telle que la segmentation sémantique.

À notre connaissance, la première méthode d'apprentissage profond utilisée pour la segmentation sémantique était la classification par patch. Inspiré par le succès des réseaux de neurones de

classification précédents, cette approche classe séparément chaque pixel en utilisant un patch d'image centré sur le pixel d'intérêt.

Ning *et al.* (2005) ont proposé un réseau de neurones convolutif permettant d'étiqueter les pixels sur des séquences d'images microscopiques de petits groupes de cellules. Leur réseau est composé de 3 couches de convolution avec des filtres de tailles (7×7, 6×6 et 6×6). Les deux premières couches de convolution sont chacune suivie d'une couche de sous-échantillonnage. Contrairement à LeNet-5, le sous-échantillonnage utilisé dans leur réseau calcule la moyenne dans un voisinage de (2×2). Cette moyenne est ensuite ajoutée à un biais entraîné, puis multipliée par un coefficient entraîné. Le résultat est enfin transmis par la fonction *tanh*. L'étiquette prédite pour chaque pixel est obtenue en appliquant les opérations de convolution et de sous-échantillonnage à une fenêtre de taille (40×40) centrée sur le pixel d'intérêt de l'image d'entrée. La couche finale comprend cinq unités, chacune représente une classe.

En utilisant la même stratégie de classification par pixel, Ciresan *et al.* (2012) ont proposé un réseau de neurones convolutif pour la segmentation des membranes de neurones biologiques dans des images microscopiques électroniques. Leur apport majeur était l'utilisation des couches de pool maximum au lieu du sous-échantillonage après chaque couche de convolution. La dernière couche est une couche entièrement connectée. Elle contient un certains nombres de neurones représentant chacune une catégorie prédéfinie (deux dans leur cas). L'application de la fonction softmax leur permet à la fin de calculer la probabilité d'appartenance de chaque pixel à une des dites catégories.

L'approche du classification par patch fonctionne bien dans de nombreuses situations, mais présente deux inconvénients majeurs. Tout d'abord, le réseau s'exécute séparément pour chaque patch, ce qui le ralentit généralement. Deuxièmement, l'utilisation de patch de grande taille nécessite l'utilisation d'un plus grand nombre de couches de pool maximum, ce qui réduit la précision de localisation, tandis que l'utilisation d'un patch de petite taille empêche le réseau d'apprendre les propriétés du contexte global. Puisque la sortie du réseau de classificateur de pixel par patch est une étiquette pour un seul pixel, la taille de l'image d'entrée doit être fixée.

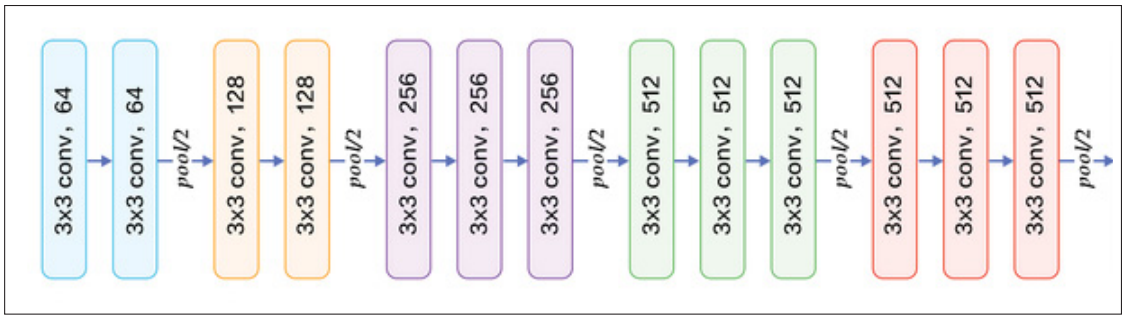

Figure 2.6 La partie convolutive (encodeur) de l'architecture VGG-16. Des filtres de taille 3×3 sont utilisés pour toutes les couches de convolution. Le nombre dans les rectangles correspond au nombre de filtres utilisés pour chaque couche. Les opérations de pool maximum sont utilisées entre les blocs de convolution pour réduire la dimension spatiale des cartes de caractéristiques. Figure adaptée de (OpenClassrooms, 2019).

Pour dépasser ces limites, Long *et al.* (2015) ont adapté certains des réseaux de classification les plus réputés (AlexNet, VGGNet et GoogLeNet) pour créer un réseau de neurones entièrement convolutif. Dans leur étude, ils ont mis au point les trois réseaux de classification différents. Les meilleurs résultats ont été obtenus avec une architecture basée sur le réseau VGG-16 qui est une version de VGGNet à 16 couches (13 couches de convolutions et 3 couches entièrement connectées). Leur idée était de conserver les couches de convolutions (figure 2.6) et de remplacer les couches entièrement connectées par des couches de convolutions avec des filtres de taille 1×1. L'utilisation des couches entièrement convolutives permet de traiter des images de taille arbitraire. Du fait que les couches de pool maximum sont utilisées, la taille de sortie de la couche finale est réduite. Pour redimensionner la carte de sortie à la taille d'origine de l'entrée, ils ont utilisé des couches de sur-échantillonnage avec une déconvolution initialisée de manière linéaire. Utiliser uniquement la sortie sur-échantillonnée de la dernière couche n'était pas suffisant pour obtenir une bonne segmentation. Pour cette raison, ils ont fusionné la sortie de la couche finale avec les cartes de caractéristiques de couches peu profondes de différents niveaux. Cette combinaison permet d'obtenir des prédictions locales tout en respectant les structures globales, ce qui permet au réseau d'obtenir des segmentations plus précises.

Au meilleur de notre connaissance, l'utilisation des couches de sur-échantillonage a été d'abord présentée par Huang *et al.* (2007) dans une architecture d'apprentissage non supervisée. Un réseau de déconvolution a également été introduit par Zeiler *et al.* (2011) pour reconstruire les images d'entrée à partir de sa représentation fonctionnelle. Zeiler & Fergus (2014) ont utilisé des couches de déconvolution dans leur travail pour comprendre le comportement des réseaux de neurones convolutionnels en visualisant des cartes de caractéristiques de différents niveaux.

Afin de mieux récupérer les pertes d'informations causées par les couches de sous-échantillonnage, les chercheurs ont poussé l'idée proposée dans le travail de Long *et al.* (2015) un peu plus loin en élargissant la partie de sur-échantillonnage des cartes de caractéristiques et en ajoutant davantage de liaisons de sauts entre les couches de différents niveaux. Ce type d'architecture est connu sous le nom de réseau encodeur-décodeur.

La première partie de l'architecture du réseau est appelée encodeur. Elle est similaire à la partie de convolution d'un réseau de classification. Des blocs de couches de convolution suivis chacun d'une couche de sous-échantillonnage sont utilisés pour convertir l'image d'entrée en cartes de caractéristiques.

La deuxième partie est le décodeur qui est une version miroir du encodeur. Elle comprend plusieurs ensembles de couches de déconvolution et de sur-échantillonnage pour récupérer les informations spatiales à partir de la sortie du encodeur. La dernière couche du décodeur est une couche de classification softmax qui produit une segmentation finale de la même taille que l'image d'entrée.

Il existe généralement des connexions d'encodeur vers décodeur pour aider le décodeur à réduire la perte d'informations et à mieux récupérer les détails de l'objet d'intérêt. De nombreuses architectures ont été développées sur la base du concept encodeur-décodeur (Badrinarayanan *et al.* (2015); Noh *et al.* (2015); Ronneberger *et al.* (2015); Badrinarayanan *et al.* (2017)). Ce qui différencie une architecture des autres est la façon dont elle relie les différents niveaux de l'encodeur avec leurs parties inversées dans le décodeur.

Badrinarayanan *et al.* (2017) ont développé une architecture de forme encodeur-décodeur, connue sous le nom SegNet, pour la segmentation multi-classes des pixels. La partie encodeur de SegNet est similaire aux 13 couches convolutives du réseau VGG-16 (figure 2.6). Elle est composée de cinq blocs. Chaque bloc est constitué de couches de convolution 2D, de normalisation par lots et d'une activation d'unité rectifiée linéaire (ReLU). La dernière couche du bloc est une couche de pool maximum qui vise à réduire la dimension spatiale des cartes de caractéristiques. Pour récupérer la haute résolution des cartes de caractéristiques, SegNet a utilisé un processus inverse de la partie encodeur en tant que décodeur avec le même nombre de blocs. Chaque bloc est constitué d'une couche de sur-échantillonnage suivie par des couches convolutives entraînées, normalisation par lots et d'une activation (ReLU). Un élément clé de l'architecture SegNet est l'utilisation des emplacements restaurés de pool maximum pour effectuer un sur-échantillonnage dans la partie décodeur. Ce processus permet de conserver des détails haute fréquence dans les images segmentées avec un faible coût de consommation de mémoire et un nombre réduit de paramètres d'apprentissage dans la phase de décodeur.

Une autre architecture de type encodeur-décodeur, connue sous le nom UNet, a été proposé par Ronneberger *et al.* (2015). UNet a montré d'excellentes performances en segmentant différentes cibles dans différentes modalités d'images médicales. L'architecture est composée de 23 couches de convolution au total. Elle consiste en un chemin de contraction (encodeur) et un chemin d'expansion (décodeur). Comme SegNet, la partie encodeur est composée de blocs de convolution répétés. Chaque bloc est constitué de deux couches de convolutions avec des filtres de taille (3×3), chacune suivie d'une activation (ReLU) et d'une opération de pool maximum de (2×2). Le nombre de cartes de caractéristiques est doublé après chaque sous-échantillonnage. La partie d'expansion de UNet est une version inversée de la partie contraction. Le nombre de cartes de caractéristiques est réduit de moitié après chaque bloc. Pour connecter la partie codeur à la partie décodeur correspondante, une copie des cartes de caractéristiques est concaténée avec les cartes correspondantes de la partie décodeur. La dernière couche utilise une convolution avec des filtres de tailles (1×1) pour fournir des cartes de classification de même nombre que les classes souhaitées. La taille de sortie est inférieure à la

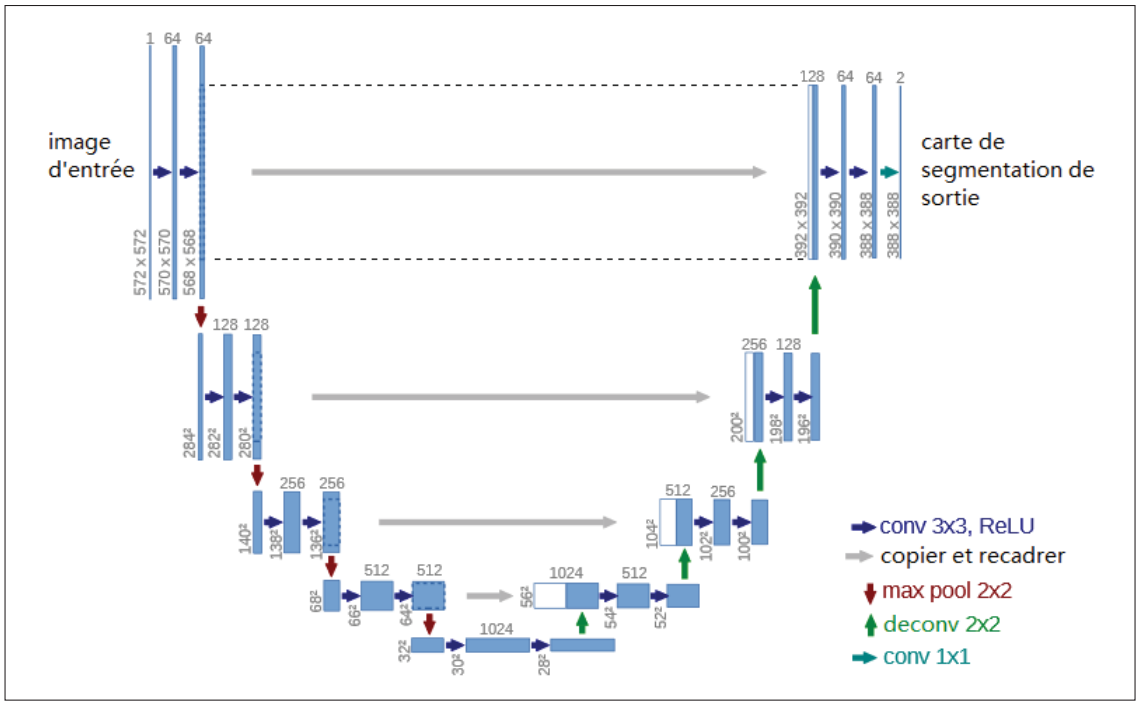

Figure 2.7 Architecture U-net. Chaque rectangle bleu correspond à une carte de caractéristiques multicanaux. Le nombre de canaux est indiqué en haut du rectangle. La taille spatiale est indiquée sur le bord inférieur gauche du rectangle. Les rectangles blancs représentent les cartes de caractéristiques copiées. Les flèches indiquent les différentes opérations. Figure reproduite et adaptée avec l'autorisation de (Ronneberger *et al.* (2015)).

taille d'entrée (voir figure 2.7) en raison de l'utilisation de la convolution sans l'ajout de pixels autour de l'image. Pour obtenir la même taille que l'entrée, l'image entière est prédite partie par partie à l'aide d'une stratégie de mosaïque de chevauchement.

Le nombre d'échantillons annotés dans le domaine médical est généralement limité car cela devrait être fait par des experts. D'autre part, les réseaux d'apprentissage dans un tel domaine nécessitent un grand nombre d'échantillons. Pour surmonter cette limite, Ronneberger *et al.* (2015) ont augmenté la taille de l'ensemble d'apprentissage en utilisant des transformations aléatoires sur les images d'entrée telles que les inversions, les distorsions, les rotations et en particulier les déformations élastiques.

Cependant, les architectures encodeur-décodeur n'étaient pas la seule solution pour réduire les inconvénients liés à l'utilisation de couches de sous-échantillonnage. Yu & Koltun (2015) ont développé un nouveau modèle permettant d'obtenir le contexte global de l'image sans réduire la taille de l'image en utilisant une nouvelle forme de convolution appelée convolution dilatée (ou convolution à trous). La convolution dilatée est utilisée pour maintenir la résolution spatiale des cartes de caractéristiques d'une couche à l'autre en augmentant simplement le champ de vision du filtre. Ce processus est effectué en insérant des zéros entre les pixels du filtre de convolution. Yu & Koltun (2015) ont proposé un réseau basé sur le modèle VGG-16 qui consiste à utiliser une série de couches de convolution dilatées de différents facteurs. Ces facteurs de dilatation augmentent suivant une fonction exponentielle allant des couches peu profondes aux couches plus profondes.

Chen *et al.* (2014) ont choisi de traiter les inconvénients des opérations de sous-échantillonnage comme une étape de post-traitement. Ils ont utilisé un réseau basé sur le modèle VGG-16 et ont proposé d'affiner la segmentation de la sortie pour récupérer les détails les plus fins. Leur idée était de combiner la carte des scores de la dernière couche de réseau avec un modèle de champ aléatoire conditionnel (CRF) entièrement connecté qui a été développé par (Krähenbühl & Koltun (2011)). Le CRF utilise une fonction d'énergie qui combine les informations d'interactions entre les paires de pixels, leurs positions et leurs intensités, avec la probabilité d'appartenance prédite par le réseau pour chaque pixel. Cette étape de post-traitement permet de récupérer les structures détaillées perdues lors de la segmentation finale.

Chen *et al.* (2018) ont proposé une nouvelle version de leur approche précédente. Plus que le VGG-16, ils ont construit une seconde architecture basée sur le réseau ResNet-101 (ResNet à 101 couches). Comme dans la première version, ils ont utilisé la convolution dilatée et le CRF comme une étape de post-traitement. L'amélioration consistait à utiliser un traitement à plusieurs échelles en transmettant les cartes de caractéristiques par une mise en pool à trous en pyramide spatiale (ASPP). Cette technique utilise des filtres dilatés pour augmenter le champ de vision lors de la convolution dans une forme de pyramide spatiale afin de capturer des cartes de caractéristiques à plusieurs échelles. ASPP utilise une série de convolutions dilatées,

de manière parallèle, avec des taux de dilatation différents. Les cartes de caractéristiques des différentes branches parallèles sont ensuite interpolées de manière bilinéaire à la résolution de l'image d'origine. La sortie est finalement générée en prenant la réponse maximale à chaque position parmi les différentes sorties de chaque branche.

Chen *et al.* (2017) ont développé une troisième version de leurs travaux précédents. Ils ont présenté différentes architectures basées sur le réseau ResNet utilisant des modules en cascade et en parallèle de convolutions à trous. Ils ont proposé un nouveau modèle de ASPP composé d'une convolution par un filtre de taille (1×1) et trois convolutions à trous par des filtres de taille (3×3) de différents facteurs de dilatation. Une normalisation par lots est utilisée après chacune des couches de convolution parallèles. Les sorties sont ensuite concaténées et transmises à une autre convolution avec un filtre de taille (1×1) pour créer la sortie finale. Cette version proposée améliore les résultats de la segmentation par rapport aux versions précédentes sans utiliser le CRF comme étape de post-traitement.

Une autre étude, présentée par Zhao *et al.* (2017), a montré qu'une bonne compréhension de la représentation du contexte général d'une scène peut conduire à un bon étiquetage des pixels. Pour employer cette idée, ils ont proposé un réseau d'analyse de scènes pyramidales (PSPNet). Leur architecture consiste en un réseau ResNet à convolution dilatée utilisé comme extracteur des cartes de caractéristiques. Ces cartes sont ensuite transmises à un module de sous-échantillonage pyramidal pour calculer la moyenne des sous-régions avec quatre échelles différentes. Chaque branche correspondant à un niveau de pyramide est traitée par une couche de convolution de taille (1×1) afin de réduire leurs dimensions. En utilisant une interpolation bilinéaire, les sorties des niveaux de pyramide sont sur-échantillonnées à la même taille que la carte de caractéristiques d'origine. Les sorties sont ensuite concaténées avec les cartes de caractéristiques initiales, ce qui permet d'obtenir une combinaison d'informations de contexte locales et globales. Enfin, une couche de convolution est utilisée pour générer les prédictions finales de pixels.

2.5 Conclusion

Dans ce chapitre, nous avons d'abord présenté les outils de base permettant de concevoir une architecture de réseau de neurones convolutif. Nous avons ensuite présenté une analyse des architectures les plus populaires utilisées pour la classification d'images telles que AlexNet, VGGNet et ResNet. Nous avons discuté les améliorations et les idées clés de chaque réseau en fonction de leur performance lors du compétition annuelle ImageNet. Enfin, nous avons passé en revue certains des réseaux populaires utilisés pour les tâches de segmentation sémantique telles que les réseaux encodeur-décodeur et les architectures basées sur la convolution dilatée. Nous avons aussi expliqué en détail les différentes opérations développées par chaque approche pour obtenir une segmentation finale de la même taille que l'entrée initiale.

CHAPITRE 3

MÉTHODOLOGIE

3.1 Introduction

Après avoir compris les principaux outils permettant de concevoir une architecture de réseau neuronal, nous proposons dans ce chapitre un réseau neuronal entièrement convolutif pour segmenter la colonne vertébrale lombaire. Comme nous l'avons expliqué au chapitre précédent, le principal problème que tous les réseaux proposés visent à résoudre est la perte de certaines informations structurelles lors de la convolution et des opérations de sous-échantillonnage, ce qui produit une solution qui n'est pas optimale de façon globale. Pour résoudre ce problème, nous allons utiliser la coupe de graphe pour affiner les prédictions finaux. L'idée est de traiter les cartes de probabilités fournies par le réseau en les interpréter sous forme de graphe et formuler le problème de segmentation d'image en tant que problème de minimisation d'énergie. Nous allons utiliser dans notre étude trois bases de données différentes, la première contient les images IRM et uniquement les annotations des vertèbres, la deuxième contient les images IRM et les annotations des disques intervertébraux et la troisième contient les images IRM et les annotations des deux structures (vertèbres et disques). Pour la segmentations binaire, nous utiliserons la coupe de graphe et pour la segmentation multi-classes, nous utiliserons l'algorithme α-expansion.

3.2 Architecture de réseau proposée

La méthode de segmentation proposée est basée sur une architecture de réseau de neurones entièrement convolutif à deux dimensions. Cette architecture est composée de 13 couches au total avec la disposition suivante : 9 couches de convolution, 3 couches entièrement connectées et la couche de classification à la fin (figure 3.1). Pour les trois premières couches, nous avons utilisé 25 filtres de taille (9×9) dans chaque couche. Dans les trois couches de convolution suivantes, nous avons utilisé 50 filtres de taille (7×7) dans chaque couche. Et enfin, pour les

trois dernières couches de convolution, 75 noyaux de taille (5×5) sont utilisés dans chaque couche.

Pour préserver la résolution spatiale, nous avons évité l'utilisation de couches de regroupement. Un pas unitaire est utilisée pour toutes les couches de convolutions. Nous avons utilisé l'unité linéaire rectifiée paramétrique (PReLU) comme fonction d'activation à la place de l'unité linéaire rectifiée (ReLU).

Comme dans les réseaux neuronaux à convolution standard, les couches entièrement connectées sont ajoutées à la fin du réseau pour coder les informations sémantiques. Cependant, pour que le réseau ne contienne que des couches convolutives, nous avons converti ces couches en un ensemble de convolutions avec des filtres de taille (1×1). Cela permet au réseau de conserver des informations spatiales et d'apprendre les paramètres de ces couches comme dans d'autres couches convolutionnelles. Les trois couches entièrement connectées sont composées de 400, 200 et 150 unités cachées.

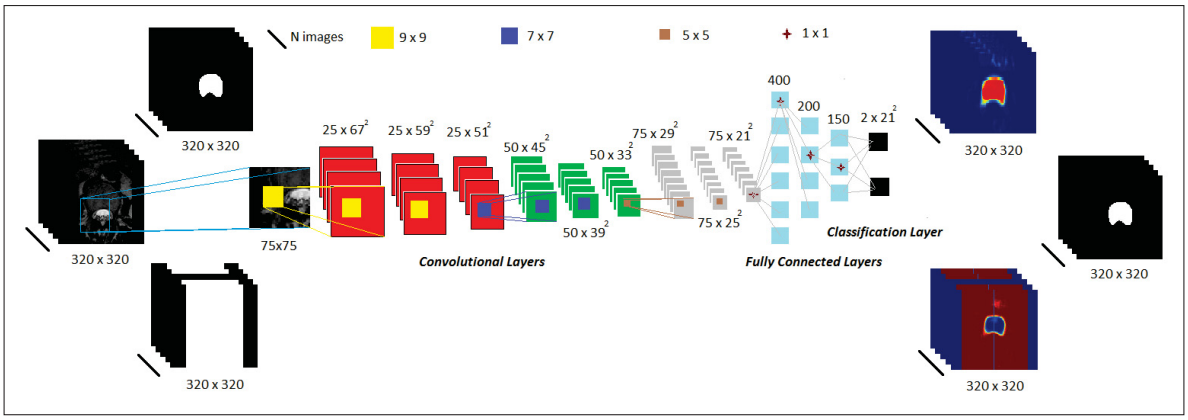

Figure 3.1 Architecture proposée.

L'un des points forts à prendre en compte lors de la construction d'un RNC consiste à initialiser correctement les poids et les biais. L'initialisation la plus simple consiste à définir tous les poids avec la valeur zéro. Dans ce cas, tous les neurones de toutes les couches effectuent le même calcul. En outre, la propagation arrière de l'algorithme d'apprentissage n'apporterait aucune

modification aux poids du réseau et le modèle deviendrait inutile. Pour surmonter ce problème, la plupart des RNC initialisent les poids de manière aléatoire au début de l'apprentissage. Dans notre modèle, nous avons utilisé l'initialisation proposée par (He *et al.* (2015)), qui a prouvé sa capacité à faire converger rapidement les architectures profondes. Dans chaque couche c, les poids sont initialisés sur la base d'une distribution gaussienne d'écart type moyenne nulle et les biais sont initialisés à zéro. La distribution d'initialisation utilisée est de la forme suivante :

$$W_c \sim \mathcal{N}\left(0, \sqrt{\frac{2}{n_c}}\right) \tag{3.1}$$

W_c indique le poids et n_c indique le nombre de connexions aux unités de la couche c. Par exemple, dans la première couche de convolution de la figure 3.1, l'entrée est une image à un seul canal (niveau de gris) et les filtres sont de taille (9×9), l'écart type dans ce cas est égal à $\sqrt{2/(1\times 9\times 9\times 1)} = 0.1571$.

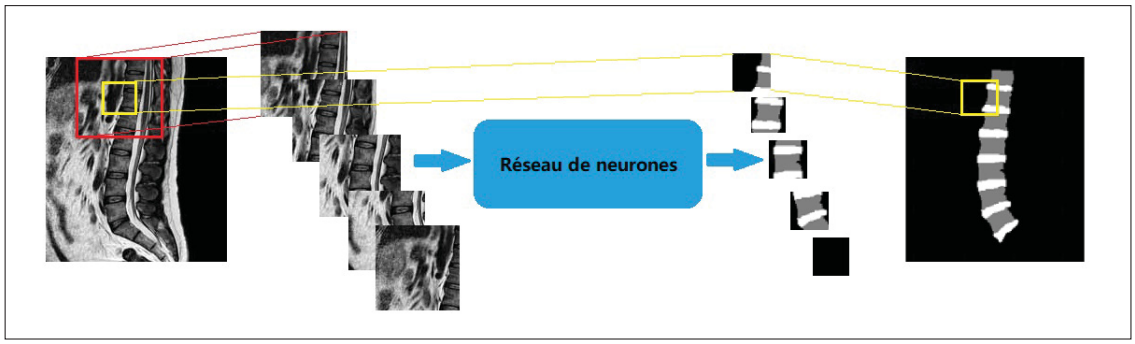

Figure 3.2 Le réseau est appris sur des patchs extraits des images d'apprentissage. La prédiction de la segmentation dans la zone jaune nécessite l'image d'entrée dans la zone rouge.

La cible que nous voulons segmenter occupe le centre des images. Pour nous assurer que notre modèle apprendra toutes les informations de la cible, nous avons utilisé un segment de taille 75×75. Chaque fois que nous glissons ce segment, l'entrée contiendra des pixels appartenant aux deux régions : l'arrière-plan et la cible. L'utilisation de segments de taille inférieure à l'image réelle fournit davantage de données d'apprentissage (figure 3.2). Cela peut être consi-

déré comme une augmentation des données d'apprentissage et peut améliorer les performances de notre modèle.

Tableau 3.1 Architecture détaillée du réseau neuronal entièrement convolutif utilisé dans notre étude. Conv2D désigne une couche convolutive à deux dimensions.

Type	Couche	taille de filtre	# de filtres	taille du sortie	taux de décrochage
couches de convolution	Conv2D	9×9×1	25	67×67×25	—
	Conv2D	9×9×1	25	59×59×25	—
	Conv2D	9×9×1	25	51×51×25	—
	Conv2D	7×7×1	50	45×45×50	—
	Conv2D	7×7×1	50	39×39×50	—
	Conv2D	7×7×1	50	33×33×50	—
	Conv2D	5×5×1	75	29×29×75	—
	Conv2D	5×5×1	75	25×25×75	—
	Conv2D	5×5×1	75	21×21×75	—
couches entièrement connectées	Conv2D	1×1×1	400	21×21×400	50%
	Conv2D	1×1×1	200	21×21×200	50%
	Conv2D	1×1×1	150	21×21×150	50%
couche de classification	Softmax	—	2	21×21×2	—

Les neurones de la dernière couche sont regroupés en N cartes de caractéristiques, où N représente le nombre de classes. La sortie est ensuite convertie en valeurs de probabilité normalisées via une fonction softmax. Les scores de répartition des classes pour chaque étiquette sont calculés comme suit :

$$P_n(x) = \frac{\exp(a_n(x))}{\sum_{n'=1}^{N} \exp(a'_n(x))} \tag{3.2}$$

$a_k(x)$ indique l'activation dans la k-ème carte de caractéristiques de la dernière couche de classification à la position de pixel x.

3.3 Fonction de perte

La fonction de perte la plus utilisée pour la segmentation de l'image est la perte d'entropie croisée par pixel. Elle est utilisée comme fonction de perte dans les réseaux de neurones qui ont des activations de type softmax dans la couche de sortie. Cette perte examine chaque pixel individuellement, en comparant l'étiquette prédite à l'étiquette cible de référence. Elle calcule la distance entre la distribution de sortie prédite par le modèle et la distribution réelle. Elle est

défini comme suit :

$$E(P',P) = -\sum_{n=1}^{N} P'_n \log P_n \tag{3.3}$$

P'_n est l'étiquette de vérité terrain du pixel d'apprentissage pour le n-ème classe et P_n est le résultat de prédiction du modèle pour le pixel d'apprentissage pour le n-ème classe.

La fonction de perte final E, est la moyenne de l'erreur pour tous les échantillons d'apprentissage d'un lot. Elle dépend de l'entrée ainsi que des paramètres du modèle d'apprentissage. Pour des prévisions précises, l'erreur calculée doit être réduite à chaque itération de l'apprentissage. Dans le réseau de neurones, cela se fait par propagation arrière. L'erreur est généralement propagée en arrière et utilisée pour modifier les poids et les biais. Ceci est fait en utilisant une fonction d'optimisation. Dans notre réseau, nous avons utilisé l'optimiseur RMSprop pour modifier les paramètres. La mise à jour vers l'arrière par RMSprop est basée sur les moyennes pondérées exponentielles du carré du gradient des paramètres précédents du réseau. Ces termes sont calculés comme suit :

$$v^{(t+1)}(\theta_i) = \gamma v^{(t)}(\theta_i) + (1-\gamma)\frac{\partial E}{\partial \theta_i}^2 \tag{3.4}$$

Le paramètre γ définit le poids entre la moyenne des valeurs précédentes et le carré de la valeur actuelle θ_i pour calculer la nouvelle moyenne pondérée. Dans notre modèle, nous définissons ce paramètre à 0,9.

$$\theta_i^{(t+1)} = \theta_i^{(t)} - \frac{\eta}{\sqrt{v^{(t+1)}(\theta_i) + \varepsilon}} \frac{\partial E}{\partial \theta_i} \tag{3.5}$$

η est le taux d'apprentissage qui représente la taille des mesures que nous prenons pour atteindre un minimum local. ε est une petite constante introduite pour éviter une division par zéro. Nous fixons le taux d'apprentissage initial à 0,001 et ε à 10^{-4}.

3.4 Réglage des hyperparamètres

3.4.1 Nombre d'époques

Le nombre d'époques correspond au nombre de fois où l'ensemble de données d'apprentissage complet passe en avant et en arrière dans le réseau neuronal. Nous avons formé notre réseau en 35 époques, chacune composée de 20 sous-époques.

3.4.2 Taille du lot d'images d'apprentissage

Dans l'apprentissage automatique, les données sont toujours trop volumineuses et nécessitent trop de mémoire si nous transmettons toutes les données dans le réseau en même temps. Pour résoudre ce problème, les données sont divisées en lots plus petits et transférées dans le réseau lot par lot. Les poids du réseau de neurones sont mis à jour à la fin de chaque étape pour les adapter aux échantillons utilisés.

À chaque sous-époque, un total de 1000 échantillons sont sélectionnés au hasard de l'ensemble d'images d'apprentissage. Ils sont traités dans le réseau avec des lots de taille 10.

3.4.3 Taux d'apprentissage

Le taux d'apprentissage contrôle la quantité de mises à jour des poids dans l'algorithme d'optimisation choisi. Dans notre modèle, le taux d'apprentissage initial est de 0,001. Il diminue d'un facteur de 2 après chaque 3 époques.

3.4.4 Décrochage

Pour améliorer les résultats de notre modèle, nous avons appliqué la régularisation par décrochage dans les dernières couches entièrement connectées avec un taux de 50%. Cela signifie que 50% des neurones dans chaque couche seront éliminés au hasard pendant l'apprentissage. Pendant la phase de prédiction, le décrochage est désactivé.

3.5 La coupe de graphe

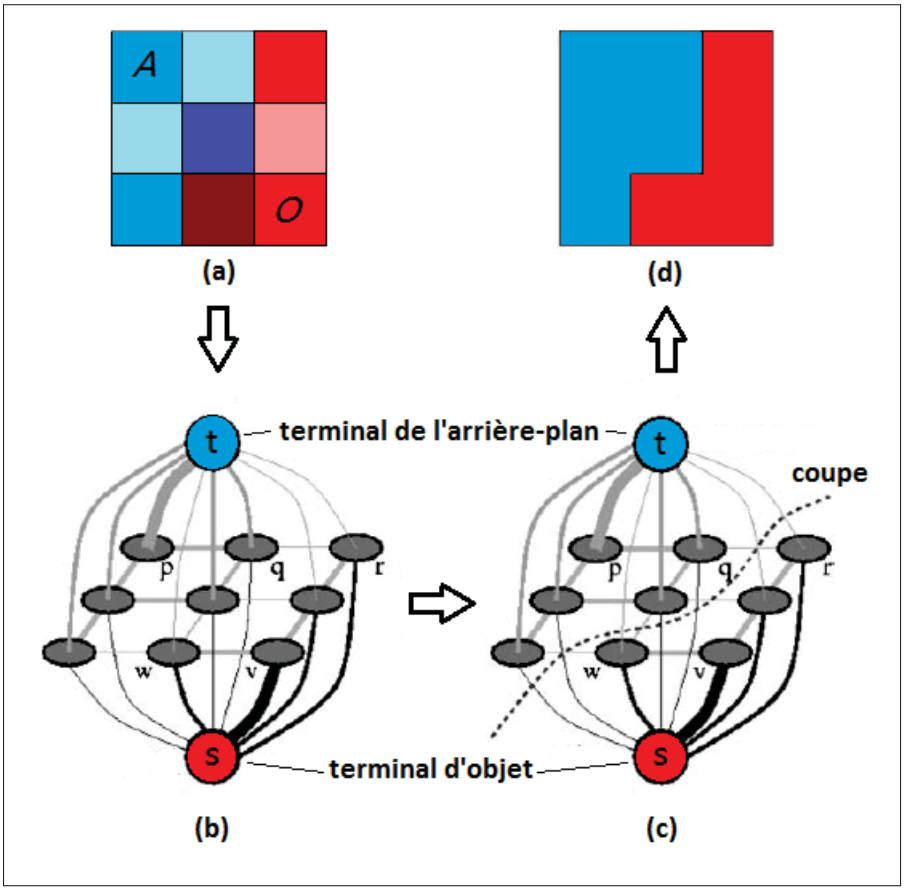

Figure 3.3 Illustration de la coupe de graphe pour la segmentation de l'image. (a) L'utilisateur fournit quelques connaissances de base en sélectionnant certains pixels appartenant à l'objet avec l'étiquette *O* et l'étiquette *A* pour l'arrière-plan. (b) Le coût de chacun des n-liens et des t-liens est reflété par son épaisseur. (c) La coupe correspond à l'énergie minimale de l'équation (3.7). (d) Résultats de segmentation. Figure reproduite et adaptée avec l'autorisation de (Boykov & Jolly (2001))

3.5.1 La théorie des graphes

La coupe de graphe est l'une des méthodes populaires utilisées pour résoudre le problème de la segmentation d'image. Elle fournit un étiquetage binaire de pixels dans une image comme appartenant à l'objet ou à l'arrière-plan. Cette méthode a été proposée par Boykov & Jolly

(2001), inspirée des travaux antérieurs de Greig *et al.* (1989). Dans leur approche, l'utilisateur doit fournir des connaissances de base sur l'objet et l'arrière-plan en attribuant une étiquette O pour l'objet et A pour l'arrière-plan comme indiqué sur la figure 3.3. Le processus de segmentation binaire par coupe de graphe consiste à considérer l'image comme un graphe et chaque pixel étant un noeud. On définit un graphe $\mathscr{G} = \{\mathscr{N}, \mathscr{E}\}$ où $\mathscr{N}$ est l'ensemble des noeuds et $\mathscr{E}$ l'ensemble des arêtes. Deux noeuds supplémentaires, appelés noeuds terminaux sont aussi considérés pour représenter l'objet O qu'on appelle la source s et l'arrière-plan A qu'on appelle le puits t. L'objectif de la segmentation est d'attribuer à chaque pixel de l'image une classe qui peut être soit O si le pixel est considéré appartenant à l'objet, soit A s'il est considéré appartenant à l'arrière-plan. Chaque paire de noeuds $(p,q) \in \mathscr{N}^2$ dans un voisinage $\mathscr{V}$ est connectée par un lien appelé n-lien. Chaque noeud $p \in \mathscr{N}$ est connecté aux noeuds terminaux s et t par deux liens respectifs appelés t-liens. Les t-liens et n-liens sont pondérés par des coûts w. La section d'un t-lien de p vers s (resp. t) permet d'attribuer l'étiquette O (resp. A) à ce pixel, définissant ainsi une segmentation de l'image. Une coupe est un sous-ensemble d'arêtes $\mathscr{C} \in \mathscr{G}$ telles que les bornes se séparent sur le graphe $\mathscr{G}(\mathscr{C}) = (\mathscr{N}, \mathscr{E} \backslash \mathscr{C})$. La coupe de segmentation entre l'objet et l'arrière-plan est tracée en trouvant la réduction du coût minimum sur le graphe (figure 3.3). Le coût d'une coupe $\mathscr{C}$ est donné par la somme des coûts des arêtes qu'elle coupe :

$$\mathscr{C} = \sum_{e \in \mathscr{E}} w_e \tag{3.6}$$

La meilleure segmentation est obtenue en calculant la réduction du coût minimum.

3.5.2 L'optimisation de l'énergie : fonction de coût

Le but de notre travail est de diviser les images en regions (classes) : objets et arrière-plan. La plupart des études qui utilisent la coupe de graphe nécessitent l'initialisation de utilisateur pour identifier des pixels appartenant à des objets et d'autres à l'arrière-plan. La segmentation se fait automatiquement jusqu'à obtention de partition optimale globale entre les objets et l'arrière-plan tout en prenant en compte l'initialisation de l'utilisateur. Cette partition optimale globale peut être atteinte en minimisant une fonction de coût. Cette fonction est définie en deux termes :

les propriétés des limites et des regions de la segmentation. Soit P l'ensemble des pixels d'une image, $\mathscr{V}$ un système de voisinage de toutes les paires $\{p,q\}$ de pixels voisins dans P et $S = \{S_1,...,S_p,...,S_{|P|}\}$ un vecteur de segmentation telque S_p spécifie l'étiquetage du pixel p. En fonction des propriétés de limite et de région de cet étiquetage S, la fonction de coût $E(S)$ est calculée comme suit :

$$E(S) = \lambda.D(S) + V(S) \tag{3.7}$$

avec

$$D(S) = \sum_{p \in P} D_p(S_p) \tag{3.8}$$

$$V(S) = \sum_{\{p,q\} \in \mathscr{V}} V_{\{p,q\}}.\delta(S_p, S_q) \tag{3.9}$$

et

$$\delta(S_p, S_q) = \begin{cases} 1, & \text{if } S_p \neq S_q \\ 0, & \text{sinon.} \end{cases} \tag{3.10}$$

Le terme D est appelé terme de données. Il mesure la cohérence entre l'étiquetage actuel S et l'image observée. D_p représente la pénalité d'associer le pixel p ayant l'étiquette S_p à l'objet ou à l'arrière plan. Les coûts des t-liens pour les deux étiquettes sont définis comme suit :

$$D_p(1) = -\ln \Pr(I_p | objet) \tag{3.11}$$

$$D_p(0) = -\ln \Pr(I_p | arriere - plan) \tag{3.12}$$

I_p représente l'intensité (niveau de gris) du pixel p dans les équations (3.11 et 3.12) . B est le terme limite qui mesure le degré de lissage de la segmentation. Les n-liens sont pondérés par un terme de régularisation conçu pour assurer la cohérence des contours dans un voisinage de pixels. Ce terme est noté par $V_{\{p,q\}} \geq 0$ dans l'équation (4.2). Il défini la pénalité pour une

discontinuité entre deux pixels voisins p et q. Ce terme est défini comme suit :

$$V_{p,q} \propto exp(-\frac{(I_p - I_q)^2}{2\sigma^2}) \tag{3.13}$$

σ est généralement une constante liée au bruit d'acquisition de l'image. Pour cela, cette fonction correspond à la répartition du bruit entre les pixels voisins. Cette fonction pénalise beaucoup les discontinuités entre les pixels d'intensités similaires lorsque $|I_p - I_q| < \sigma$. Cependant, si les pixels sont très différents, $|I_p - I_q| > \sigma$, alors la pénalité est proche de zéro. Le terme $V_{\{p,q\}}$ garantit que l'étiquetage global S est lisse. Il pénalise deux étiquettes voisines S_p et S_q si elles sont trop différentes. λ détermine la contribution relative du coût de pénalité du terme de région D par rapport au terme de propriétés des limites V. Trois contraintes sont imposées sur le terme de contour :

$$V(\alpha,\beta) = V(\beta,\alpha) \geq 0 \tag{3.14}$$

$$V(\alpha,\beta) = 0 \Leftrightarrow \alpha = \beta \; ou \; V(\alpha,\beta) \neq 0 \Leftrightarrow \alpha \neq \beta \tag{3.15}$$

$$V(\alpha,\beta) \leq V(\alpha,\gamma) + V(\gamma,\beta) \tag{3.16}$$

Le coût entre deux étiquettes différentes α et β devrait être positif et non nul (contrainte 3.14). Il est égal à zéro si les deux étiquettes sont identiques (contrainte 3.15). La dernière contrainte définit la règle du triangle qui consiste qu'une coupe courte est toujours moins chère ou égale que de prendre tout le chemin, c'est (contrainte 3.16).

L'algorithme standard de coupe de graphe est capable de trouver la solution optimale en cas d'étiquetage binaire pour la segmentation en deux classes. Toutefois, en ce qui concerne les problèmes d'étiquetage multiple où le nombre de classes est supérieur à deux, la minimisation de la fonction énergétique de la coupe de graphe standard devient NP-difficile et ne peut être appliquée directement pour trouver la solution optimale. Boykov *et al.* (2001) ont développé deux algorithmes de mouvement, α-expansion et α-β-swap, pour traiter les problèmes de segmentation multi-étiquetage.

3.6 α-expansion

L'algorithme α-expansion traite le problème de la segmentation multi-étiquettes en tant que sous-séquences de problème binaire. À chaque itération, l'algorithme attribue l'étiquette α à l'une des classes et prend toutes les autres comme une classe d'étiquette $\bar{\alpha}$. L'idée principale de l'algorithme est de segmenter successivement tous les pixels α et $\bar{\alpha}$ avec des coupes de graphes. La région d'étiquette α ne puisse que se développer. À chaque itération, l'étiquette des pixels ne peut être modifiée qu'en α ou conserve son étiquette. L'algorithme tente chaque classe possible pour α jusqu'à ce qu'il converge.

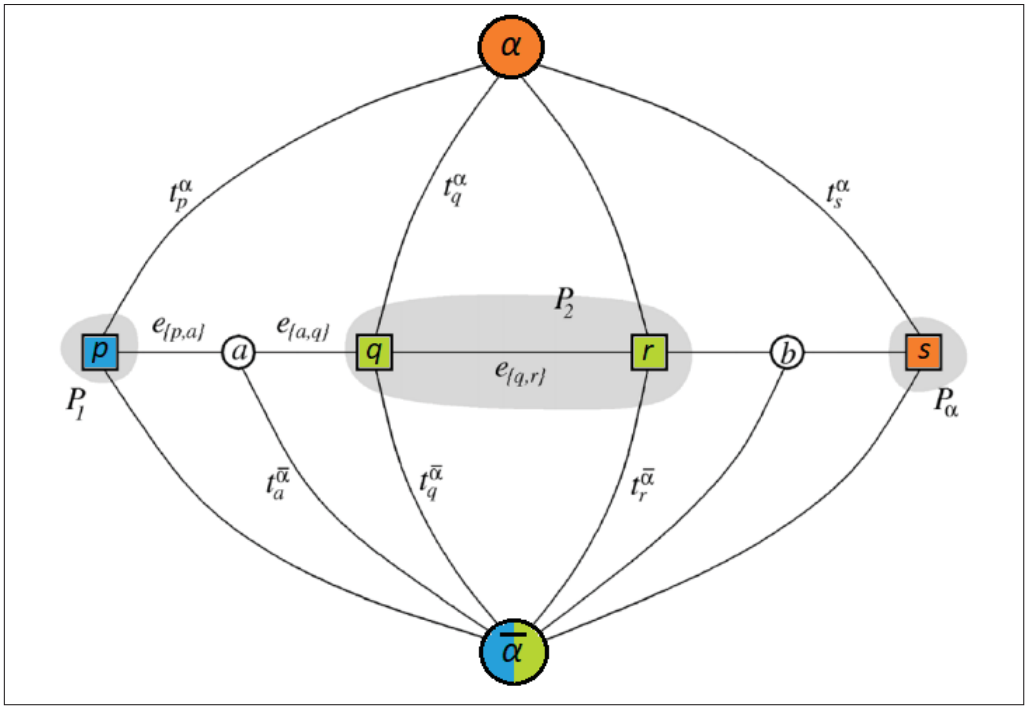

Figure 3.4 Illustration du graphe α-expansion pour une image 1D. L'ensemble des pixels de l'image est $P = \{p,q,r,s\}$ et la partition courante est $P = \{P_1, P_2, P_\alpha\}$ où $P_1 = \{p\}$ (bleu), $P_2 = \{q,r\}$ (vert) et $P_\alpha = \{s\}$ (orange). Deux nœuds auxiliaires a et b sont introduits entre les pixels voisins qui possèdent des étiquettes différentes. Figure reproduite et adaptée avec l'autorisation de (Boykov *et al.*, 2001).

Dans l'algorithme 3.1, une seule exécution des étapes 5 à 8 est appelée une itération et une exécution des étapes 3 à 12 est un cycle. À chaque cycle, l'algorithme effectue une itération pour

Algorithme 3.1 α-expansion

Input : *Commencez par un étiquetage arbitraire S*
Output : *étiquetage optimal* S^*
étiquette changée = faux
for *tous* $\alpha \in S$ **do**
    Calculer le $\alpha-$expansion $\widehat{S}$ avec la plus faible énergie $E(\widehat{S})$ via des coupes de graphe
    if $E(\widehat{S}) \leq E(S)$ **then**
        $S = \widehat{S}$
        étiquette changée = vrai
    end
end
if *étiquette changée* **then**
    goto 3
end

chaque étiquette. Un cycle réussit si un étiquetage strictement meilleur est trouvé à n'importe quelle itération. L'algorithme s'arrête si aucune autre amélioration n'est possible.

3.7 Post-traitement

Bien que les segmentations obtenues avec notre approche soient généralement lisses et proches des annotations manuelles, de petites régions isolées peuvent parfois apparaître dans les résultats de segmentation finaux. Les algorithmes de régularisation utilisées dans les expériences binaires et multi-classes peuvent éliminer ces régions. Dans certains cas, plus que les régions parasites, les résultats peuvent montrer une vertèbre ou un disque supplémentaire. Dans notre étude, nous visons à segmenter seulement sept régions de chaque structure. La coupe de graphe ne peut qu'améliorer les segmentations et ne peut pas supprimer ces régions supplémentaires. En tant qu'étape de post-traitement, nous supprimons ces régions en ne conservant que les sept composants connectés les plus grands de chaque structure.

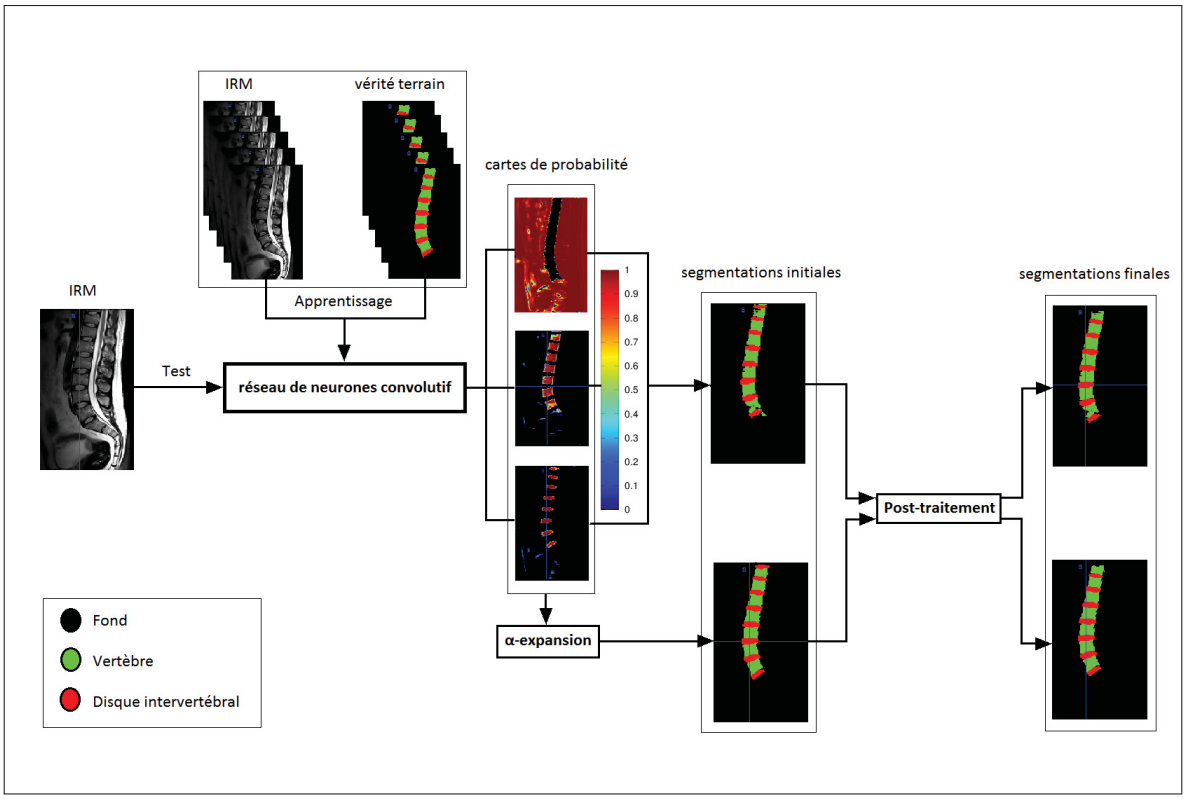

Figure 3.5 Organigramme de la méthode proposée

3.8 Conclusion

Dans ce chapitre, nous avons présenté les étapes suivies pour la segmentation des vertèbres et des disques de la colonne vertébrale à partir d'images IRM. La méthode proposée est basée sur la combinaison d'un réseau de neurones convolutif avec une régularisation par la coupe de graphe. Tout d'abord, nous avons présenté les détails de l'architecture du réseau de neurones proposée. Nous avons décrit ensuite les algorithmes de régularisation que nous allons utilisés pour améliorer encore les résultats de la segmentation fournis par notre réseau de neurones. Nous avons commencé par décrire l'algorithme de coupe de graphe, qui sera utilisé pour la segmentation binaire et puis l'algorithme α-expansion pour la segmentation multi-classes. Presque toutes les approches mentionnés dans la littérature, qui utilisent ces algorithmes de régularisation, nécessitent l'initialisation manuelle de la part d'un utilisateur. Par contre, dans

notre approche, nous allons utiliser les cartes de probabilité de segmentation générées par notre réseau de neurones comme initialisation.

CHAPITRE 4

SIMULATIONS ET RÉSULTATS

4.1 Introduction

Le chapitre précédent a présenté les éléments qui constituent l'approche de la segmentation automatique de la colonne lombaire développée au cours de cette étude. Dans ce chapitre, nous abordons d'abord les détails de la mise en oeuvre du réseau. Deuxièmement, nous décrivons les paramètres d'évaluation utilisés pour évaluer les résultats de la segmentation. Pour démontrer l'efficacité de notre approche, nous présenterons dans la suite les résultats expérimentaux de trois bases de données d'images IRM 3D différentes. La première base de données contient la segmentation manuelle des vertèbres. La seconde contient les annotations manuelles des disques intervertébraux. Enfin, nous évaluerons notre méthode sur une nouvelle base de données que nous avons construite et qui contient les annotations manuelles des deux structures : les vertèbres et les disques. Pour les évaluations quantitatives, nous utiliserons le coefficient de similarité de Dice et la distance de Hausdorff entre nos résultats de segmentation et les annotations de référence. Nous présenterons également une évaluation qualitative et une comparaison entre nos résultats et ceux de méthodes populaires partageant leurs résultats expérimentaux sur les mêmes bases de données.

4.2 Détails d'implémentation

La méthode de segmentation proposée dans notre étude est basée sur une architecture de réseau de neurones entièrement convolutif à deux dimensions. Cette architecture est composée de 13 couches au total avec la disposition suivante : 9 couches de convolution, 3 couches entièrement connectées et la couche de classification à la fin (figure 3.1). Pour les trois premières couches, nous avons utilisé 25 filtres de taille 9×9 dans chaque couche. Dans les trois couches de convolution suivantes, nous avons utilisé 50 filtres de taille 7×7 dans chaque couche. Et enfin, pour

les trois dernières couches de convolution, 75 noyaux de taille 5×5 sont utilisés dans chaque couche.

Pour préserver la résolution spatiale, nous avons utilisé un pas unitaire pour toutes les opérations de convolution et nous avons évité l'utilisation de couches de sous-échantillonnage. En tant que fonction d'activation, nous avons utilisé l'unité linéaire rectifiée paramétrique (PReLU), qui fonctionne mieux que l'unité linéaire rectifiée (ReLU) standard.

Les trois couches entièrement connectées sont composées de 400, 200 et 150 unités cachées, respectivement. Ces couches sont suivies d'une couche de classification finale, qui produit les cartes de probabilité pour chacune des 3 classes : les vertèbres, les disques intervertébraux et l'arrière-plan.

Au lieu d'utiliser l'image 3D entier comme entrée, notre réseau extrait de l'image originale des segments bidimensionnels de petite taille et les utilise comme entrée. Cette stratégie réduit les besoins en mémoire et augmente considérablement le nombre d'échantillons d'apprentissage.

L'optimisation des paramètres réseau est effectuée via l'optimiseur RMSprop en utilisant l'entropie croisée comme fonction de coût. Pour initialiser les poids de notre réseau, nous avons utilisé une distribution gaussienne à moyenne nulle d'un écart type de $\sqrt{2/n_c^l}$, où n_c^l indique le nombre de connexions aux unités de la couche l. Nous avons fixé l'élan à 0,6 et le taux d'apprentissage initial à 0,001. Le réseau est régularisé en utilisant un décrochage avec un taux de 50% sur les couches entièrement connectées.

L'un des principaux paramètres de la mise en œuvre d'un réseau de neurones est le nombre d'époques. Le seul moyen de connaître le nombre optimal consiste à tracer la précision d'apprentissage du réseau en fonction du nombre d'époques. La figure 4.1 montre que la précision pendant l'apprentissage du réseau sur les données d'apprentissage et de validation commence à saturer après 35 époques. Nous avons donc choisi, pour toutes les expériences suivantes, d'arrêter le processus d'apprentissage à ce point. Chaque époque est composée de 20 sous-époques.

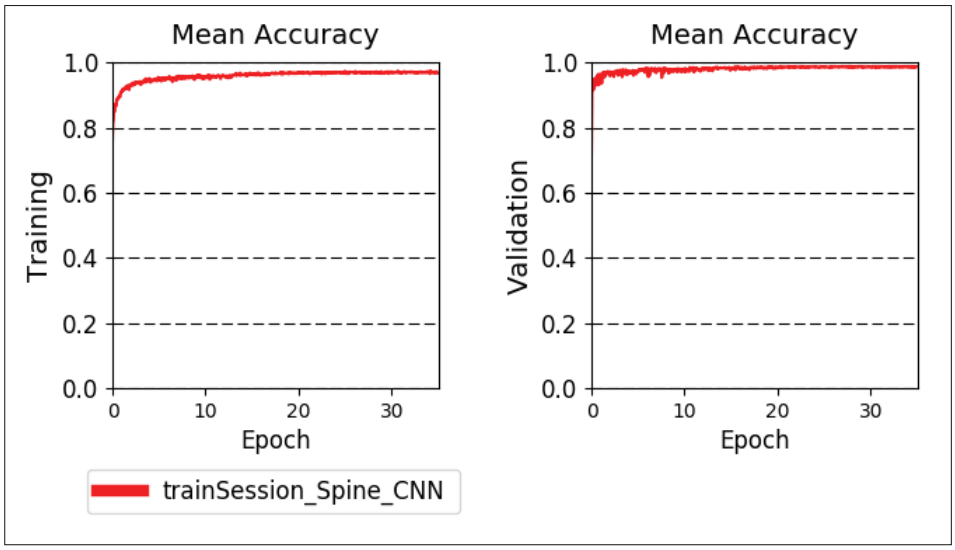

Figure 4.1 La précision d'apprentissage et de la validation de notre réseau neuronal en fonction du nombre d'époques

À chaque sous-époque, un total de 1000 échantillons sont sélectionnés de manière aléatoire à partir des images d'apprentissage et traités par lots de taille 10.

Pour mettre en œuvre notre réseau, nous avons adapté l'architecture du RNC développé par (Kamnitsas *et al.* (2017)) en utilisant la bibliothèque Theano. Le modèle a été formé et testé sur une unité de traitement graphique GPU NVIDIA Tesla P-100. Les algorithmes de la coupe de graphe et α-expansion ont été développés sur Matlab. L'ordinateur utilisé pour exécuter ces algorithmes est un processeur Intel® Core i7-4790 à 3,6 GHz.

L'apprentissage de notre réseau sur la base de données des vertèbres prend environ 20 minutes par époque et environ 12 heures au total. La segmentation d'un sujet nécessite environ 30 secondes. La mise en œuvre de la coupe de graphe prend environ une minute pour chaque sujet. L'apprentissage sur la base de données de disques intervertébraux prend environ 17 minutes par époque. La segmentation finale prend environ une minute par sujet. Finalement, l'apprentissage sur la base de données multi-classes, contenant les deux structures, prend environ 45 minutes par époque et environ un jour au total. La segmentation finale prend environ une minute par sujet.

4.3 Métriques d'évaluation

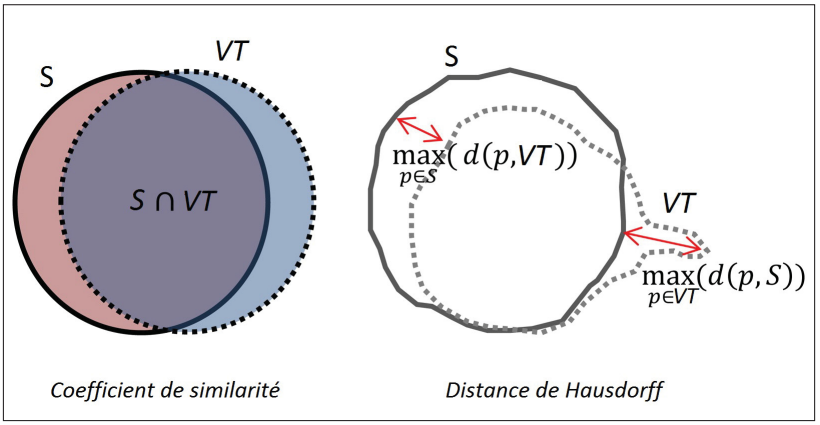

Figure 4.2 Illustration du coefficient de similarité et de la distance de Hausdorff.

Les segmentations obtenus par notre méthode sont comparées aux segmentations manuelles effectuées par un expert en radiologie médicale. Pour évaluer ces résultats, nous avons utilisé le coefficient de similarité de Dice (CSD) et la distance de Hausdorff (HD).

4.3.1 Coefficient de similarité de Dice

Le coefficient de similarité de Dice (CSD) mesure le chevauchement entre les régions du résultat de segmentation de la méthode proposée et la vérité terrain d'une image qui est utilisée comme référence. Il est calculé par l'équation suivante :

$$CSD = \frac{2 \times |S \cap VT|}{|S| \cup |VT|} \tag{4.1}$$

où S représente le résultat de la segmentation et VT la vérité de terrain générée manuellement par les radiologues. La valeur CSD varie de 0 à 1. Un CS plus grand signifie une meilleure segmentation (Zijdenbos *et al.* (1994)).

4.3.2 Distance de Hausdorff

La distance de Hausdorff (HD) est utilisée pour indiquer si deux formes sont identiques et, si elles sont différentes, elle fournit une mesure de distance symétrique de l'écart maximal entre deux contours étiquetés qui sert à quantifier ces différences. La distance de Hausdorff est la plus grande distance minimale entre la segmentation S et la vérité terrain VT. La distance minimale est la norme euclidienne entre un pixel p qui appartient au premier contour et le point le plus proche du second contour. Autrement dit, pour chaque point de S, il faut trouver la distance minimale à VT, et vice versa.

$$HD(S,VT) = max(max_{p\in S}\, d(p,VT), max_{p\in VT}\, d(p,S)) \tag{4.2}$$

où,

$$d(p,S) = min_{p'\in S}||p - p'|| \tag{4.3}$$

4.4 Résultats de la segmentation des vertèbres

Pour l'évaluation quantitative de la présente méthode sur les 23 images IRM, le coefficient de similarité et la distance de Hausdorff entre la segmentation automatique et la segmentation de la vérité terrain sont calculés sur des volumes 3D.

4.4.1 Bases de données

La base de données, accessible au public [1], est constituée de 23 images IRM 3D spin-écho pondérées en T2 de façon sagittale provenant de 23 sujets différents. Les images ont été acquises à l'aide d'un scanner Siemens Tesla 1.5 et ré-échantillonnées à une taille de voxel de 2 $\times$ 1,25 $\times$ 1,25 mm^3. Toutes les images sont échantillonnées pour avoir les mêmes tailles de 39 $\times$ 305 $\times$ 305 voxels. Chaque image contient au moins 7 corps vertébraux de la colonne vertébrale inférieure (T11 - L5). Pour chaque corps vertébral, 161 vertèbres au total, une segmentation

1. https ://zenodo.org/record/22304

manuelle de référence est fournie sous la forme d'un masque binaire. Toutes les images et tous les masques binaires sont stockés dans le format de fichier NIFTI (Neuroimaging Informatics Technology Initiative) (Chu *et al.* (2015b)).

4.4.2 Les résultats de la segmentation du RNC

La performance de segmentation des corps vertébraux a été évaluée en utilisant 22 sujets pour l'apprentissage et un sujet pour le test, de sorte que le réseau a été formé 23 fois et à chaque fois l'évaluation est faite sur le sujet restant. Les masques de segmentation des vertèbres obtenus ont été comparés aux annotations de référence en se basant sur deux facteurs : le coefficient de similarité de Dice (CSD) et la distance de surface de Hausdorff (HD). Pour la segmentation des vertèbres, un CSD moyen de 94,65$\pm$ 1,08% et une HD de 2.1 $\pm$ 0.37 mm (moyenne $\pm$ Std) ont été obtenus. Les résultats détaillés sont présentés dans le tableau 4.1 pour tous les sujets. De plus, nous avons présenté un tableau récapitulatif 4.4 des résultats obtenus par nos méthodes ainsi que des résultats obtenus par d'autres études populaires qui ont été évaluées sur la même base de données.

Tableau 4.1 Résultats de la segmentation obtenus par notre RNC sur 23 images MR 3D. L'apprentissage a été effectué sur 22 sujets en laissant chaque fois un sujet pour le test.

Sujets	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**	**#11**	**#12**
CSD (%)	95.74	95.13	95.61	91.66	95.06	94.91	95.31	94.28	95.86	95.39	93.55	94.68
HD (mm)	2.00	2.00	2.00	3.75	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
#13	**#14**	**#15**	**#16**	**#17**	**#18**	**#19**	**#20**	**#21**	**#22**	**#23**	**Moyenne $\pm$ Std**	
93.39	95.18	95.45	92.06	94.96	95.06	94.42	94.51	94.52	94.87	95.42	**94.65 $\pm$ 1.08**	
2.00	2.00	2.00	2.50	2.00	2.00	2.00	2.00	2.00	2.00	2.00	**2.10 $\pm$ 0.37**	

Une de ces études a été présentée par (Chu *et al.* (2015a)). Les auteurs ont utilisé une méthode de régression et de classification unifiée des forêts aléatoires basée sur l'apprentissage pour la localisation et la segmentation des vertèbres. Ils ont utilisé la même validation croisée que nous. Ils ont formé le réseau 23 fois en utilisant 22 sujets pour l'apprentissage et le sujet restant pour le test. Le temps de calcul moyen pour segmenter une image était d'environ 1,3 minute,

tandis que pour notre étude, il ne faut que 30 secondes pour segmenter un sujet. De plus, la moyenne du coefficient de similarité de segmentation rapportée par leur étude était de 88,72%, ce qui est inférieur au résultat obtenu par notre réseau de 5,93%.

Les distances de Hausdorff obtenues par (Chu *et al.* (2015a)) varient de 4,3 à 8,3 *mm*. Cela signifie que la distance entre la surface de l'annotation manuelle et la surface du résultat de la segmentation est toujours supérieure à 3 pixels pour chaque sujet. Dans notre étude, les distances de Hausdorff obtenues varient de 2 à 3,75 *mm*. Cela signifie que dans le cas le plus complexe, nous avons obtenu une différence de 3 pixels. De plus, une distance de Hausdorff de 2 *mm*, qui représente une différence d'un pixel, est obtenue pour 21 sujets.

Une autre étude a été proposée par (Korez *et al.* (2016)) pour la segmentation automatique des vertèbres. Leur méthode consiste à utiliser un réseau de neurones convolutif 3D pour générer des cartes de probabilité pour les corps vertébraux. Ces cartes sont ensuite intégrées à des modèles déformables pour améliorer la précision et la robustesse de la segmentation primaire. Cette étude a été évaluée sur la même base de données que celle utilisée par (Chu *et al.* (2015a)). Pour l'apprentissage de leur réseau, la base de données a été divisée en 11 images d'apprentissage et 12 images de test. Pour s'assurer de tester sur tous les sujets, cette validation croisée a été répétée cinq fois. Les détails de la validation, le temps de calcul et le coefficient de similarité par sujet n'étaient pas rapportés dans cette étude. Korez *et al.* (2016) ont signalé un coefficient de similarité de 93,4%, que nous avons dépassé par 1,25%, en ne prenant en compte que les résultats de segmentation primaires obtenus par notre réseau. Pour examiner également les performances de notre réseau par rapport à cette étude, nous avons suivi une validation croisée similaire en prenant la moitié des sujets pour l'apprentissage et le reste des sujets pour le test. Après avoir effectué cette validation croisée trois fois, nous avons obtenu un coefficient de similarité global de 94,01%.

4.4.3 Les résultats de la segmentation par la coupe de graphe

Pour choisir les meilleurs termes de régularisation λ et σ, nous avons effectué la coupe de graphe sur tous les différents sujets en utilisant différentes valeurs de λ allant de 0.1 à 0.9 avec une incrémentation de 0.1, et différentes valeurs de σ allant de 100 à 1000 avec une incrémentation de 100.

Tableau 4.2 Les CSD des résultats de segmentation obtenus par l'algorithme de coupe de graphe en utilisant différentes paires de termes de régularisation.

Sujet 14											
λ \ σ	**10**	**100**	**200**	**300**	**400**	**500**	**600**	**700**	**800**	**900**	**1000**
0.1	95.21	95.28	95.32	95.34	95.38	95.39	95.40	95.41	95.41	95.34	95.43
0.2	95.17	95.29	95.22	95.34	95.37	95.41	95.42	95.46	95.47	95.47	95.48
0.3	94.90	94.80	95.22	95.42	95.30	95.33	95.32	95.35	95.36	95.41	95.42
0.4	94.58	94.80	95.40	95.11	95.14	95.21	95.23	95.27	95.29	95.30	95.34
0.5	94.30	94.50	94.71	94.86	94.96	95.01	95.04	95.11	95.16	95.18	95.22
0.6	93.84	94.24	94.52	94.63	94.81	94.87	94.90	94.98	95.02	95.05	95.08
0.7	93.84	95.01	94.36	94.49	94.63	94.73	94.90	94.83	95.02	94.92	94.96
0.8	94.09	93.87	94.14	94.28	94.42	94.58	94.60	94.71	94.74	94.79	94.83
0.9	94.71	93.73	93.93	94.07	94.27	94.43	94.50	94.51	94.58	94.65	94.68
Sujet 15											
0.1	95.45	95.59	95.64	95.68	95.71	95.71	95.72	95.72	95.71	95.72	95.73
0.2	95.41	95.63	95.75	95.79	95.80	95.83	95.84	95.82	95.84	95.84	95.83
0.3	95.27	95.65	95.78	95.80	95.89	95.87	95.87	95.85	95.86	95.87	95.88
0.4	95.03	95.65	95.78	95.80	95.89	95.90	95.90	95.91	95.87	95.86	95.84
0.5	94.86	95.43	95.64	95.73	95.79	95.82	95.81	95.85	95.86	95.87	95.88
0.6	94.06	95.31	95.55	95.66	95.70	95.75	95.79	95.79	95.80	95.80	95.81
0.7	94.06	94.87	95.86	95.56	95.40	95.67	95.70	95.54	95.73	95.74	95.74
0.8	94.06	95.00	95.55	95.45	95.54	95.59	95.61	95.68	95.67	95.67	95.67
0.9	94.06	94.87	95.25	95.56	95.40	95.54	95.56	95.54	95.56	95.58	95.58
Sujet 16											
0.1	93.27	92.48	92.40	92.33	92.30	92.27	92.24	92.21	92.21	92.19	92.20
0.2	92.97	92.71	92.56	92.42	92.35	92.28	92.21	92.21	92.17	92.14	92.14
0.3	93.16	92.81	92.58	92.41	92.30	92.23	92.17	92.10	92.07	92.07	92.05
0.4	93.27	92.75	92.47	92.29	92.17	92.02	91.92	91.90	91.87	91.87	91.84
0.5	93.33	92.68	92.38	92.21	92.09	91.96	91.88	91.83	91.76	91.74	91.69
0.6	93.37	92.64	92.29	92.08	91.97	91.79	91.76	91.71	91.67	91.57	91.50
0.7	93.35	92.53	92.22	91.96	91.87	91.81	91.68	91.55	91.47	91.45	91.38
0.8	93.27	92.47	92.11	91.95	91.77	91.71	91.56	91.43	91.34	91.34	91.23
0.9	93.13	92.28	91.96	91.84	91.55	91.46	91.30	91.16	91.21	91.15	91.17

Le tableau 4.2 montre les coefficients de similarité de Dice des résultats de segmentation des sujets 14, 15 et 16, obtenus par des différentes paires de termes de régularisation (λ , σ). Ce tableau montre clairement que le meilleur coefficient de similarité, qui indique le résultat optimal de segmentation, dépend du sujet étudié et diffère d'un sujet à l'autre. Les cases rouges dans le tableau indiquent les meilleurs CSD pour chaque sujet. Prenons l'exemple du sujet 14 sur lequel le meilleur résultat est obtenu en prenant $\lambda = 0.2$ et $\sigma = 1000$ comme paramètres pour l'algorithme de la coupe de graphe. Pour le sujet 16, la segmentation la plus précise est obtenue avec $\lambda = 0.6$ et $\sigma = 10$.

Dans notre étude, les termes de régularisation utilisés pour la coupe de graphe ne doivent pas dépendre uniquement du sujet étudié. Ces paramètres doivent rester fixes lorsque nous appliquons la coupe de graphe pour tous les sujets et ne doivent pas être modifiés d'un sujet à l'autre. La raison est que, dans tous les défis de segmentation des images médicale, les sujets de test sont toujours conservés uniquement pour la compétition et seuls les sujets d'apprentissage seront disponibles pour les concurrents. Donc, pour choisir la meilleure paire de valeurs λ et σ, nous avons calculé pour chaque combinaison possible de ces paramètres la moyenne des coefficients de similarité des résultats de segmentation de tous les sujets afin d'identifier la paire donnant la moyenne la plus élevée.

Tableau 4.3 Les moyennes des CSD des résultats de segmentation de tous les sujets obtenus par l'algorithme de coupe de graphe en utilisant différentes paires de termes de régularisation. Les cases rouges représentent le meilleur CSD pour chaque sujet.

λ \ σ	**10**	**100**	**200**	**300**	**400**	**500**	**600**	**700**	**800**	**900**	**1000**
0.1	94.66	94.74	94.79	94.78	94.82	94.81	94.81	94.80	94.80	94.81	94.81
0.2	94.65	94.74	94.80	94.83	94.84	94.85	94.84	94.83	94.84	94.84	94.84
0.3	94.58	94.64	94.72	94.78	94.79	94.79	94.79	94.80	94.80	94.82	94.83
0.4	94.37	94.55	94.64	94.67	94.70	94.72	94.74	94.75	94.74	94.76	94.76
0.5	94.11	94.29	94.41	94.49	94.60	94.61	94.63	94.66	94.67	94.67	94.69
0.6	93.80	94.06	94.29	94.42	94.44	94.50	94.53	94.57	94.58	94.58	94.59
0.7	93.69	94.04	94.20	94.32	94.37	94.43	94.47	94.49	94.50	94.51	94.52
0.8	93.43	93.78	94.06	94.19	94.23	94.31	94.36	94.38	94.42	94.42	94.44
0.9	93.26	93.65	93.89	94.06	94.13	94.20	94.25	94.29	94.33	94.34	94.36

Le tableau 4.3 présente les moyennes des coefficients de similarité de Dice, obtenues par l'algorithme de coupe de graphe, en utilisant les différents combinaison des paramètres λ et σ. La moyenne optimale, marquée dans le tableau par la case rouge, a été obtenue en prenant $\lambda = 0,2$ et $\sigma = 500$. Ces deux termes ont été utilisés dans les expériences suivantes pour la segmentation des vertèbres.

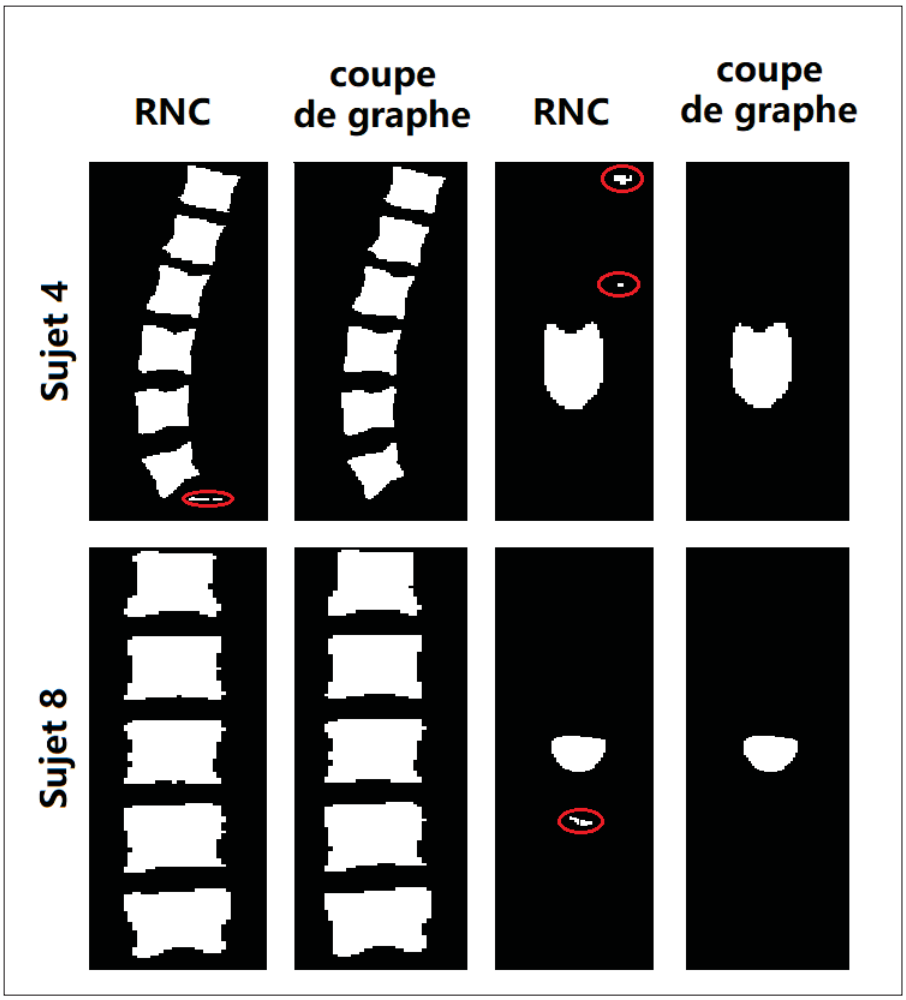

Figure 4.3 L'amélioration de l'utilisation de la coupe de graphe sur les prédictions obtenues par notre RNC sur deux sujets différents. (Première ligne) Résultats de segmentation du sujet 4 sur les vues sagittale et axiale. (Deuxième ligne) Résultats de segmentation du sujet 8 sur les vues coronale et axiale.

Dans la figure 4.3, nous présentons les résultats de la segmentation de deux sujets différents sur des coupes bidimensionnelles dans des vues différentes. Dans la première ligne, nous présentons les résultats de la segmentation sur le sujet 4 dans les vues sagittale et axiale. Ce sujet a une scoliose en haut à gauche qui rend difficile la segmentation de la vertèbre supérieure T11 qui n'est pas en position normale. Le coefficient de similarité de la segmentation obtenu sur ce sujet est le plus bas. La deuxième ligne montre les résultats de la segmentation sur le sujet 8 sur les vues coronale et axiale. Le but de cette figure est de montrer l'amélioration de l'utilisation de la coupe de graphe sur les cartes de probabilité fournies par notre RNC. Dans chaque ligne, la première image de chaque vue montre le résultat de la segmentation obtenu par notre réseau. La seconde image est le résultat de l'application de l'algorithme de coupe de graphe sur les cartes de probabilité fournies par notre RNC. Pour le sujet 4, le coefficient de similarité entre la référence manuelle et la segmentation obtenue par notre RNC est de 90,97%, ce qui est considérablement supérieur au 85,1% obtenu par (Chu *et al.* (2015a)). Après l'application de la coupe de graphe aux cartes de probabilité fournies pour ce sujet, nous avons obtenu un coefficient de similarité de 91,54%.

La plus grande amélioration de l'utilisation de notre algorithme a été obtenue sur le sujet 8. Le coefficient de similarité entre la segmentation de référence manuelle et le résultat de la segmentation obtenu par notre RNC est de 94,05%, ce qui est supérieur au 87,1% obtenu par (Chu *et al.* (2015a)). Après l'utilisation de l'algorithme de coupe de graphe, nous avons obtenu un coefficient de similarité de 94,72%.

La figure montre aussi la capacité de l'algorithme de coupe de graphe à enlever les petites régions qui n'appartiennent pas à la texture des vertèbres (entourées en rouge). La plupart des travaux précédents ont utilisé des opérations morphologiques simples pour éliminer ces régions, ce qui pourrait être plus facile et plus rapide que d'utiliser des algorithmes de minimisation d'une fonction d'énergie. L'utilisation de l'algorithme de coupe de graphe proposée dans notre étude ne vise pas uniquement à éliminer ces petites régions, mais également à rendre les résultats obtenus par notre RNC plus précis et les formes des corps vertébraux plus lisses.

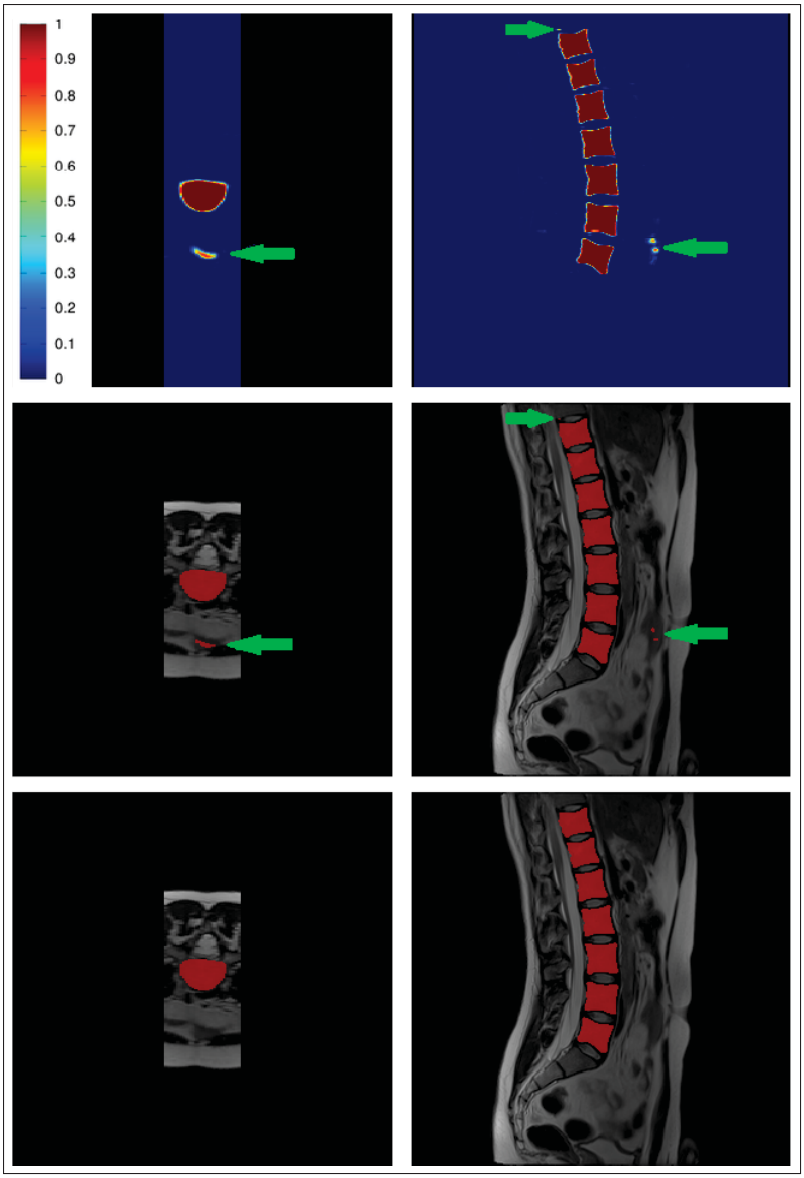

Figure 4.4 Résultats visuels des différentes étapes de notre méthode sur les vues axiales et sagittales. (Première ligne) La carte de probabilités fournie par notre RNC. (Deuxième ligne) Le résultat de la segmentation du RNC. (Troisième ligne) La segmentation finale obtenue par la coupe de graphe en utilisant la carte de probabilités fournie par le RNC comme initialisation. Les flèches vertes dans indiquent les faux pixels prédits en tant que vertèbres. Ces zones ont été supprimées en appliquant la coupe de graphe.

Pour montrer quantitativement cette amélioration par rapport aux opérations morphologiques, nous avons utilisé le même post-traitement que les travaux précédents en supprimant les fausses régions et en ne conservant que les sept plus grandes régions correspondant aux sept vertèbres. Ce post-traitement a pu améliorer le coefficient de similarité global de 94,54% à 94,65%, ce qui représente une augmentation de seulement 0,11%. D'autre part, l'utilisation de notre algorithme de coupe de graphe a conduit à une amélioration de 94.54% à 94,85%. L'amélioration obtenue par notre algorithme varie d'un sujet à l'autre. Elle peut être importante ou insignifiante en fonction du sujet étudié. Prenons comme exemple le sujet 8, le coefficient de similarité après l'élimination des régions indésirables est de 94,28%, ce qui est inférieur au résultat final obtenu par la coupe de graphe de 0,44%. Cependant, pour le sujet 23, le coefficient de similarité n'a augmenté que de 0,01%, ce qui signifie que l'algorithme n'a apporté aucune amélioration.

La figure 4.4 illustre les étapes de notre approche pour la segmentation du sujet 8. Sur la première ligne de la figure, nous montrons les cartes de probabilité, générées par notre réseau, sur les vues axiale et sagittale. Les pixels rouges indiquent des valeurs de probabilité proches de 1, qui font référence au corps des vertèbres. Les pixels bleus, proches de 0, qui font référence à l'arrière-plan. Nous pouvons clairement remarquer la présence de fausses prédictions, qui sont marquées par des flèches vertes sur la figure. Ces mauvaises prédictions ont été segmentées en tant que corps vertébral dans le résultat final généré par notre réseau. Cela illustre la nécessité de rectifier la segmentation du réseau et de supprimer ces régions indésirables. La dernière ligne de la figure montre le résultat de la segmentation après l'application de coupe de graphe en utilisant la carte de probabilité fournie par notre RNC comme initialisation. Nous pouvons remarquer que les régions de bruit ont été supprimées et que seules les sept vertèbres apparaissent sur la segmentation finale.

Le coefficient de similarité global atteint par notre étude, après avoir enlevé les fausses régions des résultats obtenus par notre réseau et en ne conservant que les sept régions vertébrales, est de 94,65%. La suppression des régions fausses dans les résultats de segmentation obtenus par la coupe de graphe améliore le coefficient de similarité global seulement de 0,04%, ce qui prouve l'efficacité de l'algorithme pour fournir une segmentation précise sans régions indésirables.

Les résultats détaillés de la segmentation par sujet, obtenus par notre RNC et par l'algorithme de coupe de graphe avant et après le post-traitement, sont présentés dans le tableau 4.5.

Tableau 4.4 Les moyennes des coefficients de similarité des résultats de la segmentation des vertèbres de nos méthodes et d'autres études populaires obtenues sur la même base de données d'images.

Méthodes	**Type de validation croisée**	**coefficient de similarité global**
Chu *et al.* (2015a)	22 sujets pour l'apprentissage et un pour le test	88.72%
Korez *et al.* (2016)	5 × (11 sujets pour l'apprentissage et 12 pour le test)	93.40%
UNet	22 sujets pour l'apprentissage et un pour le test	87.89%
RNC	3 × (11 sujets pour l'apprentissage et 12 pour le test)	**94.01%**
RNC	22 sujets pour l'apprentissage et un pour le test	**94.65%**
RNC et coupe de graphe	22 sujets pour l'apprentissage et un pour le test	**94.89%**

La moyenne du coefficient de similarité des résultats de la segmentation obtenus par Korez *et al.* (2016) est de 93,4%, ce que nous avons dépassé de 1,14 %, en ne prenant en compte que les résultats de la segmentation obtenus par notre RNC. En appliquant l'algorithme de coupe de graphe aux cartes de probabilité fournies par notre réseau, nous avons obtenu un coefficient de similarité moyen de 94,85%, ce qui est supérieur de 0,31% à notre résultat primaire et de 1,45% au résultat global présenté par (Korez *et al.* (2016)).

Pour comparer les résultats obtenus par notre réseau, nous avons aussi utilisé une version UNet, qui est disponible en ligne [2], et l'avons adapté à la tâche de segmentation des vertèbres. UNet a été initialement publiée pour la segmentation biomédicale et s'est révélé utile à d'autres domaines tel que la segmentation d'images satellite. Ce réseau a la capacité d'apprendre à partir de faibles quantités de données d'apprentissage, ce qui est toujours le cas dans le domaine médical. De nombreux codes de projet, qui ont utilisé l'architecture UNet pour la segmentation, sont disponibles en ligne dans Github. Une grande partie de ces codes de projet concerne la segmentation d'images médicales, mais il existe d'autres projets ayant pour objectif la segmentation d'images RVB et satellite. Le code que nous avons utilisé est publié sur GitHub. Il vient avec une bonne description concernant l'installation et la configuration des logiciels utilisés. L'implémentation du code est générale et il est facile à adapter à nos fins. Nous n'avons

2. https ://github.com/zhixuhao/unet

Tableau 4.5 Les résultats de la segmentation de chaque sujet de la base de données de vertèbres, obtenus par nos différentes méthodes et par le travail de Chu *et al.* (2015a).

Sujets	Chu *et al.* (2015a)	UNet	RNC	RNC post-traité	coupe de graphe	coupe de graphe post-traitée
Sujet 1	88.60	89.79	95.69	95.74	96.06	96.09
Sujet 2	86.90	90.04	94.98	95.13	95.38	95.42
Sujet 3	86.70	90.66	95.55	95.61	96.00	96.00
Sujet 4	85.10	84.19	90.97	91.66	91.54	91.91
Sujet 5	88.40	91.77	94.93	95.06	95.31	95.31
Sujet 6	89.20	89.39	94.82	94.91	94.92	94.95
Sujet 7	88.20	83.06	95.30	95.31	95.90	95.90
Sujet 8	87.10	89.52	94.05	94.28	94.72	94.72
Sujet 9	88.30	90.44	95.81	95.86	96.09	96.09
Sujet 10	89.50	88.24	95.39	95.39	95.59	95.59
Sujet 11	89.50	78.22	93.43	93.55	93.75	93.78
Sujet 12	89.90	87.42	94.64	94.68	94.88	94.88
Sujet 13	86.60	84.69	93.38	93.39	93.76	93.76
Sujet 14	88.50	87.64	95.16	95.18	95.41	95.41
Sujet 15	88.00	87.31	95.38	95.45	95.83	95.86
Sujet 16	87.30	88.23	92.04	92.06	92.28	92.28
Sujet 17	91.80	87.61	94.80	94.96	95.20	95.25
Sujet 18	91.00	89.12	94.95	95.06	95.11	95.14
Sujet 19	87.30	84.33	94.03	94.42	94.28	94.51
Sujet 20	91.20	90.09	94.41	94.51	94.66	94.66
Sujet 21	87.90	89.37	94.47	94.52	94.37	94.37
Sujet 22	91.90	90.65	94.84	94.87	95.05	95.05
Sujet 23	91.70	89.72	95.42	95.42	95.43	95.43
Moyenne (%)	**88.72 $\pm$ 1.82**	**87.89 $\pm$ 3.07**	**94.54 $\pm$ 1.14**	**94.65 $\pm$ 1.05**	**94.85 $\pm$ 1.12**	**94.89 $\pm$ 1.07**

modifié que quelques hyperparamètres et certaines configurations pour l'adapter aux images que nous allons utiliser dans notre projet.

Le code UNet que nous avons utilisé prend comme entrée des images carrées bidimensionnelles de taille 128×128 avec une plage d'intensité de pixel allant de 0 à 255. Dans la base de données de vertèbres, la plage d'intensité de pixel varie de -251 à 1500 et la taille spatiale des images est de 39×305. Donc, pour utiliser ce code, les images doivent être normalisées et redimensionnées. Nous avons d'abord normalisé les images afin que les intensités de pixels soient comprises entre 0 et 255, puis nous avons divisé le volume 3D de chaque sujet en images 2D. Pour redimensionner les images, nous avons coupé la région contenant les vertèbres et centré chaque tranche dans une matrice nulle de taille 128×128. Vu que l'UNet que nous avons utilisé génère un masque 2D de la même taille que l'image d'entrée, nous avons concaténé pour chaque sujet toutes les images de sortie afin d'obtenir le résultat final de la segmentation en 3D.

Le code a été développé avec Keras. L'apprentissage a été exécuté sur 35 époques en utilisant une entropie croisée binaire comme fonction de perte. Les sorties du réseau sont des probabi-

lités sigmoïdes sur une carte de taille 128×128. La sortie finale est convertie en une image binaire étiquetée à un seuil de 0,5. Le réseau a été appris 23 fois, à chaque fois nous avons utilisé 22 sujets d'apprentissage en laissant un seul sujet pour le test.

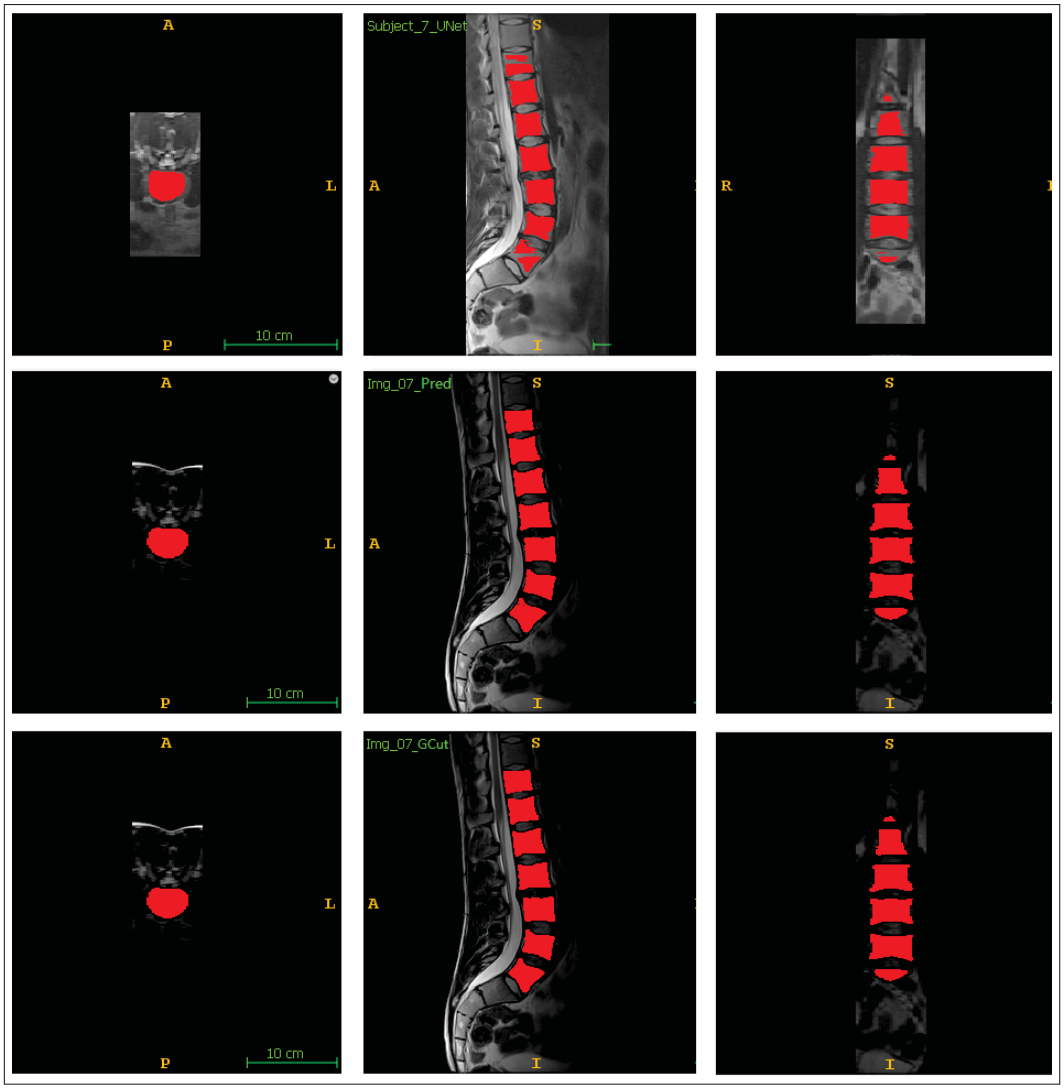

Figure 4.5 Exemple de résultats de segmentation de vertèbres pour le sujet 7 selon différentes méthodes. (Première ligne) Résultat UNet. (Deuxième ligne) Résultat de notre RNC. (Dernière ligne) Résultats de la coupe du graphique en utilisant comme initialisation les cartes de probabilité générées par notre RNC.

En terme de durée, le code UNet était plus rapide que notre RNC. Le réseau a pris quelques heures pour l'apprentissage et quelques secondes pour le test. Qualitativement, les résultats de

segmentation sont acceptables et montrent clairement la forme des corps vertébraux et l'anatomie de la colonne. Quantitativement, la moyenne des coefficients de similarité des segmentations obtenues par UNet est de 87,89%, ce qui est inférieur à notre résultat de 94,89% avec une différence drastique de 7%. Les résultats de segmentation pour chaque sujet sont rapportés dans le tableau 4.5.

La figure 4.5 illustre les résultats visuels de la segmentation du sujet 7 sur les trois vues spatiales selon différentes méthodes. La première ligne de la figure montre le résultat de la segmentation obtenu par l'architecture UNet. Ce réseau était capable d'identifier la forme de la colonne vertébrale, mais le résultat n'était pas précis et certaines vertèbres n'étaient pas complètement segmentées (les vertèbres T11 et L5 dans ce cas). Un coefficient de similarité de 83,06% a été obtenu pour ce sujet. La deuxième ligne de la figure montre le résultat de la segmentation obtenu par notre RNC. Les vertèbres sont bien identifiées et la segmentation ne contient aucune fausse prédiction. Le coefficient de similarité obtenu par cette segmentation est de 95,31%, ce qui représente une amélioration de 12,25% par rapport à UNet. L'algorithme de coupe de graphe améliore également le résultat de la segmentation primaire obtenue par notre réseau, et le coefficient de similarité final obtenu est de 95,9%.

4.5 Résultats de la segmentation des disques intervertébraux

4.5.1 Base de données

La base de données du défi MICCAI-2015 est constituée d'images IRM tridimensionnelles spin-écho pondérées en T2 de 15 patients différents, chacun contenant 7 disques intervertébraux de la colonne vertébrale inférieure (T11 - L5). Chaque patient a été scanné avec un scanner IRM 1,5 Tesla de Siemens (Siemens Magnetom Sonata, Siemens Healthcare, Erlangen, Allemagne). Pour chaque disque, une segmentation manuelle de référence est fournie sous la forme d'un masque binaire. Toutes les images et tous les masques binaires sont stockés dans le format de fichier NIFTI.

4.5.2 Les résultats de la segmentation du RNC

Notre méthode a été évaluée sur une base de données accessible au public qui a été transmise aux concurrents du défi de segmentation des disques intervertébraux MICCAI-CSI-2015 en tant que base de données d'apprentissage. Les méthodes ayant participé à la compétition ont été évaluées sur une nouvelle base de données composée de 10 sujets. Comme nous ne disposons pas de cette base de données de test, nous avons évalué les performances de notre réseau en nous basant uniquement sur la base de données d'apprentissage. Nous avons utilisé 14 sujets comme données d'apprentissage pour notre réseau, est ensuite validé sur le sujet restant. Nous avons répété cette validation croisée 15 fois pour segmenter à chaque fois un sujet. Pour la segmentation des disques, un CSD moyen de 91.56$\pm$ 2.25% et une distance HD moyenne de 3.64 $\pm$ 3.03 *mm* (moyenne $\pm$ Std) ont été obtenus. Les coefficients de similarité de Dice des résultats de segmentation de tous les sujets sont présentés dans le tableau 4.6.

Tableau 4.6 Résultats de la segmentation des disques obtenus par notre RNC sur les 15 sujets. À chaque fois, 14 sujets sont utilisés pour l'apprentissage et un sujet pour le test.

Sujets	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**
CSD (%)	90.91	92.51	92.71	85.77	93.10	92.62	87.76	92.13
HD (mm)	2.80	2.36	2.36	12.56	2.36	2.50	9.97	2.36
#9	**#10**	**#11**	**#12**	**#13**	**#14**	**#15**	**Moyenne $\pm$ Std**	
93.63	94.46	91.04	93.00	90.35	91.58	91.90	**91.56 $\pm$ 2.25**	
2.36	2.00	2.50	2.50	2.80	2.67	2.50	**3.64 $\pm$ 3.03**	

Plus que l'évaluation sur la base de données de test, 6 équipes qui ont participé à la compétition MICCAI-2015, ont également évalué leur approche sur la base de données d'apprentissage et publié les résultats de segmentations obtenus dans leurs articles. La même validation croisée a été utilisée par toutes ces méthodes, ce qui nous permet de comparer équitablement nos résultats de segmentation avec leurs résultats.

Les coefficients de similarité moyens, obtenus par les différentes équipes, varient de 88% à 92,5%, ce qui est acceptable pour la pratique clinique (Zheng *et al.* (2017)).

Korez *et al.* (2015a) ont présenté une approche de segmentation utilisant un modèle déformable basé sur un descripteur de contexte de similarité. Ils ont obtenu les meilleurs résultats de segmentation sur les deux bases de données de test du défi MICCAI-2015. Ils ont également obtenu le résultat de segmentation le plus élevé sur les données d'apprentissage avec un coefficient de similarité global de 92,52%. Wang & Forsberg (2015) ont obtenu le deuxième meilleur résultat de segmentation sur les images d'apprentissage avec un CSD moyen de 91%. La CSD moyen obtenu par notre RNC est inférieure au premier meilleur résultat par 0,96% et supérieure au deuxième résultat par 0,56%. Sur la base des résultats finaux du concours, la deuxième place a été remportée par le travail proposé par (Chu *et al.* (2015c)). Cette méthode a également été évaluée sur la base de données d'apprentissage et a obtenu un CSD global de 88%, ce qui est nettement inférieur à notre résultat par 3,56%.

Cette méthode a obtenu les résultats de segmentation les plus faibles sur les sujets 4 et 13, avec un CSD de 85% et 83% respectivement. Dans notre cas, le résultat de segmentation le plus faible a été obtenu sur le sujet 4 avec un CSD de 85,77%. Pour le sujet 13, nous avons obtenu un CSD de 90,35%, ce qui est nettement supérieur à leur résultat par 7,35%. Les résultats de la segmentation des sujets 4 et 13, obtenus par notre réseau, sont illustrés à la figure 4.6. Dans chaque ligne de la figure, nous avons montré les cartes de probabilité générées par notre RNC et le résultat final de la segmentation pour chaque sujet. Nous pouvons clairement remarquer qu'il n'y a pas de probabilité élevée que les pixels du disque (T11-T12) appartiennent à la structure des disques intervertébraux. Cette fausse prédiction a provoqué l'absence de ce disque dans le masque de segmentation final. Ce sujet contient une scoliose gauche supérieure, ce qui rend difficile la détection et la segmentation du disque supérieur. Pour le sujet 13, les sept disques intervertébraux sont détectés et segmentés. Le coefficient de similarité entre notre résultat de segmentation et l'annotation manuelle est de 90,35%. Quantitativement, ce résultat peut être acceptable, mais qualitativement, on peut remarquer facilement le manque de pixels dans la structure de certains disques. La correction de ces erreurs peut alors améliorer la segmentation et la rendre plus précise.

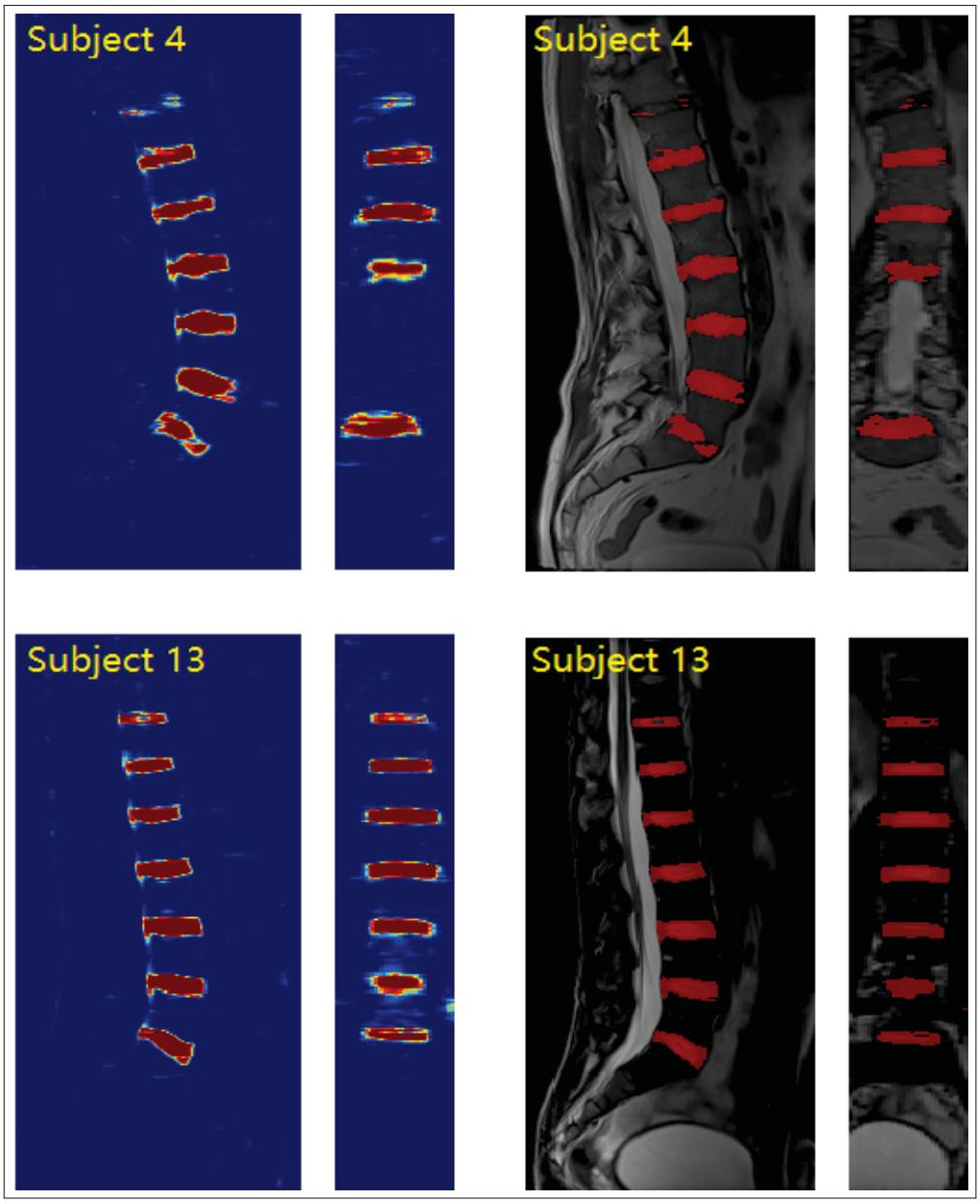

Figure 4.6 Les cartes de probabilité (à gauche) et les résultats de segmentation (à droite) obtenus par notre RNC pour les disques intervertébraux du sujet 4 (première ligne) et du sujet 13 (deuxième ligne), en vues mi-sagittale et mi-coronale.

4.5.3 Les résultats de la segmentation par la coupe de graphe

En suivant la même approche utilisée pour la tâche de segmentation des vertèbres, nous avons appliqué l'algorithme de coupe de graphe pour améliorer les résultats de segmentation pri-

maires des disques intervertébraux en utilisant les cartes de probabilité générées par notre réseau comme initialisation. Pour choisir les paramètres de notre algorithme, nous avons essayé différentes combinaisons de paires de termes de régularisation sur tous les sujets d'apprentissage. La paire de termes qui nous a donné le coefficient de similarité global le plus élevé est ($\lambda = 0,3$ et $\sigma = 500$). Ces termes sont utilisés dans notre algorithme pour les expériences suivantes. Les résultats de la segmentation de chaque sujet, obtenus par l'algorithme de coupe de graphe, sont présentés dans le tableau 4.7.

Tableau 4.7 Résultats de la segmentation des disques obtenus par l'algorithme coupe de graphe sur les 15 sujets.

Sujets	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**
CSD (%)	91.83	93.46	93.26	86.26	93.61	92.87	87.67	92.59
HD (mm)	3.44	2.36	2.36	25.59	2.36	2.50	15.91	2.36
#9	**#10**	**#11**	**#12**	**#13**	**#14**	**#15**	**Moyenne $\pm$ Std**	
93.69	94.89	91.79	93.25	91.24	92.31	92.87	**92.11 $\pm$ 2.21**	
2.36	2.00	2.67	2.50	3.20	2.67	2.50	**4.99 $\pm$ 6.44**	

L'application de l'algorithme de coupe du graphe pour la segmentation des disques intervertébraux en utilisant les cartes de probabilité générées par notre RNC comme initialisation a amélioré le coefficient de similarité moyen par 0,55%. Cette amélioration diffère d'un sujet à l'autre, par exemple pour le sujet 1, le CSD de la segmentation a augmenté par 0,91%, tandis que pour le sujet 9, le CSD n'a progressé que de 0,06%.

Après avoir utilisé l'algorithme de coupe de graphe, le coefficient de similarité le plus faible a été obtenu sur le même sujet 4. L'algorithme a prouvé son efficacité en augmentant le CSD moyen de la segmentation de 0,49%, mais il a également manqué d'identifier et segmenter le disque supérieur. Les six autres disques ont été bien segmentés sans aucun pixel manquant comme dans le résultat précédent, et la forme de chaque disque est devenue lisse (voir figure 4.7).

Pour le sujet 4, la carte de probabilité et la segmentation obtenues par notre RNC, n'ont pu identifier que quelques pixels du disque supérieur (T11-T12). Alors que pour l'algorithme de

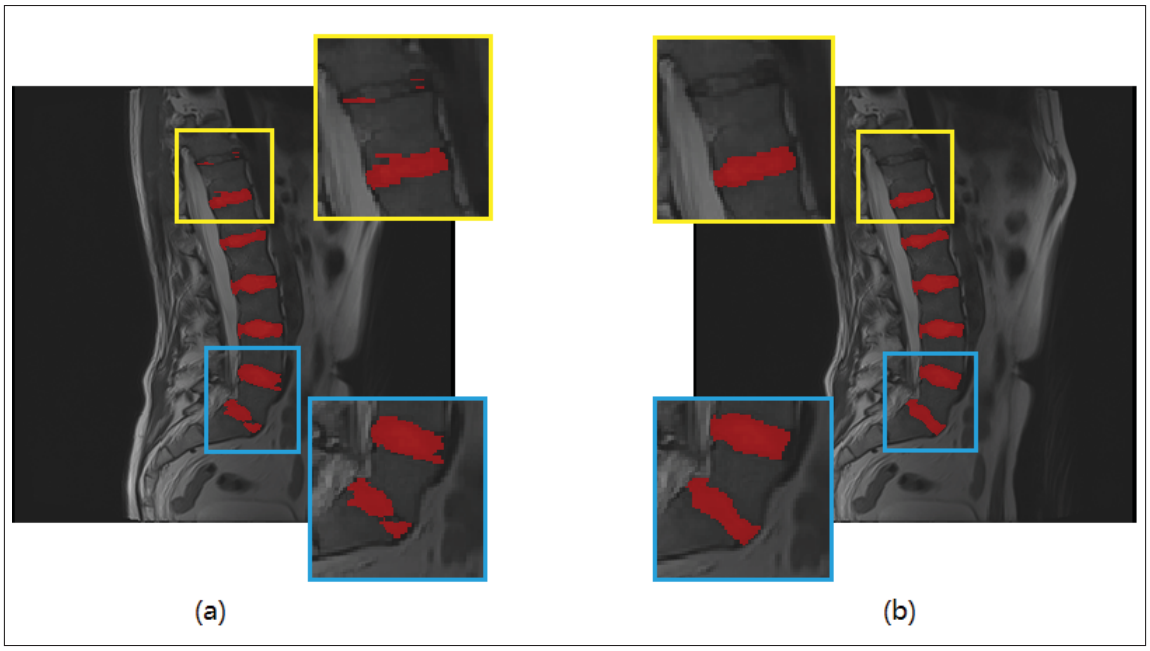

Figure 4.7 Résultats de la segmentation obtenus sur le sujet 4 (vue mi-sagittale). (a) Résultat de la segmentation à l'aide de notre RNC. (b) Résultat de la segmentation en utilisant l'algorithme de coupe de graphe.

coupe de graphe, ces pixels ont été éliminés et le résultat final n'a pas été en mesure d'identifier aucun pixel de ce disque. La raison en est que l'objectif principal de notre algorithme est d'affiner et de corriger le résultat précédent, en rendant la segmentation plus lisse et en complétant les pixels manquants. En d'autres termes, si l'initialisation est si petite comparée à la taille réelle de la région d'intérêt, l'algorithme ignorera ces parties et, au lieu de terminer la segmentation, il supprimera ces petites régions en les traitant comme des pixels de bruit ou des structures d'arrière-plan.

Pour résumer, nous avons pu améliorer nos résultats en utilisant l'algorithme de coupe de graphe. Le coefficient de similarité de Dice global que nous avons atteint sur ces données est de 92,11%, ce qui est inférieur au meilleur résultat par 0,41%. Ce résultat est acceptable si nous le comparons aux résultats obtenus par les autres études et, au meilleur de notre connaissance, seuls nous et l'étude de (Korez *et al.* (2015a)) avons pu atteindre un CSD global supérieure à 92%. Les résultats de segmentation obtenus par nos méthodes, ainsi que ceux obtenus par les équipes ayant participé à la compétition MICCAI-2015, sont résumés dans le tableau 4.8.

Tableau 4.8 Les moyennes des coefficients de similarité des résultats de la segmentation des disques intervertébraux obtenus par nos méthodes ainsi que par d'autres études populaires. Les résultats sont obtenus sur la même base de données en utilisant la même validation croisée.

Méthodes	**Type**	**coefficient de similarité global**
Wang & Forsberg (2015)	enregistrement multi-atlas	91.00 ± 0.01 %
Chu *et al.* (2015c)	méthode basée sur l'apprentissage	88.00 ± 3.70 %
Hutt *et al.* (2015)	CRF à base de super-voxel	90.00 ± 3.00 %
Urschler *et al.* (2015)	forêts de régression et contours actifs	89.10 ± 2.90 %
Korez *et al.* (2015a)	modèle déformable	92.52 ± 2.64 %
Neubert *et al.* (2015)	modèles de forme active	89.60 ± 0.03 %
RNC	réseau de neurones convolutif	$\mathbf{91.56 \pm 2.25}$ %
RNC + Coupe de graphe	réseau de neurones convolutif et coupe de graphe	$\mathbf{92.11 \pm 2.21}$ %

Les organisateurs de la compétition MICCAI-2015 ont publié deux bases de données de test (test1, test2) composées chacune de 5 images IRM 3D. Les annotations manuelles associées à ces bases de données n'ont pas été fournies afin d'évaluer équitablement les méthodes des équipes participantes. Nous avons utilisé nos méthodes pour segmenter les disques intervertébraux sur ces images, mais nous n'avons pas pu évaluer quantitativement nos résultats en raison de l'indisponibilité des annotations manuelles. Pour cette raison, nous allons montrer que des résultats visuels de la segmentation de chaque sujet sur des coupes mi-sagittales.

Pour les expériences sur les données de test, nous avons utilisé les 15 premiers sujets comme base de données d'apprentissage pour notre réseau, puis nous avons validé notre méthode sur les deux bases de données de test. Pour l'algorithme de coupe de graphe, nous avons utilisé les mêmes paramètres que les expériences précédentes. Dans la figure 4.8, nous montrons les résultats de la segmentation obtenus par la coupe de graphe sur les 10 sujets des deux bases de données de test (test1 et test2). La partie à gauche de figure présente les résultats de segmentation obtenus sur les sujets de la base de données test1 sur des coupes mi-sagittales, et la partie a droite présente les résultats de segmentation sur les sujets de la base de données test2. Ces résultats visuels montrent que les sept disques ont été bien détectés et segmentés sur tous les sujets de la base de données de test. Ils montrent également l'efficacité de notre algorithme pour générer une segmentation lisse pour chaque disque.

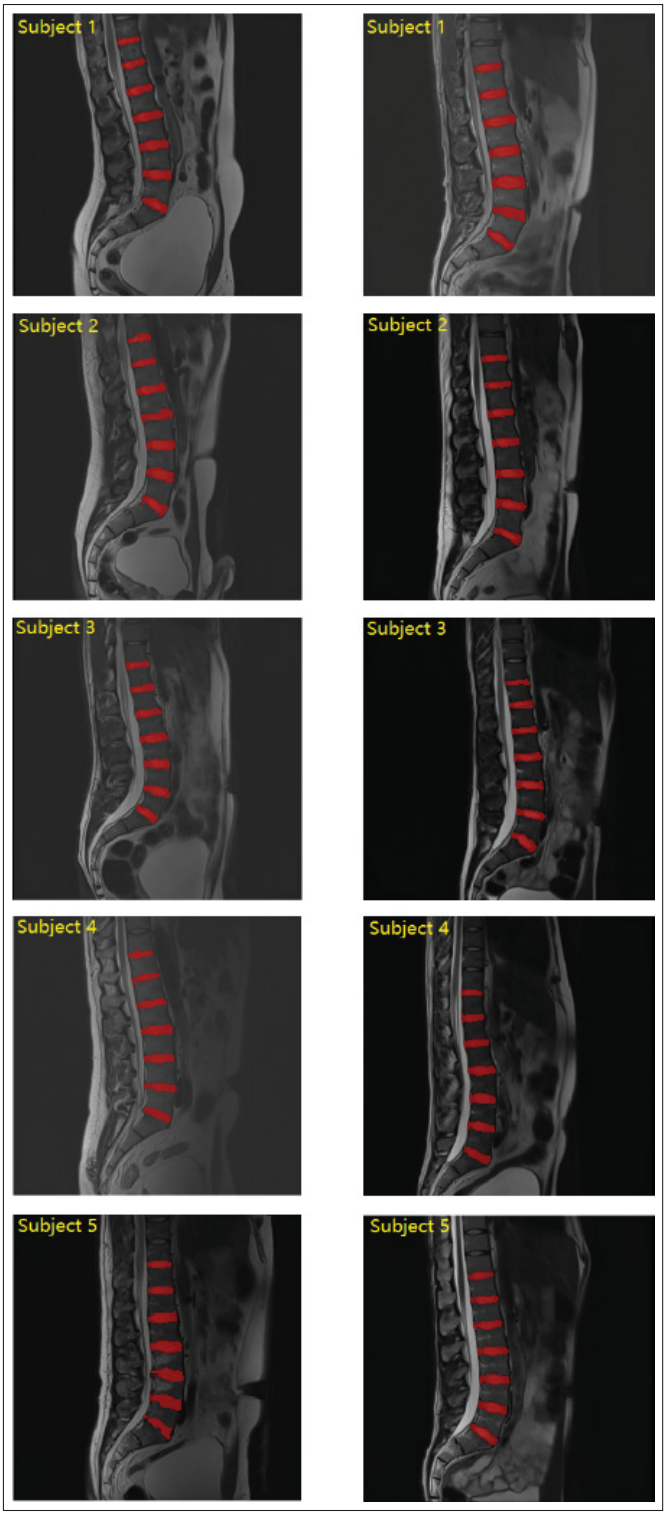

Figure 4.8 Résultats de segmentation par la coupe de graphe obtenus sur les bases de données test1 (à gauche) et test2 (à droite), visualisés sur des coupes mi-sagittales.

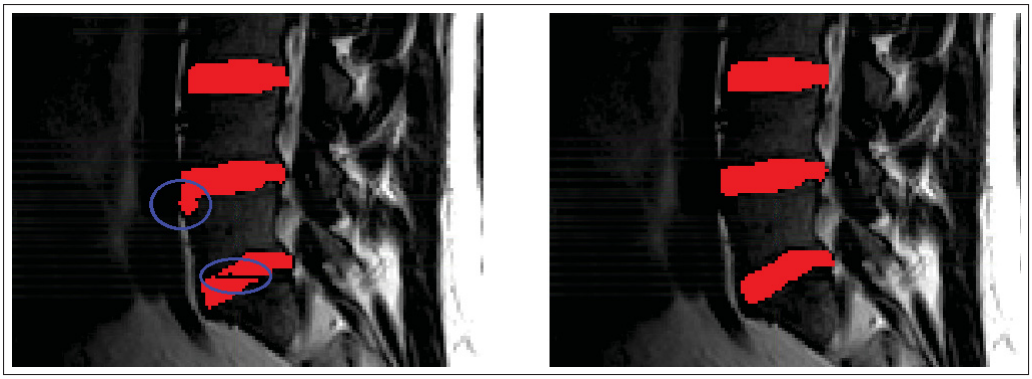

Figure 4.9 Exemple de segmentation des disques lombaire du sujet 4 de la base de données test1. Le résultat à gauche est la segmentation obtenue par notre RNC. Le résultat à droite est la segmentation obtenue par la coupe de graphe.

Dans la figure 4.9, nous présentons les résultats de la segmentation de trois disques lombaires (S1-L5, L5-L4 et L4-L3) du sujet 4 de la base de données test1. À gauche, nous montrons le résultat de segmentation obtenu par notre RNC. À droite, nous montrons le résultat de segmentation après avoir utilisé l'algorithme de coupe de graphe sur les cartes de probabilité générées par notre réseau pour ce sujet. La segmentation obtenue par notre RNC montre une hernie discale au niveau du disque L5-L4 (entourée en bleu), qui est une fausse prédiction dans ce cas. Nous pouvons également constater la manque de certains pixels au niveau du disque inférieur. En appliquant l'algorithme de coupe de graphe sur les cartes de probabilité générées par notre RNC, le résultat est devenu plus précis et les faux pixels ont été éliminés. L'algorithme a également corrigé la segmentation du disque inférieur en rajoutant les pixels manquants.

4.6 Resultats de la segmentation multi-classes

4.6.1 Base de données

La base de données est constituée de 10 images IRM 3D spin-écho pondérées en T2 de façon sagittale provenant de 10 sujets différents. Les images ont été acquises à l'aide d'un scanner Siemens Tesla 1.5 et ré-échantillonnées à une taille de voxel de $2 \times 1{,}25 \times 1{,}25\ mm^3$. Chaque image contient au moins 7 corps vertébraux de la colonne vertébrale inférieure (T11 - L5). Une

segmentation de référence manuelle est fournie pour chaque sujet où les étiquettes sont codées avec des valeurs numériques (0, 1 et 2). Chaque référence manuelle contient 7 vertèbres et 7 disques, l'arrière-plan est étiqueté par zéro, les vertèbres par 1 et les disques par 2. Toutes les images et les références manuelles sont stockées au format de fichier NIFTI.

4.6.2 Les résultats de la segmentation du RNC

Dans les deux sections précédentes, la segmentation de la colonne vertébrale a été traitée comme une tâche de classification en deux classes. Certaines études ont montré que l'utilisation de deux classes seulement présentait certains inconvénients (Moran *et al.* (2018); Whitehead *et al.* (2018)), notamment le déséquilibre du nombre de pixels entre les deux classes. Ceci est dû au fait que la zone spatiale de la colonne vertébrale dans l'image médicale est petite comparée à la zone de l'arrière-plan. En conséquence, les modèles basés sur l'apprentissage acquerront plus de connaissances sur l'arrière-plan que la colonne vertébrale. Sur la base de cette problématique, nous allons examiner si la segmentation simultanée des deux structures, vertèbres et disques, est possible et si elle peut améliorer les résultats de segmentation ou l'aggraver.

Dans cette partie, les résultats de segmentation souhaités doivent inclure 3 étiquettes représentant les trois structures de notre image (vertèbres, disques et arrière-plan). Nous devons d'abord adapter notre RNC afin qu'il fournisse des masques multi-classes. Pour ce faire, il suffit de changer un seul paramètre dans notre code pour que notre réseau traite les tâches multi-classes. Tous les autres détails d'implémentation restent les mêmes.

Pour l'évaluation de notre approche, nous avons utilisé 9 sujets comme données d'apprentissage pour notre réseau, est ensuite validé sur le sujet restant. Nous avons répété cette validation croisée 10 fois pour segmenter à chaque fois un sujet. Pour la segmentation des vertèbres, un CSD moyen de 94.18$\pm$ 1.94% et une HD moyenne de 1.56 $\pm$ 1.02 *mm* (moyenne $\pm$ Std) ont été obtenus. Pour les disques, nous avons obtenu un CSD moyen de 91.81$\pm$ 2.92% et une HD moyenne de 4.62 $\pm$ 7.08 *mm*. Les coefficients de similarité de Dice des résultats de segmentation de tous les sujets sont présentés dans le tableau 4.9.

Tableau 4.9 Résultats de segmentation obtenus par notre RNC sur les 10 sujets multi-classes.

Sujets		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	**Moyenne**
Vertèbres	**CSD (%)**	95.62	94.75	95.67	88.95	95.37	94.00	93.04	93.69	95.44	95.29	**94.18 ± 1.94**
	HD (mm)	1.41	1.41	1.00	4.58	1.00	1.41	1.41	1.41	1.00	1.00	**1.56 ± 1.02**
Disques	**CSD (%)**	91.47	93.12	93.42	84.44	92.91	93.09	88.48	92.80	93.77	94.64	**91.81 ± 2.92**
	HD (mm)	2.24	1.41	1.41	24.79	1.41	1.73	9.00	1.41	1.41	1.41	**4.62 ± 7.08**

Au meilleur de nos connaissances et sur la base de notre expérience, un coefficient de similarité supérieur à 94% pour la segmentation des vertèbres et un CSD supérieur à 91% pour la segmentation des disques, sont cliniquement acceptables et peuvent être considérés comme des résultats précis. Les résultats de la segmentation, obtenus pour les vertèbres et les disques, confirment que notre approche est capable de produire des résultats acceptables pour la tâche de segmentation multi-classes.

Pour examiner l'hypothèse que la segmentation multi-classes peut améliorer les résultats par rapport à la segmentation binaire, nous avons procédé comme pour les expériences précédentes en segmentant les vertèbres et les disques séparément. Pour ce faire, nous allons utiliser deux réseaux, chacun permettant de segmenter une seule structure. Pour la base d'apprentissage de chaque réseau, nous devons disposer des images IRM et de leurs annotations manuelles respectives de la structure cible. Pour avoir ces références binaires, nous avons divisé les annotations manuelles multi-classes de chaque sujet afin de disposer séparément de deux masques binaires, correspondant chacun à l'une des deux structures. Pour la mise en œuvre des réseaux, nous avons suivi la même approche que les expériences binaires précédentes. Nous avons utilisé neuf sujets à chaque fois et les annotations manuelles binaires respectives comme base de données d'apprentissage, puis nous avons évalué le sujet restant. Les résultats de la segmentation pour tous les sujets sont obtenus après avoir répété cette validation croisée dix fois. Les coefficients de similarité de la segmentation binaire des vertèbres et des disques intervertébraux pour les dix sujets sont présentés dans le tableau 4.10.

En utilisant uniquement les annotations manuelles des vertèbres pour l'apprentissage du réseau, nous avons obtenu un coefficient de similarité moyen des résultats de segmentation de

93.04%. Dans le cas de la segmentation des disques, nous avons atteint un CSD de 90,08%. En utilisant la segmentation multi-classes, le CSD moyen a augmenté de 0.38% pour les résultats de segmentation des vertèbres et de 1,05% pour les disques. Ces résultats ont clairement prouvé que, sur la base du même nombre de données d'apprentissage, l'utilisation d'annotations multi-classes peut améliorer les résultats de segmentation de chaque structure.

Tableau 4.10 Les coefficients de similarité des résultats de segmentation des vertèbres et des disques en utilisant des réseaux binaires et multi-classes.

méthodes / sujets	segmentation binaire				segmentation multi-classes			
	vertèbres		disques		vertèbres		disques	
	RNC	RNC + post-traitement	RNC	RNC + post-traitement	RNC	RNC + post-traitement	RNC	RNC + post-traitement
Sujet 1	95.17	95.72	89.58	89.72	95.40	95.62	91.43	91.47
Sujet 2	92.39	94.82	91.14	91.97	92.59	94.75	91.91	93.12
Sujet 3	89.77	95.23	88.82	92.15	92.05	95.67	89.61	93.42
Sujet 4	88.37	88.59	83.35	83.27	88.62	88.95	84.46	84.44
Sujet 5	93.00	95.49	92.14	92.17	94.85	95.37	92.51	92.91
Sujet 6	93.91	94.12	91.81	92.00	93.88	94.00	93.08	93.09
Sujet 7	92.59	93.00	86.21	86.06	92.83	93.04	88.91	88.48
Sujet 8	94.55	94.57	91.33	92.04	93.61	93.69	92.25	92.80
Sujet 9	95.76	96.00	93.09	93.28	95.26	95.44	93.73	93.77
Sujet 10	94.84	95.13	93.34	94.16	95.15	95.29	93.39	94.64
Moyenne (%)	93.04 $\pm$ 2.27	94.27 $\pm$ 2.06	90.08 $\pm$ 3.03	90.68 $\pm$ 3.25	93.42 $\pm$ 1.96	94.18 $\pm$ 1.94	91.13 $\pm$ 2.67	91.81 $\pm$ 2.92

Le tableau 4.10 présente également les résultats après le post-traitement des segmentations primaires obtenues par les différents réseaux. Comme nous l'avons déjà mentionné, le post-traitement que nous utilisons supprime toutes les fausses régions de la segmentation et ne retient que les sept grandes régions qui représentent les structures cibles.

Le post-traitement a amélioré le CSD moyen de pour la segmentation des vertèbres de 1,23% pour le cas binaire et seulement de 0,76% pour la cas multi-classes. Pour la segmentation des disques intervertébraux, l'amélioration du CSD a été presque similaire pour les deux tâches, soit 0,6% pour la segmentation binaire et 0,68% pour le cas multi-classes.

La figure 4.10 présente des exemples de segmentation du sujet 1 sur des vues mi-sagittales en utilisant deux RNC binaires et un RNC multi-classes. Les première et deuxième images de la première colonne montrent respectivement les résultats de la segmentation des vertèbres et des disques en utilisant séparément deux RNC binaires. La dernière image de la première colonne montre le résultat de la combinaison des deux segmentations binaires. La dernière image de la deuxième colonne montre le résultat de la segmentation obtenu par le réseau multi-classes.

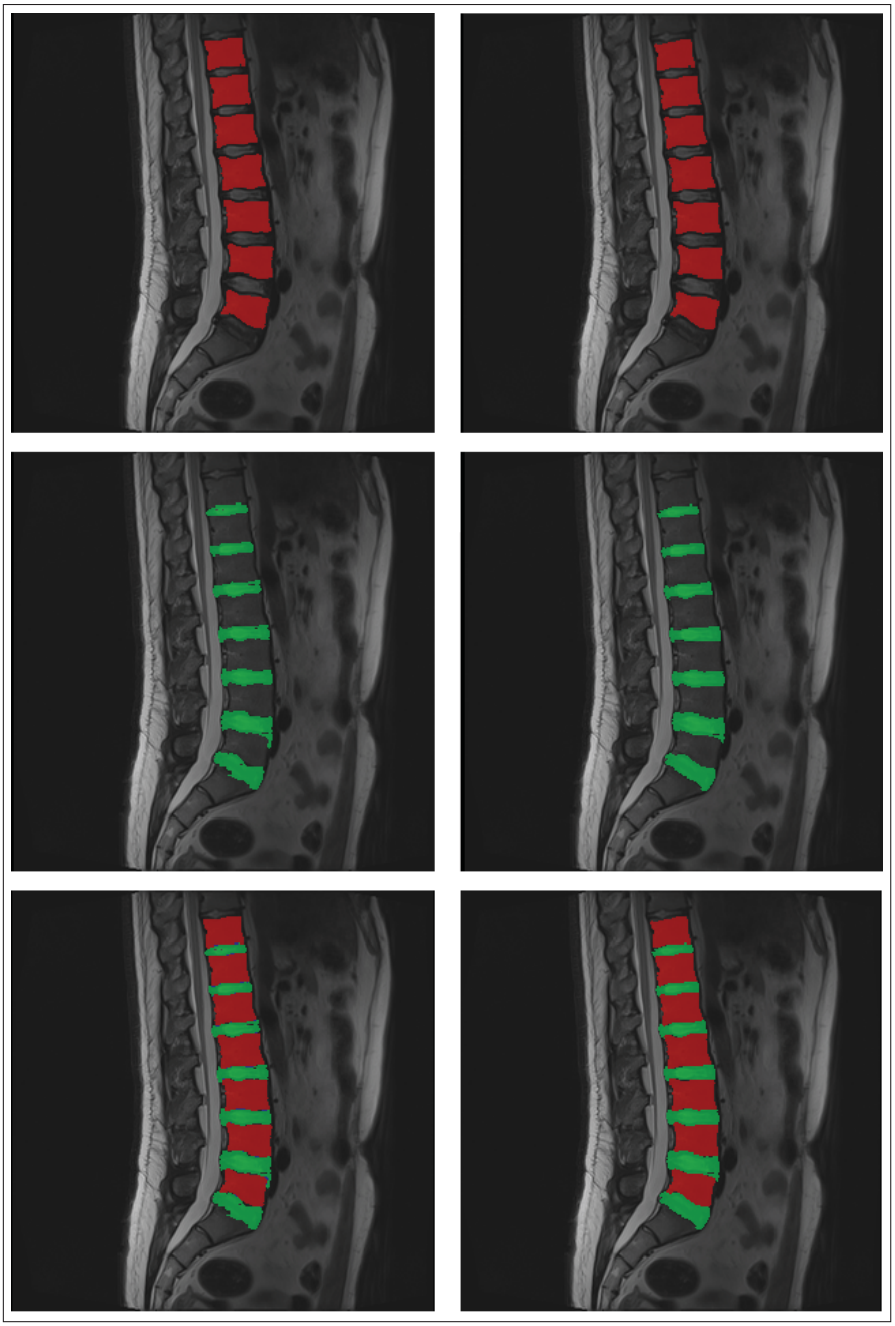

Figure 4.10 Résultats de segmentation des vertèbres (première ligne) et des disques (deuxième ligne), en utilisant deux RNC binaires (première colonne) et un RNC multi-classes (deuxième colonne), sur des vues mi-sagittales.

Les première et deuxième images de cette colonne montrent les résultats de la segmentation de chaque structure en scindant le résultat multi-classes en deux segmentations binaires distinctes.

La segmentation des vertèbres pour ce sujet en utilisant la segmentation multi-classes n'a pas été améliorée, ce qui est également le cas pour presque tous les sujets. Alors que, pour les disques, la figure montre que le résultat de la segmentation a été amélioré qualitativement et que la forme de chaque corps intervertébral est plus lisse que celle du résultat binaire. Le CSD du résultat de la segmentation du sujet 1, obtenu par le RNC binaire, a été de 89,72%. En utilisant le réseau multi-classes, le résultat de la segmentation s'est amélioré et le CSD a augmenté de 1,75%.

En combinant les résultats binaires des deux structures, les résultats de segmentation obtenu pour les 10 sujets atteint un CSD moyen de 92,48%. En utilisant des images multi-classes pour l'apprentissage de notre RNC, les segmentations complètes se sont améliorées et le CSD moyen atteint 93%, ce qui représente une augmentation de 0,52%.

4.6.3 Les résultats de la segmentation par α-expansion

En suivant le même démarche utilisé pour l'algorithme de coupe de graphe, nous avons appliqué l'algorithme α-expansion afin d'améliorer les résultats de segmentation primaires, en utilisant les cartes de probabilité générées par notre RNC multi-classes comme initialisation. Pour choisir les paramètres optimaux pour notre algorithme, nous avons essayé différentes combinaisons de paires de termes régularisations sur tous les sujets de notre base de données. La paire de termes qui nous a donné le coefficient de similarité de Dice global le plus élevé est (λ = 1.1 et σ = 190). Ces termes sont utilisés comme paramètres fixes dans l'algorithme pour les expériences suivantes. Les résultats de la segmentation de chaque sujet, obtenus par l'algorithme α-expansion, sont présentés dans le tableau 4.11.

L'algorithme α-expansion a amélioré les résultats de la segmentation des vertèbres pour tous les sujets et le CSD moyen a été plus précis que le résultat obtenu par le RNC de 0,45%. Pour le cas des disques, l'algorithme a légèrement amélioré les résultats de la segmentation et le CSD moyen est resté pratiquement identique (91,10%). Après le post-traitement de la segmentation des corps vertébraux, la différence entre le CSD moyen atteint par le RNC et le CSD obtenu par

Tableau 4.11 Les coefficients de similarité des résultats de segmentation des vertèbres et des disques en utilisant des réseaux binaires et multi-classes.

méthodes / sujets	segmentation par coupe de graphe				segmentation par α-expansion			
	vertèbres		disques		vertèbres		disques	
	résultat primaire	résultat post-traité	résultat primaire	résultat post-traité	résultat primaire	résultat post-traité	résultat primaire	résultat post-traité
Sujet 1	95.91	96.09	91.09	91.10	95.39	95.39	91.13	91.14
Sujet 2	92.96	95.11	92.77	93.26	93.26	94.96	93.13	93.28
Sujet 3	90.50	95.64	90.02	92.92	93.47	95.65	89.02	92.44
Sujet 4	88.73	88.65	84.44	84.42	88.94	88.94	84.94	84.94
Sujet 5	93.33	95.70	93.22	93.22	95.18	95.18	93.17	93.17
Sujet 6	94.26	94.30	92.79	92.79	94.22	94.22	92.27	92.27
Sujet 7	93.26	93.47	86.66	86.62	93.08	93.08	87.63	87.49
Sujet 8	95.01	95.01	92.59	92.94	94.19	94.19	92.62	92.62
Sujet 9	96.21	96.29	93.61	93.61	95.62	95.62	93.76	93.76
Sujet 10	95.05	95.31	94.26	94.81	95.33	95.44	93.31	94.37
Moyenne (%)	93.52 ± 2.25	94.56 ± 2.02	91.15 ± 3.06	91.57 ± 3.18	93.87 ± 1.87	94.27 ± 1.94	91.10 ± 2.30	91.55 ± 2.85

l'algorithme, a été réduite à 0,07%. Pour les disques, les segmentations post-traitées générées par le RNC ont été quantitativement plus précises que celles obtenues par l'algorithme.

En termes qualitatifs, les résultats de segmentation obtenus par l'algorithme α-expansion ne contiennent pas de pixels parasites. Le post-traitement des résultats primaires n'a aucune incidence sur la segmentation des sept sujets (voir tableau 4.11). Pour les trois autres sujets, leurs résultats de segmentation montrent quelques régions supplémentaires qui couvrent certaines parties du corps vertébral ou discale de la colonne. Dans notre étude, nous visons uniquement à segmenter sept régions de chaque structure. C'est pourquoi nous avons éliminé les parties supplémentaires des résultats de segmentation obtenus. Ces parties indésirables appartiennent aux vertèbres S1 et T10, et aux disques S2-S1 et T11-T10. La figure 4.11 montre des exemples de la présence de régions supplémentaires dans les résultats de la segmentation des sujets 2 et 3, sur des vues sagittales, qui ont été supprimées après le post-traitement.

L'utilisation de l'algorithme α-expansion sur les prédictions de notre réseau n'améliore pas nos résultats de façon significative. La seule raison est que notre réseau de neurones à convolution nous a donné les meilleurs résultats possibles sur ces sujets. Les cartes de probabilité, utilisées comme initialisation pour l'algorithme, ont été obtenues après avoir formé notre réseau à 35 époques. Pour montrer l'efficacité de l'algorithme, nous avons utilisé pour toutes les expériences suivantes les cartes de probabilité résultant de la première époque. Les résultats obtenus par notre RNC aussi que de l'algorithme α-expansion, sur les dix sujets multi-classes, sont présentés dans le tableau 4.12.

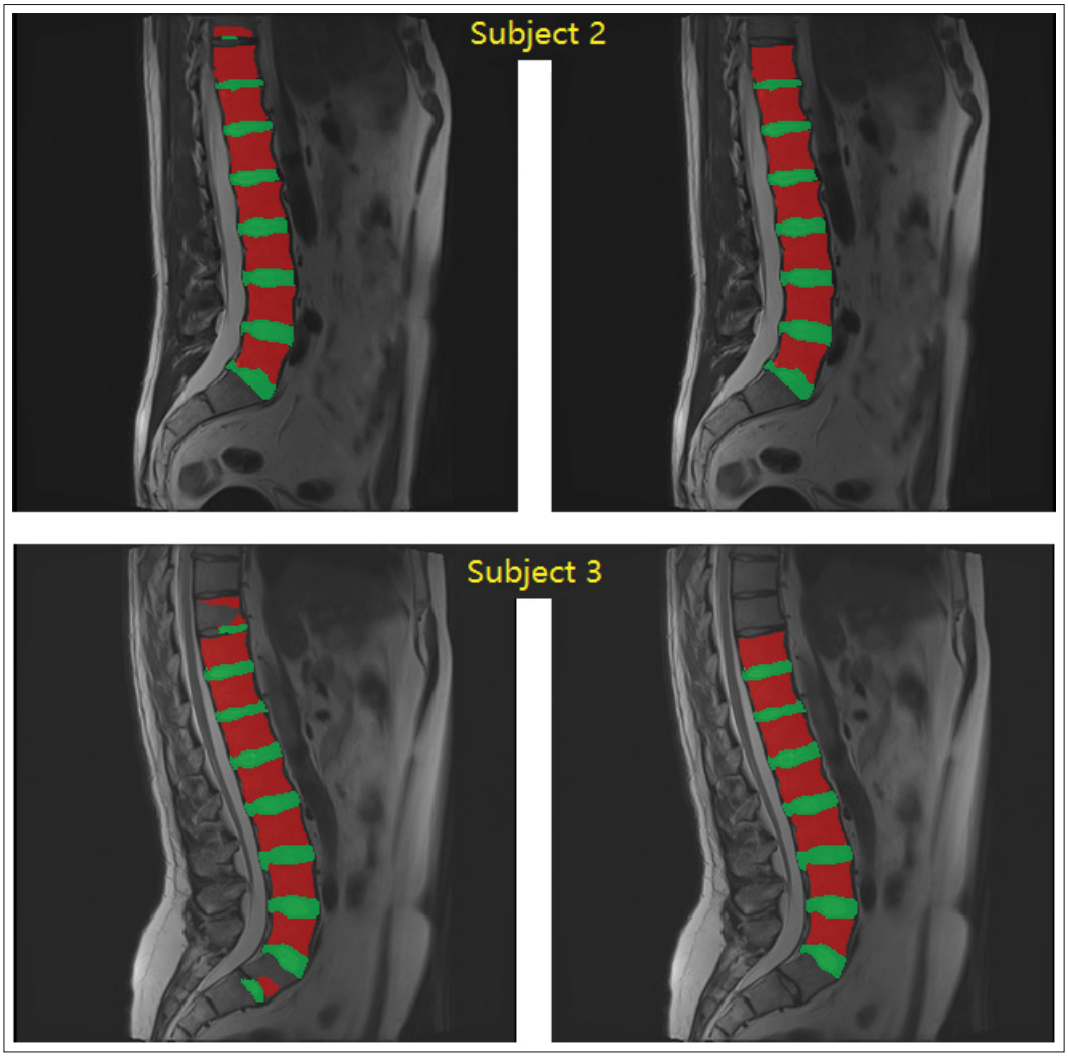

Figure 4.11 Résultats de segmentation des sujets 2 et 3 sur des vues mi-sagittales. (À gauche) Résultats de segmentation obtenus par l'algorithme α-expansion. (À droite) Résultats de segmentation post-traités.

Ces résultats montrent clairement l'effet de l'algorithme sur les prédictions obtenues par notre RNC. L'algorithme a permis d'augmenter le coefficient de similarité de Dice global pour la segmentation des vertèbres de 7.04%, et de 5.8% pour la segmentation des disques. La figure 4.12 montre la différence entre les résultats de segmentation obtenus par le RNC et par l'algorithme α-expansion pour le sujet 1.

Bien que l'algorithme a amélioré les segmentations obtenues par le RNC, ces résultats ne peuvent pas être considérés comme des segmentations précises vu que les coefficients de simi-

Tableau 4.12 Les CSD des résultats de segmentation obtenus par notre RNC et par l'algorithme α-expansion après l'apprentissage du réseau pour une seule époque.

Résultats obtenus par notre RNC											
Sujets	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	**Moyenne**
Vertèbres (%)	86.73	82.58	83.85	73.33	84.31	87.02	78.44	87.54	78.17	89.56	**83.15 $\pm$ 4.83**
Disques (%)	80.20	73.64	84.48	70.91	87.16	84.95	76.84	78.26	86.35	88.45	**81.12 $\pm$ 5.75**
Résultats obtenus par l'algorithme α-expansion											
Vertèbres (%)	93.21	93.05	93.73	68.33	92.16	91.58	91.76	92.59	92.05	93.43	**90.19 $\pm$ 7.32**
Disques (%)	87.70	78.90	91.41	79.12	90.21	90.44	76.83	88.04	92.62	93.97	**86.92 $\pm$ 5.95**

larité obtenus pour chacune des deux structures restent faibles et non acceptables cliniquement. Pour la segmentation des vertèbres, le coefficient de similarité moyen obtenu est inférieur de 4,08% à celui obtenu en utilisant comme initialisation les cartes de probabilité générées par le RNC après l'apprentissage du réseau pendant 35 époques. Pour les disques, le coefficient de similarité moyen obtenu est de 86.92%, soit 4,63% inférieur à celui obtenu en utilisant comme initialisation les cartes de probabilité générées après 35 époques. La raison de ces faibles résultats est que les algorithmes tels que la coupe de graphe et α-expansion dépendent fortement de l'initialisation. En d'autres termes, une mauvaise initialisation entraînera une mauvaise segmentation et une bonne initialisation conduira à une segmentation précise.

La figure 4.13 montre les résultats obtenus par l'algorithme d'expansion alpha en utilisant deux initialisations différentes. Les cartes de probabilité à gauche, qui dans ce cas représentent une mauvaise initialisation, ont conduit à une segmentation imprécise, tandis que les cartes de probabilité à droite, qui peuvent être considérées comme une bonne initialisation, ont permis d'obtenir la meilleure segmentation pour ce sujet.

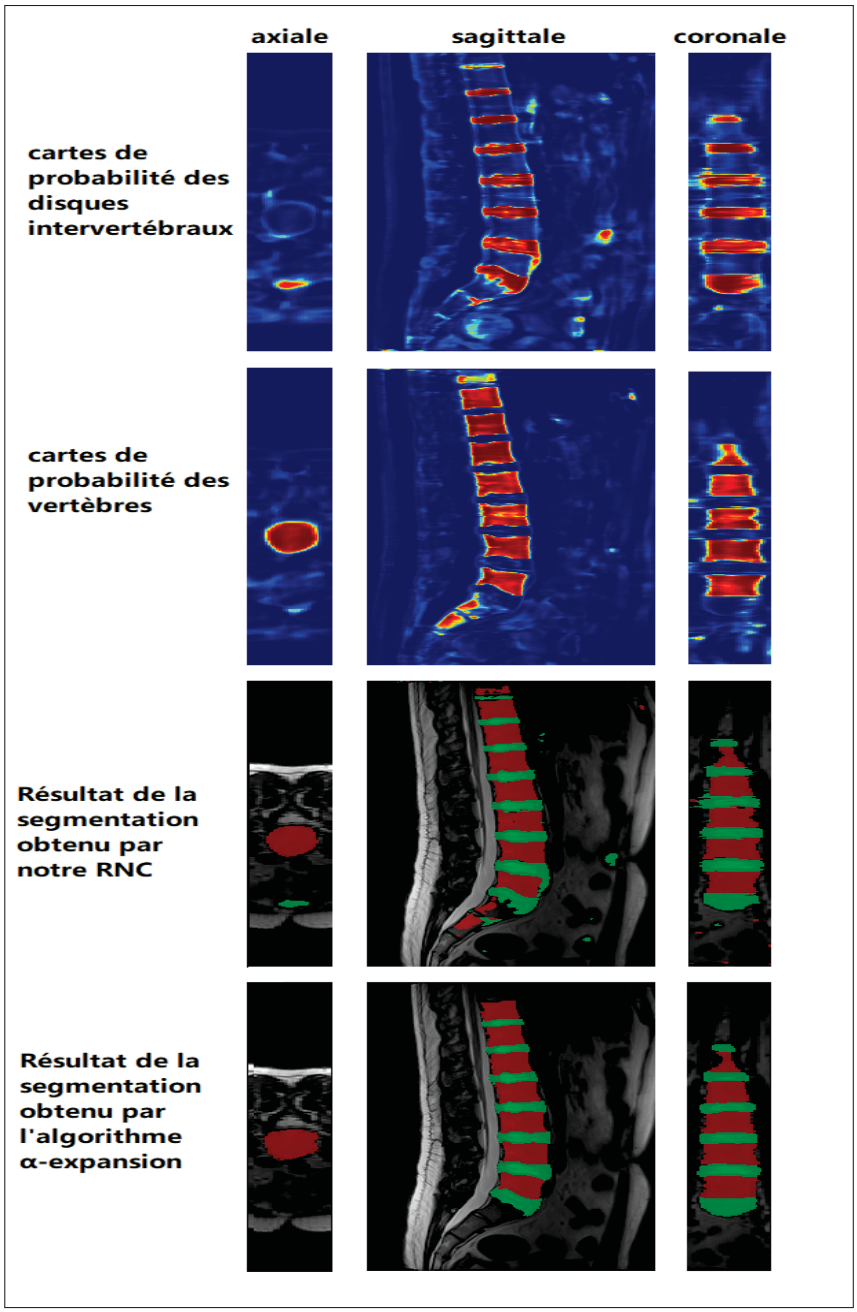

Figure 4.12 Résultats de segmentation multi-classes du sujet 1 après l'apprentissage du RNC pour une époque. (Première et deuxième lignes) Cartes de probabilité des disques et des vertèbres. (Troisième ligne) Résultat obtenu par le RNC. (Quatrième ligne) Résultat obtenu par l'algorithme α-expansion.

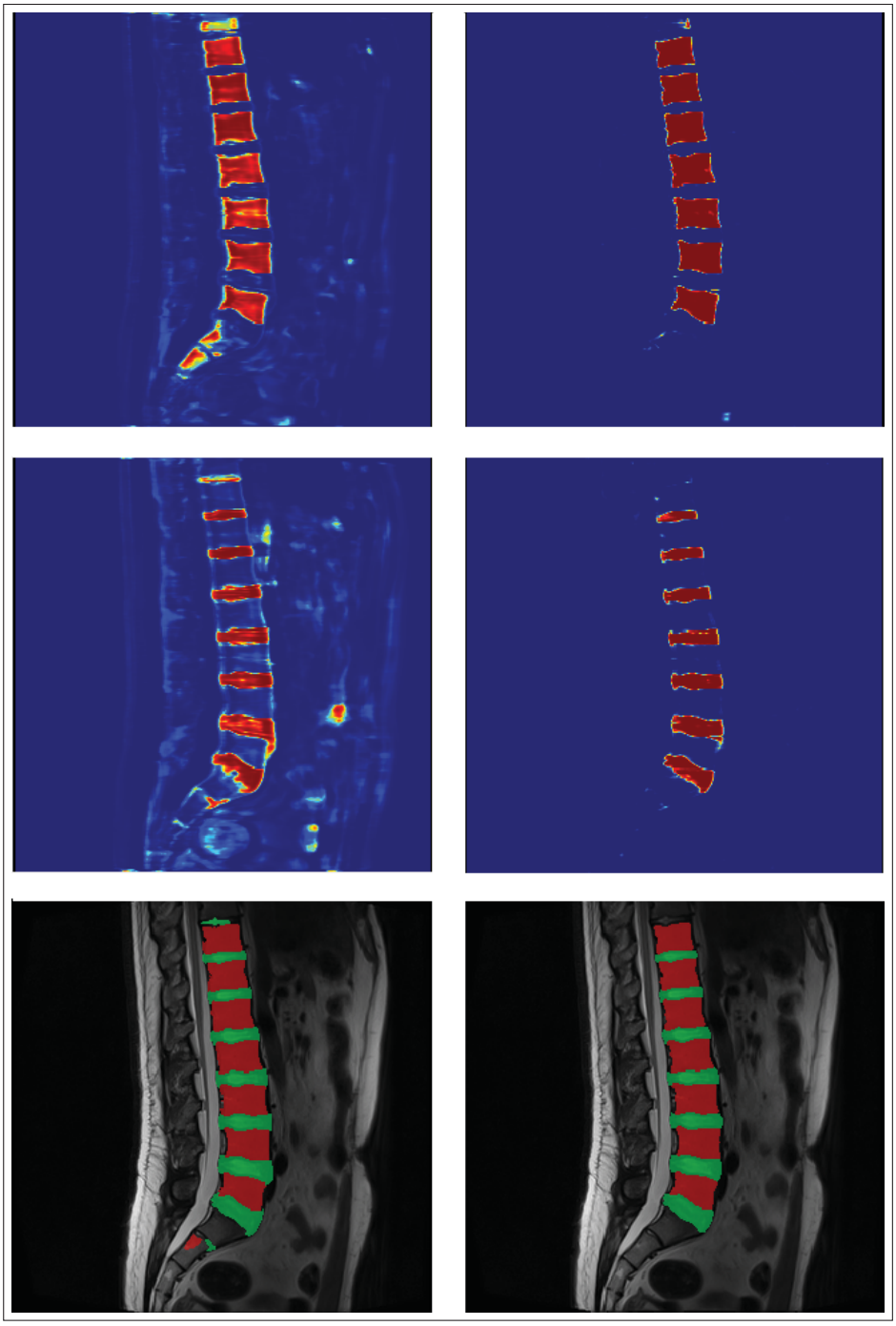

Figure 4.13 Résultats de segmentation multi-classes obtenus par l'algorithme α-expansion en utilisant deux différentes initialisations. (Première et deuxième lignes) Cartes de probabilité des vertèbres et des disques. (Troisième ligne) Résultat de segmentation.

4.7 Conclusion

Dans ce chapitre, nous avons démontré l'efficacité de notre méthode proposée sur la base des résultats de l'évaluation. Nous avons utilisé le coefficient de similarité de Dice (CSD) et la distance de Hausdorff (HD) pour évaluer les résultats de segmentation obtenus. Trois bases de données ont été utilisées dans notre étude. La première base de données est composée de 23 sujets sur lesquels nous avons utilisé un RNC binaire pour segmenter les vertèbres. La deuxième base de données est composée de 15 sujets sur lesquels nous avons utilisé un RNC binaire pour segmenter les disques intervertébraux. La dernière base de données est composée de 10 sujets sur lesquels nous avons utilisé un RNC multi-classes pour segmenter les deux structures ; vertèbres et disques. Pour affiner les résultats de la segmentation, nous avons utilisé l'algorithme de coupe de graphe pour les tâches binaires et l'algorithme α-expansion pour la tâche la segmentation multi-classes. Pour la base de données des vertèbres, nous avons atteint un CSD moyen de 94,89%. Pour la base de données des disques intervertébraux, nous avons obtenu un CSD moyen de 92,11%. Pour la segmentation multi-classes, nous avons obtenu un CSD moyen de 94,27% pour les vertèbres et de 91,81% pour les disques.

CHAPITRE 5

CONCLUSION ET TRAVAUX FUTURS

5.1 Conclusion

Dans ce mémoire, nous avons proposé une approche automatique et efficace pour la segmentation des vertèbres et des disques, qui combine les réseaux de neurones convolutifs à une régularisation par minimisation d'énergie.

Nous avons commencé par former deux réseaux de neurones entièrement convolutifs sur des patches d'image IRM bidimensionnels afin de fournir une segmentation binaire pour chacune des deux structures, les vertébrés et les disques. Nous avons ensuite utilisé l'algorithme de coupe de graphe afin d'améliorer les résultats primaires obtenus par les réseaux de neurones. Un tel algorithme nécessite une initialisation qui, dans la plupart des cas, est fournie manuellement par l'utilisateur. Dans notre travail, nous avons utilisé les cartes de probabilité générées par les réseaux de neurones comme données d'initialisation.

Pour le cas de la segmentation des deux structures combinées, nous avons formé un troisième réseau de neurones basé sur les données contenant les annotations manuelles des vertèbres et des disques en même temps. Nous avons ensuite appliqué l'algorithme α-expansion pour améliorer les résultats primaires obtenus par le réseau.

Les résultats expérimentaux ont montré la capacité de ces réseaux à segmenter les deux structures d'une façon efficace et cliniquement acceptable. L'utilisation de l'algorithme de coupe de graphe ont amélioré d'avantage ces résultats quantitativement, en donnant des coefficients de similarité élevés et plus proches des segmentations manuelles faites par les experts, et qualitativement en donnant des résultats finaux avec un minimum d'erreur et en rendant les formes des structures d'intérêt plus lisses.

Pour le cas multi-classes, les résultats expérimentaux ont prouvé que l'utilisation des images combinant les deux structures ont donné des résultats meilleurs que les résultats de la segmen-

tation des structures séparées. L'utilisation de l'algorithme α-expansion a fourni des résultats de segmentation lisse et dépourvu de pixels parasites. Ces résultats sont, cependant, moins précis quantitativement que ceux obtenus par le réseau.

5.2 Travaux futurs

La similarité des tissus des vertèbres, des disques et des structures qui l'entourent est l'une des difficultés auxquelles nous sommes confrontés dans la segmentation des images médicales. Cela entraînera une faible discrimination entre les structures d'intérêt et les pixels d'arrière-plan. Notre modèle proposé a classé certains pixels d'arrière-plan en vertèbres (ou disques) et certains pixels vertébraux en arrière-plan.

Ainsi, l'un des travaux à venir consiste à **développer un modèle plus robuste** afin de disposer d'une capacité de discrimination plus puissante permettant de différencier les pixels des structures d'intérêt de celles de l'arrière-plan.

La convolution avec un filtre de petite taille est préférable pour extraire les informations locales, mais pour capturer les informations distribuées dans un contexte global, il est préférable d'utiliser un filtre de plus grande taille. Pour tirer avantage des deux types de filtres, nous pouvons **utiliser les blocs *Inception*** développé par (Szegedy *et al.* (2015)), dans lesquels des convolutions avec des filtres de tailles différentes fonctionnent dans la même couche de manière parallèle. Un nouveau modèle d'*Inception* a également été introduit par (Dolz *et al.* (2018b)). Il contient des blocs convolutifs dilatés en parallèle avec différents taux de dilatation. Les convolutions dilatées aident à capturer plus efficacement le contexte global, ce qui permet d'abandonner l'utilisation des couches de max-pooling.

Nous pouvons également **utiliser des connexions de saut** dans l'architecture de notre réseau, qui connectent les sorties des couches peu profondes à l'entrée des couches suivantes, pour récupérer les informations perdues lors du passage des cartes de caractéristiques à travers plusieurs couches.

Dans notre travail, la segmentation d'une image IRM 3D est réalisée en traitant des images bidimensionnels de manière indépendante, ce qui pourrait constituer une utilisation non optimale des données d'images médicales. Pour incorporer les informations contextuelles 3D des structures d'intérêt, nous proposons comme futurs travaux d'**utiliser des filtres tridimensionnels** et de comparer leurs impacts par rapport aux résultats obtenus avec des filtres 2D. Une des raisons qui a découragé les chercheurs d'utiliser des filtres 3D est le coût élevé en calcul dû au nombre accru de paramètres. Pour surmonter cette limitation, nous pouvons **utiliser les trois images 2D orthogonales des différentes vues comme entrée** pour notre réseau, autrement nous pouvons **utiliser des filtres 3D de petite taille**.

BIBLIOGRAPHIE

Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., Deng, L., Penn, G. & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, 22(10), 1533–1545.

Adams, R. & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6), 641–647.

Al Arif, S. M. M. R., Knapp, K. & Slabaugh, G. (2018). Shape-Aware Deep Convolutional Neural Network for Vertebrae Segmentation. *Computational Methods and Clinical Applications in Musculoskeletal Imaging*, pp. 12–24.

Andersson, G. B. (1999). Epidemiological features of chronic low-back pain. *The lancet*, 354(9178), 581–585.

Andrej Karpathy, S. C. C. (2018, Janvier). CS231n : Convolutional Neural Networks for Visual Recognition. Repéré à http://cs231n.github.io/neural-networks-1/.

Aslan, M. S., Ali, A., Chen, D., Arnold, B., Farag, A. A. & Xiang, P. (2010). 3D vertebrae segmentation using graph cuts with shape prior constraints. *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 2193–2196.

Athertya, J. S. & Kumar, G. S. (2016). Fuzzy Clustering Based Segmentation Of Vertebrae in T1-Weighted Spinal MR Images. *CoRR*, abs/1605.02460.

Ayed, I. B., Punithakumar, K., Garvin, G., Romano, W. & Li, S. (2011). Graph cuts with invariant object-interaction priors : application to intervertebral disc segmentation. *Biennial International Conference on Information Processing in Medical Imaging*, pp. 221–232.

Ayed, I. B., Punithakumar, K., Minhas, R., Joshi, R. & Garvin, G. J. (2012). Vertebral body segmentation in MRI via convex relaxation and distribution matching. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 520–527.

Badrinarayanan, V., Handa, A. & Cipolla, R. (2015). Segnet : A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv :1505.07293*.

Badrinarayanan, V., Kendall, A. & Cipolla, R. (2017). Segnet : A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481–2495.

Been, E. & Kalichman, L. (2014). Lumbar lordosis. *The Spine Journal*, 14(1), 87–97.

Bezdek, J. C., Ehrlich, R. & Full, W. (1984). FCM : The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3), 191–203.

Boykov, Y., Veksler, O. & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11), 1222–1239.

Boykov, Y. Y. & Jolly, M.-P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 1, 105–112.

Brown, A., Angus, D. & Chen, S. (2005). *Costs and Outcomes of Chiropractic Treatment for Lower Back Pain*. Canadian Coordinating Office for Health Technology Assessment.

Castro-Mateos, I., Pozo, J. M., Lazary, A. & Frangi, A. F. (2014). 2D segmentation of intervertebral discs and its degree of degeneration from T2-weighted magnetic resonance images. *Medical Imaging 2014 : Computer-Aided Diagnosis*, 9035, 903517.

Castro-Mateos, I., Pozo, J. M., Lazary, A. & Frangi, A. F. (2016). Automatic construction of patient-specific finite-element mesh of the spine from IVDs and vertebra segmentations. *Medical Imaging 2016 : Biomedical Applications in Molecular, Structural, and Functional Imaging*, 9788, 97881U.

Chen, H., Dou, Q., Wang, X., Qin, J., Cheng, J. C. & Heng, P.-A. (2016). 3D fully convolutional networks for intervertebral disc localization and segmentation. *International Conference on Medical Imaging and Virtual Reality*, pp. 375–382.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv :1412.7062*.

Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv :1706.05587*.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. (2018). 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.

Chou, R., Qaseem, A., Snow, V., Casey, D., Cross, J. T., Shekelle, P. & Owens, D. K. (2007). Diagnosis and treatment of low back pain : a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Annals of internal medicine*, 147(7), 478–491.

Chu, C., Belavỳ, D. L., Armbrecht, G., Bansmann, M., Felsenberg, D. & Zheng, G. (2015a). Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method. *PloS one*, 10(11), e0143327.

Chu, C., Belavy, D. L., Armbrecht, G., Bansmann, M., Felsenberg, D. & Zheng, G. (2015b). Annotated T2-weighted MR images of the Lower Spine Annotated T2-weighted MR images of the Lower Spine.

Chu, C., Yu, W., Li, S. & Zheng, G. (2015c). Localization and segmentation of 3D intervertebral discs from MR images via a learning based method : a validation framework. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 141–149.

Chuang, K.-S., Tzeng, H.-L., Chen, S., Wu, J. & Chen, T.-J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *computerized medical imaging and graphics*, 30(1), 9–15.

Ciresan, D., Giusti, A., Gambardella, L. M. & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, pp. 2843–2851.

Cireşan, D. C., Meier, U., Gambardella, L. M. & Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, 22(12), 3207–3220.

Coillard, C. & Rivard, C. H. (1996). Vertebral deformities and scoliosis. *European Spine Journal*, 5(2), 91–100.

Coleman, G. B. & Andrews, H. C. (1979). Image segmentation by clustering. *Proceedings of the IEEE*, 67(5), 773–785.

Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, 61(1), 38–59.

Dagenais, S., Caro, J. & Haldeman, S. (2008). A systematic review of low back pain cost of illness studies in the United States and internationally. *The spine journal*, 8(1), 8–20.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7), 1160–1169.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009). Imagenet : A large-scale hierarchical image database. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255.

Dey, V., Zhang, Y. & Zhong, M. (2010). A review on image segmentation techniques with remote sensing perspective.

Dolz, J., Desrosiers, C. & Ayed, I. B. (2018a). 3D fully convolutional networks for subcortical segmentation in MRI : A large-scale study. *NeuroImage*, 170, 456–470.

Dolz, J., Desrosiers, C. & Ayed, I. B. (2018b). IVD-Net : Intervertebral disc localization and segmentation in MRI with a multi-modal UNet. *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*, pp. 130–143.

Donelson, R., McIntosh, G. & Hall, H. (2012). Is it time to rethink the typical course of low back pain ? *PM&R*, 4(6), 394–401.

Dong, X. & Zheng, G. (2015). Automated 3D lumbar intervertebral disc segmentation from MRI data sets. Dans *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging* (pp. 131–142). Springer.

Forsberg, D. (2015). Atlas-based segmentation of the thoracic and lumbar vertebrae. Dans *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging* (pp. 215–220). Springer.

Freburger, J. K., Holmes, G. M., Agans, R. P., Jackman, A. M., Darter, J. D., Wallace, A. S., Castel, L. D., Kalsbeek, W. D. & Carey, T. S. (2009). The rising prevalence of chronic low back pain. *Archives of internal medicine*, 169(3), 251–258.

Frymoyer, J. W. (1988). Back pain and sciatica. *New England Journal of Medicine*, 318(5), 291–300.

Georges Dolisi, M. (2019, Janvier). Torsiscoliose Torsi-scoliose. Repéré à https://www.medelli.fr/glossaire-medical/torsiscoliosetorsi-scoliose/.

Ghosh, S. & Chaudhary, V. (2014). Supervised methods for detection and segmentation of tissues in clinical lumbar MRI. *Computerized medical imaging and graphics*, 38(7), 639–649.

Ghosh, S., Raja'S, A., Chaudhary, V. & Dhillon, G. (2011). Automatic lumbar vertebra segmentation from clinical CT for wedge compression fracture diagnosis. *Medical Imaging 2011 : Computer-Aided Diagnosis*, 7963, 796303.

Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587.

Gonzalez, R. C. & Woods, R. E. (2002). Thresholding. *Digital Image Processing*, 595–611.

Graves, A. & Jaitly, N. (2014). Towards end-to-end speech recognition with recurrent neural networks. *International Conference on Machine Learning*, pp. 1764–1772.

Greig, D. M., Porteous, B. T. & Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 271–279.

Gross, D. P., Ferrari, R., Russell, A. S., Battié, M. C., Schopflocher, D., Hu, R. W., Waddell, G. & Buchbinder, R. (2006). A population-based survey of back pain beliefs in Canada. *Spine*, 31(18), 2142–2145.

Guzmán, J., Esmail, R., Karjalainen, K., Malmivaara, A., Irvin, E. & Bombardier, C. (2001). Multidisciplinary rehabilitation for chronic low back pain : systematic review. *Bmj*, 322(7301), 1511–1516.

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A. et al. (2014). Deep speech : Scaling up end-to-end speech recognition. *arXiv preprint arXiv :1412.5567*.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M. & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35, 18–31.

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Delving deep into rectifiers : Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.

He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hille, G., Glaßer, S. & Tönnies, K. (2016). Hybrid level-sets for vertebral body segmentation in Clinical Spine MRI. *Procedia Computer Science*, 90, 22–27.

Hoy, D., Brooks, P., Blyth, F. & Buchbinder, R. (2010). The epidemiology of low back pain. *Best practice & research Clinical rheumatology*, 24(6), 769–781.

Hoy, D., Bain, C., Williams, G., March, L., Brooks, P., Blyth, F., Woolf, A., Vos, T. & Buchbinder, R. (2012). A systematic review of the global prevalence of low back pain. *Arthritis & Rheumatology*, 64(6), 2028–2037.

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., Barendregt, J. et al. (2014). The global burden of low back pain : estimates from the Global Burden of Disease 2010 study. *Annals of the rheumatic diseases*, 73(6), 968–974.

Hu, J., Shen, L. & Sun, G. (2017). Squeeze-and-excitation networks. *arXiv preprint arXiv :1709.01507*.

Huang, F. J., Boureau, Y.-L., LeCun, Y. et al. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. *2007 IEEE conference on computer vision and pattern recognition*, pp. 1–8.

Huang, J., Jian, F., Wu, H. & Li, H. (2013). An improved level set method for vertebra CT image segmentation. *Biomedical engineering online*, 12(1), 48.

Hutt, H., Everson, R. & Meakin, J. (2015). 3d intervertebral disc segmentation from mri using supervoxel-based crfs. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 125–129.

Janssens, R., Zeng, G. & Zheng, G. (2017). Fully Automatic Segmentation of Lumbar Vertebrae from CT Images using Cascaded 3D Fully Convolutional Networks. *arXiv preprint arXiv :1712.01509*.

Ji, X., Zheng, G., Belavy, D. & Ni, D. (2016a). Automated intervertebral disc segmentation using deep convolutional neural networks. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 38–48.

Ji, X., Zheng, G., Liu, L. & Ni, D. (2016b). Fully automatic localization and segmentation of intervertebral disc from 3d multi-modality mr images by regression forest and cnn. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 92–101.

Kamdi, S. & Krishna, R. (2012). Image segmentation and region growing algorithm. *International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume*, 2.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D. & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, 36, 61–78.

Kaplan, K. M., Spivak, J. M. & Bendo, J. A. (2005). Embryology of the spine and associated congenital abnormalities. *The Spine Journal*, 5(5), 564–576.

Katz, J. N. (2006). Lumbar disc disorders and low-back pain : socioeconomic factors and consequences. *JBJS*, 88, 21–24.

Kaur, D. & Kaur, Y. (2014). Various image segmentation techniques : a review. *International Journal of Computer Science and Mobile Computing*, 3(5), 809–814.

Knecht, C., Humphreys, B. K. & Wirth, B. (2017). An Observational Study on Recurrences of Low Back Pain During the First 12 Months After Chiropractic Treatment. *Journal of Manipulative & Physiological Therapeutics*, 40(6), 427–433.

Korez, R., Ibragimov, B., Likar, B., Pernuš, F. & Vrtovec, T. (2015a). Deformable model-based segmentation of intervertebral discs from MR spine images by using the SSC descriptor. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 117–124.

Korez, R., Ibragimov, B., Likar, B., Pernuš, F. & Vrtovec, T. (2015b). A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation. *IEEE transactions on medical imaging*, 34(8), 1649–1662.

Korez, R., Likar, B., Pernuš, F. & Vrtovec, T. (2016). Model-Based Segmentation of Vertebral Bodies from MR Images with 3D CNNs (pp. 433–441). Cham : Springer International Publishing.

Korez, R., Ibragimov, B., Likar, B., Pernuš, F. & Vrtovec, T. (2017). Intervertebral disc segmentation in MR images with 3D convolutional networks. *Medical Imaging 2017 : Image Processing*, 10133, 1013306.

Krähenbühl, P. & Koltun, V. (2011). Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, pp. 109–117.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pp. 1097–1105.

Lacasse, A., Roy, J.-S., Parent, A. J., Noushi, N., Odenigbo, C., Pagé, G., Beaudet, N., Choinière, M., Stone, L. S., Ware, M. A. et al. (2017). The Canadian minimum dataset for chronic low back pain research : a cross-cultural adaptation of the National Institutes of Health Task Force Research Standards. *CMAJ open*, 5(1), E237.

Law, M. W., Tay, K., Leung, A., Garvin, G. J. & Li, S. (2013). Intervertebral disc segmentation in MR images using anisotropic oriented flux. *Medical image analysis*, 17(1), 43–61.

Lawrence, S., Giles, C. L., Tsoi, A. C. & Back, A. D. (1997). Face recognition : A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1), 98–113.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.

Lessmann, N., van Ginneken, B. & Išgum, I. (2018). Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images. *Medical Imaging 2018 : Image Processing*, 10574, 1057408.

Lidgren, L. (2003). The bone and joint decade 2000-2010. SciELO Public Health.

Lin, M., Chen, Q. & Yan, S. (2013). Network in network. *arXiv preprint arXiv :1312.4400*.

Liu, C. & Zhao, L. (2018). Intervertebral Disc Segmentation and Localization from Multimodality MR Images with 2.5 D Multi-scale Fully Convolutional Network and Geometric Constraint Post-processing. *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*, pp. 144–153.

Long, J., Shelhamer, E. & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.

Lonstein, J. E. (1999). Congenital spine deformities : scoliosis, kyphosis, and lordosis. *Orthopedic Clinics*, 30(3), 387–405.

Lonstein, J. E. & Carlson, J. (1984). The prediction of curve progression in untreated idiopathic scoliosis during growth. *Journal of Bone and Joint Surgery-Series A*, 66(7), 1061–1071.

Luoma, K., Riihimäki, H., Luukkonen, R., Raininko, R., Viikari-Juntura, E. & Lamminen, A. (2000). Low back pain in relation to lumbar disc degeneration. *Spine*, 25(4), 487–492.

Maas, A. L., Hannun, A. Y. & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*, 30(1), 3.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. et al. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10), 1993–2024.

Michopoulou, S. K., Costaridou, L., Panagiotopoulos, E., Speller, R., Panayiotakis, G. & Todd-Pokropek, A. (2009). Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. *IEEE transactions on Biomedical Engineering*, 56(9), 2225–2231.

Modic, M. T., Steinberg, P. M., Ross, J. S., Masaryk, T. J. & Carter, J. R. (1988). Degenerative disk disease : assessment of changes in vertebral body marrow with MR imaging. *Radiology*, 166(1), 193-199.

Moran, S., Gaonkar, B., Whitehead, W., Wolk, A., Macyszyn, L. & Iyer, S. S. (2018). Deep learning for medical image segmentation–using the IBM TrueNorth neurosynaptic system. *Medical Imaging 2018 : Imaging Informatics for Healthcare, Research, and Applications*, 10579, 1057915.

Nair, V. & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.

Neubert, A., Fripp, J., Engstrom, C., Schwarz, R., Lauer, L., Salvado, O. & Crozier, S. (2012). Automated detection, 3D segmentation and analysis of high resolution spine MR images using statistical shape models. *Physics in Medicine & Biology*, 57(24), 8357.

Neubert, A., Fripp, J., Chandra, S. S., Engstrom, C. & Crozier, S. (2015). Automated intervertebral disc segmentation using probabilistic shape estimation and active shape models. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 150–158.

Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L. & Barbano, P. E. (2005). Toward automatic phenotyping of developing embryos from videos. *IEEE Transactions on Image Processing*, 14(9), 1360–1371.

Noh, H., Hong, S. & Han, B. (2015). Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE international conference on computer vision*, pp. 1520–1528.

Norouzi, A., Rahim, M. S. M., Altameem, A., Saba, T., Rad, A. E., Rehman, A. & Uddin, M. (2014). Medical image segmentation methods, algorithms, and applications. *IETE Technical Review*, 31(3), 199–213.

NVIDIA, C. (2018, Janvier). Feature extraction using convolution. Repéré à https://developer.nvidia.com/discover/convolution/.

OpenClassrooms, K. N. (2019, Janvier). Classez et segmentez des données visuelles. Repéré à https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles/5097666-tp-implementez-votre-premier-reseau-de-neurones-avec-keras/.

Pal, N. R. & Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy systems*, 3(3), 370–379.

Panjabi, M. M. (2003). Clinical spinal instability and low back pain. *Journal of electromyography and kinesiology*, 13(4), 371–379.

Parkhi, O. M., Vedaldi, A., Zisserman, A. et al. (2015). Deep Face Recognition. *BMVC*, 1(3), 6.

Pengel, L. H., Herbert, R. D., Maher, C. G. & Refshauge, K. M. (2003). Acute low back pain : systematic review of its prognosis. *Bmj*, 327(7410), 323.

Pereira, S., Pinto, A., Alves, V. & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on medical imaging*, 35(5), 1240–1251.

Pham, D. L., Xu, C. & Prince, J. L. (2000). Current methods in medical image segmentation. *Annual review of biomedical engineering*, 2(1), 315–337.

Pinterest. (2019, Janvier). Best Gifts for a Nurse ! Repéré à https://www.pinterest.ca/pin/347340189990089767/.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once : Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788.

Ren, S., He, K., Girshick, R. & Sun, J. (2015). Faster r-cnn : Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, pp. 91–99.

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241.

Rubin, D. I. (2007). Epidemiology and risk factors for spine pain. *Neurologic clinics*, 25(2), 353–371.

Sahoo, P. K., Soltani, S. & Wong, A. K. (1988). A survey of thresholding techniques. *Computer vision, graphics, and image processing*, 41(2), 233–260.

Schneider, S., Schmitt, H., Zoller, S. & Schiltenwolf, M. (2005). Workplace stress, lifestyle and social factors as correlates of back pain : a representative study of the German working population. *International archives of occupational and environmental health*, 78(4), 253–269.

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. et al. (2015). Going deeper with convolutions.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI*, 4, 12.

Themis Medica, D. F. P. (2019, Mars). Anatomie de la colonne vertébrale. Repéré à https://chirurgie-dos.com/canal-lombaire-etroit-stenose-foraminale/anatomie-de-colonne-vertebrale-canal-lombaire-etroit-stenose-foraminale/.

Urschler, M., Hammernik, K., Ebner, T. & Štern, D. (2015). Automatic intervertebral disc localization and segmentation in 3d mr images based on regression forests and active contours. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 130–140.

Vania, M., Mureja, D. & Lee, D. (2017). Automatic Spine Segmentation using Convolutional Neural Network via Redundant Generation of Class Labels for 3D Spine Modeling. *arXiv preprint arXiv :1712.01640*.

Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 1, I–I.

Wang, C. & Forsberg, D. (2015). Segmentation of intervertebral discs in 3D MRI data using multi-atlas based registration. *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pp. 107–116.

Whitehead, W., Moran, S., Gaonkar, B., Macyszyn, L. & Iyer, S. (2018). A deep learning approach to spine segmentation using a feed-forward chain of pixel-wise convolutional networks. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 868–871.

Wiggins, G. C., Shaffrey, C. I., Abel, M. F. & Menezes, A. H. (2003). Pediatric spinal deformities. *Neurosurgical focus*, 14(1), 1–14.

Withey, D. J. & Koles, Z. J. (2008). A review of medical image segmentation : methods and available software. *International Journal of Bioelectromagnetism*, 10(3), 125–148.

Xu, B., Wang, N., Chen, T. & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv :1505.00853*.

Yu, F. & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv :1511.07122*.

Zaitoun, N. M. & Aqel, M. J. (2015). Survey on image segmentation techniques. *Procedia Computer Science*, 65, 797–806.

Zeiler, M. D. & Fergus, R. (2014). Visualizing and understanding convolutional networks. *European conference on computer vision*, pp. 818–833.

Zeiler, M. D., Taylor, G. W., Fergus, R. et al. (2011). Adaptive deconvolutional networks for mid and high level feature learning. *ICCV*, 1(2), 6.

Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. (2017). Pyramid scene parsing network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890.

Zheng, G., Chu, C., Belavỳ, D. L., Ibragimov, B., Korez, R., Vrtovec, T., Hutt, H., Everson, R., Meakin, J., Andrade, I. L. et al. (2017). Evaluation and comparison of 3D intervertebral disc localization and segmentation methods for 3D T2 MR data : A grand challenge. *Medical image analysis*, 35, 327–344.

Zhu, S. C. & Yuille, A. (1996). Region competition : Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 18(9), 884–900.

Zhu, X., He, X., Wang, P., He, Q., Gao, D., Cheng, J. & Wu, B. (2016). A method of localization and segmentation of intervertebral discs in spine MRI based on Gabor filter bank. *Biomedical engineering online*, 15(1), 32.

Zijdenbos, A. P., Dawant, B. M., Margolin, R. A. & Palmer, A. C. (1994). Morphometric analysis of white matter lesions in MR images : method and validation. *IEEE transactions on medical imaging*, 13(4), 716–724.

Zukic, D., Vlasák, A., Dukatz, T., Egger, J., Horínek, D., Nimsky, C. & Kolb, A. (2012). Segmentation of Vertebral Bodies in MR Images. *VMV*, pp. 135–142.

Zukić, D., Vlasák, A., Egger, J., Hořínek, D., Nimsky, C. & Kolb, A. (2014). Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images. *Computer Graphics Forum*, 33(6), 190–204.