TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AP	Arrière-Plan
CHU	Centre Hospitalier Universitaire
CNN	Convolutional Neural Network
CV	Corps vertébral
DICOM	Digital Imaging and Communications in Medicine
DIV	Disque Intervertébral
ELU	Exponential Linear Units
FCM	Fuzzy C-Mean
FCN	Fully Connected Network
GPU	Graphics Processing Unit
HOG	Histogram of Oriented Gradients
IRM	Imagerie par Résonance Magnétique
ITK	Insight Segmentation and Registration Toolkit
MLP	Multi-Layer Perceptron
NIFTI	Neuroimaging Informatics Technology Initiative
NPY	Numpy
PCA	Principal Component Analysis
ReLU	Rectified Linear Units
scSE	Spatial and Channel Squeeze and Excitation
SE	Squeeze and Excitation
SVM	Support Vector Machine

INTRODUCTION

La scoliose idiopathique est une déformation tridimensionnelle importante de la colonne vertébrale dont les causes sont inconnues. Cette déformation entraîne une torsion et une courbure anormales de la colonne vertébrale dont les conséquences se manifestent par des déformations morphologiques qui varient selon sa gravité. Nous parlons de douleurs rachidiennes chroniques, inconfort, mobilité réduite, diminution de la qualité de vie en cas de scoliose sévère. Dans de telles circonstances, un traitement chirurgical est préconisé. La chirurgie minimalement invasive est une technique qui limite le traumatisme opératoire en réduisant l'incision à quelques centimètres. Sa mise en pratique vise à accélérer la récupération post-opératoire, réduire la douleur et le risque d'infections. Par contre, la visibilité du chirurgien est considérablement réduite. Les systèmes de navigation par ordinateur représentent une solution intéressante pour guider le chirurgien dans sa tâche. La discectomie par thoracoscopie est une procédure chirurgicale minimalement invasive du rachis employée dans le traitement de la scoliose. Elle consiste à réaliser une résection de certains DIVs pour relâcher la pression créée par la torsion de la colonne vertébrale. Ensuite, des vis sont placées sur les vertèbres qui vont créer un mouvement mécanique qui va redresser la colonne de manière continue durant plusieurs mois. Le système de navigation par ordinateur requis pour effectuer la discectomie implique l'utilisation d'une caméra endoscopique qui prend des images de la colonne vertébrale du patient en temps réel pour mettre à jour un modèle préopératoire 3D obtenu par résonance magnétique suite à un recalage 3D/2D. La mise en place du modèle préopératoire 3D implique une segmentation des CVs et des DIVs. L'imagerie par résonance magnétique (IRM) est une modalité qui offre une bonne visualisation de ces deux structures. Segmenter l'IRM pourrait fournir des informations quantitatives relatives aux DIVs et aux CVs, comme leur position, angle de rotation, taille, pourcentage de tissu restant, etc. Toutefois, la segmentation des vertèbres et des structures voisines en IRM reste une tâche difficile, ce qui est principalement dû au mauvais contraste entre

les os et les tissus mous. De plus, l'IRM tend à générer des images dont l'intensité est non homogène.

La littérature propose plusieurs méthodes pour la segmentation des DIVs et CVs à partir d'IRM. Toutefois, ces approches sont pour la plupart destinées à des contextes d'application différents de la scoliose. Le changement de morphologie dû à la scoliose fait que ces approches ne prennent pas en considération la déformation de ces deux structures et ne sont donc pas assez robustes pour effectuer leur segmentation. En parallèle, l'analyse d'images médicales a grandement profité de l'émergence de l'intelligence artificielle et plus particulièrement de l'apprentissage profond et sa capacité à résoudre des problèmes complexes de reconnaissance de formes. Cette capacité est conditionnelle à la disponibilité d'un grand nombre de données annotées, ce qui est très peu évident à obtenir dans le contexte médical et les données de patients scoliotiques en IRM ne font pas exception à la règle. D'ailleurs, ce projet fait suite à une étude menée par Chevrefils (2010) qui a proposé une méthode de segmentation des DIVs en IRM en ayant accès aux mêmes données des patients scoliotiques utilisées dans son travail.

Le but de ce projet est de proposer une méthode basée sur un apprentissage profond pour une segmentation simultanée des DIVs et CVs à partir d'IRM sur le plan sagittal de patients atteints de scoliose, quel qu'en soit le degré de sévérité. Pour ce faire, nous nous appuyons sur les avancées des réseaux de neurones convolutifs et les architectures récemment proposées pour réaliser de la segmentation sémantique d'images médicales. Les questions de recherche sont :

- Est-il possible avec peu de données de patients scoliotiques et un nombre conséquent de sujets non scoliotiques d'entraîner un réseau de neurones convolutif (CNN) à correctement segmenter les DIVs et CVs ?
- Quelle est la meilleure stratégie d'apprentissage dans ce cas ?

- Est-ce qu'une approche multi-classes est plus efficace qu'une segmentation binaire des deux structures ?

Nous posons l'hypothèse qu'un apprentissage avec des images de patients non scoliotiques permet d'extraire les caractéristiques générales des deux structures et une bonne représentation des IRMs. Cette information sera transférée à un second entraînement en utilisant cette fois un petit jeu de données de patients scoliotiques. Nous supposons aussi qu'il est utile de doter le modèle d'un mécanisme de recalibration de l'information apprise au fil de l'entraînement et ainsi maximiser la qualité de l'apprentissage et préserver les caractéristiques importantes. Nous supposons que le bloc *Spatial and Channel Squeeze and Excitation* (scSE) (Roy *et al.*, 2018) est un bon mécanisme pour gérer cela. Nous posons aussi l'hypothèse qu'une segmentation multi-classe serait plus judicieuse qu'une segmentation binaire en raison de la forte corrélation de localisation entre les deux structures.

Le reste de ce mémoire est subdivisé en 5 chapitres :

- **Le chapitre 1** passe en revue les connaissances médicales en lien avec ce projet et les connaissances théoriques en apprentissage profond pour la classification et segmentation d'images.
- **Le chapitre 2** revoit les méthodes classiques de segmentation des DIVs et des CVs à partir d'IRM et celles qui sont basées sur l'apprentissage profond pour mettre en relief la contribution de ce mémoire.
- **Le chapitre 3** présente notre méthode basée sur un apprentissage profond de segmentation simultanée des DIVs et CVs à partir d'IRM de patients atteints ou non de scoliose idiopathique. Un pré-traitement est d'abord appliqué sur l'ensemble des images pour réduire le bruit, apparier les contrastes des images issues des deux bases de données et augmenter le nombre de données d'entraînement via des stratégies d'augmentation de données géométriques. Le modèle est entraîné via une architecture complètement convolutive (FCN) de

type encodeur-décodeur où un bloc scSE est introduit. Le coefficient de Kappa de Cohen et l'entropie croisée forment la fonction objectif à minimiser. Le modèle est entraîné en deux temps. Une première passe est effectuée en utilisant les images de sujets non scoliotiques. Le modèle est ensuite ré-entraîné sur les images de patients scoliotiques en lui transférant les connaissances précédemment apprises. Contrairement à ce qui a été proposé dans la littérature, notre méthodologie met l'accent essentiellement sur la scoliose et au mécanisme d'attention sur les caractéristiques pertinentes via le bloc scSE et la gestion du déséquilibre de classes où l'arrière-plan de l'image est plus dominant que les DIVs et CVs à travers le coefficient de Kappa de Cohen.

- **Le chapitre 4** présente et analyse les résultats obtenus par notre méthode. Quatre variantes de notre modèle ont d'abord été entraînées et comparées entre elles. Les résultats montrent que le transfert d'apprentissage permet au modèle d'étendre sa représentation des deux structures DIVs et CVs aux déformations qu'ils peuvent subir. Quant au bloc scSE, il permet de doter le modèle d'une capacité de généralisation et de cibler les caractéristiques discriminantes des DIVs et des CVs. Nous démontrons aussi à travers une comparaison entre notre modèle et deux autres entraînés à réaliser de manière séparée une segmentation des DIVs pour l'un et une segmentation des CVs pour l'autre, qu'une segmentation multi-classes est plus efficace.
- **Le chapitre 5** conclut le mémoire et discute des perspectives d'amélioration.

CHAPITRE 1

REVUE DES CONNAISSANCES

Dans ce chapitre, nous introduisons les concepts et états de l'art des principales thématiques reliés à notre projet. Nous commençons par une revue des connaissances médicales (Section 1.1). Ensuite, nous nous focalisons sur l'aspect théorique de l'apprentissage automatique et particulièrement de l'apprentissage profond (Section 1.2).

1.1 Revue des connaissances médicales

Dans cette section nous nous intéressons à la colonne vertébrale et plus particulièrement à l'anatomie des vertèbres (Section 1.1.1.1) et des DIVs (Section 1.1.1.2), à la scoliose idiopathique et ses traitements (1.1.2) et enfin à l'IRM (Section 1.1.3) et son utilisation dans le contexte clinique de la scoliose.

1.1.1 Anatomie de la colonne vertébrale

La colonne vertébrale est une structure complexe de l'anatomie humaine qui relie la base du crâne au bassin permettant un support de l'ensemble du corps. Elle est composée de 24 corps osseux empilés appelés *vertèbres*. Ces dernières sont classifiées en 3 catégories selon leur région d'appartenance (Figure 1.1). Dans le sens du haut jusqu'au bas du dos, on compte : 7 vertèbres cervicales (C1-C7), 12 vertèbres thoraciques (T1-T12) et 5 vertèbres lombaires (L1-L5) s'ajoute le sacrum. Les vertèbres sont séparées par une structure cartilagineuse nommée *disque intervertébral*. Il en existe donc 23 dont 6 se situent dans la région cervicale, 12 dans la région thoracique et 5 dans la région lombaire (Keenan, 2015). Les principales fonctions de la colonne vertébrale sont la protection de la moelle épinière, le support de grosses structures osseuses comme le bassin, les côtes, la ceinture scapulaire, etc. La colonne vertébrale permet aussi de contrôler le mouvement du corps, lui permettant une certaine flexibilité tout en maintenant sa stabilité. À partir du plan coronal, une colonne vertébrale saine est aperçue comme

étant droite et symétrique. Or, dans un plan sagittal, on peut mieux apercevoir sa courbure naturelle qui se décompose en 2 courbures : *cyphose* dans la région thoracique et *lordose* dans les régions cervicale et lombaire (Figure 1.1 vue latérale). D'ailleurs ce sont ces deux courbures qui maintiennent l'équilibre du corps et absorbent le choc.

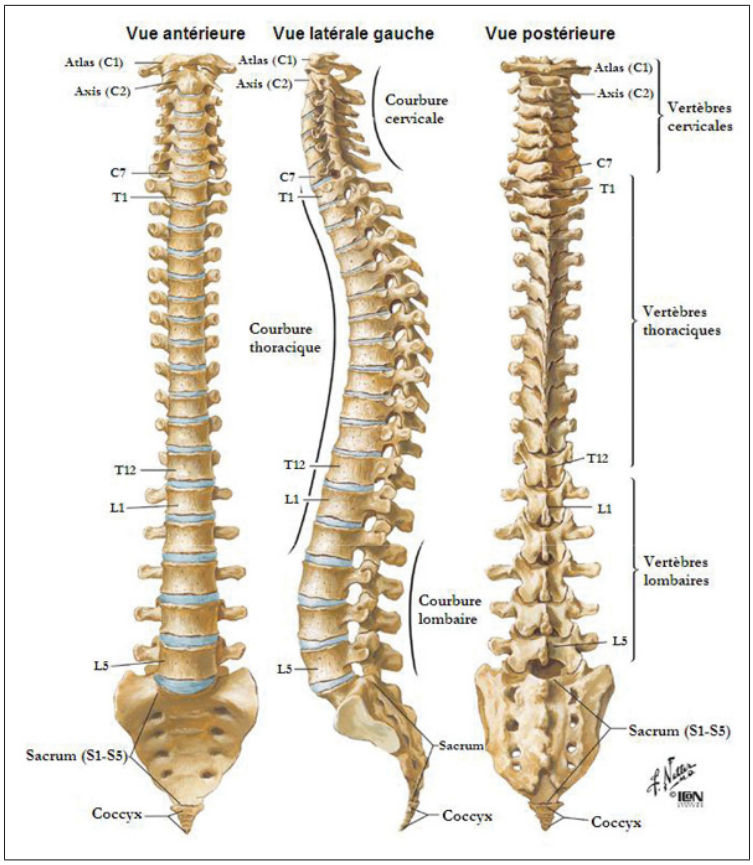


Figure 1.1 Illustration d'une colonne vertébrale humaine. Tirée de Chevrefils (2010)

1.1.1.1 Vertèbre

Une vertèbre (Figure 1.3) est composée de 2 parties essentielles : une partie antérieure appelée *corps vertébral* qui circonscrit un espace nommé trou vertébral et une partie postérieure appelée *arc neural* qui est constituée d'une paire de lames, d'une paire de pédicules et soutient un processus épineux, deux processus transverses et quatre processus articulaires. Du à la

superposition des vertèbres, les CVs forment un axe solide pour l'appui du tronc et les trous vertébraux créent le canal rachidien qui protège la moelle épinière.

Le CV est de forme cylindrique et constitue la plus grande structure d'une vertèbre. Ses faces supérieures et inférieures appelées plateaux vertébraux sont aplaties, présentent de petites aspérités et un anneau autour de leur circonférence. Elles sont en contact avec un disque intervertébral et servent à l'alimenter. À l'avant, le corps est de forme convexe horizontalement et concave verticalement. À l'arrière, il est légèrement concave horizontalement et plat verticalement.

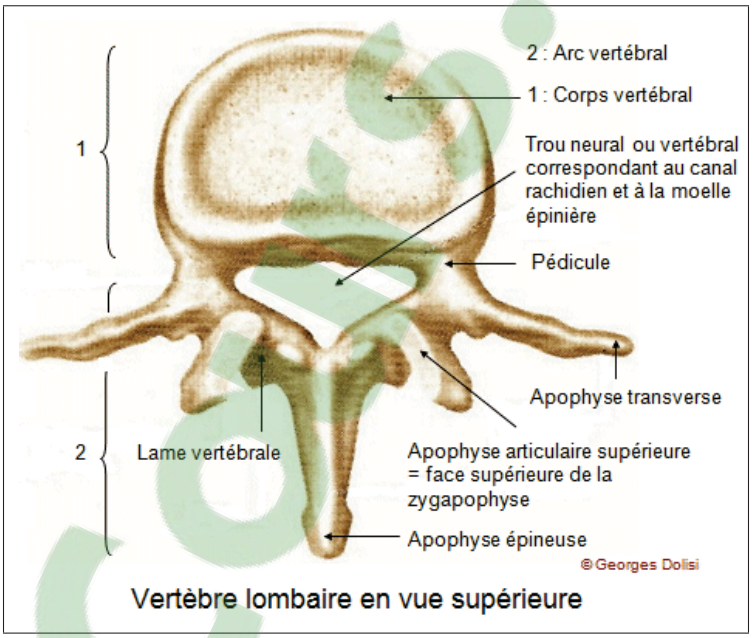


Figure 1.2 Illustration d'une vertèbre humaine. 1 est la partie antérieure et 2 est la partie postérieure. Tirée de Chevrefils (2010)

1.1.1.2 Disque intervertébral

Un disque intervertébral repose sur les plateaux vertébraux qui l'entourent et forme une articulation cartilagineuse qui permet un mouvement limité entre les CVs. Chaque disque a une structure vasculaire constituée d'un anneau fibreux externe appelé annulus fibrosus qui entoure

une substance gélatineuse interne appelée nucleus pulposus.L'annulus fibrosus se compose de plusieurs lames de fibrocartilage fait de collagène. Le nucleus pulposus est une structure molle de forme sphérique qui contient deux constituantes principales : les fibres de collagènes de types 2 qui se prolongent dans les plaques cartilagineuses et les protéoglycans. Il est donc composé de 80% d'eau, 15% de protéogyclan et 5% de fibre collagène. Cependant, sa composition extra cellulaire varie avec l'âge. Le nucleus pulposus est chargé de protéger le disque et les vertèbres supérieure et inférieure en distribuant la pression uniformément et empêchant toute contrainte qui pourrait les endommager. Il agit donc d'un absorbeur de chocs.

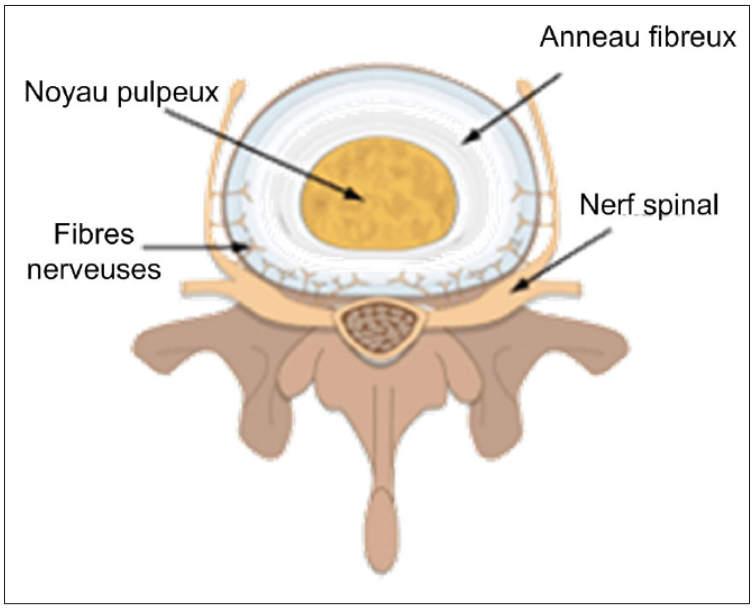


Figure 1.3 Illustration d'un disque intervertébral humain. Tirée de Chevrefils (2010)

1.1.2 Scoliose idiopathique de l'adolescent

La scoliose (figure 1.4) est une déformation complexe tridimensionnelle de la colonne vertébrale qui cause une torsion et une déformation du thorax, de l'abdomen et parfois du bassin. Chaque vertèbre affiche une orientation et une position dans l'espace ainsi qu'une certaine morphologie. Dans le plan sagittal, on peut apercevoir une cambrure, un dos plat ou creux. Dans le plan frontal, on peut constater une déviation par rapport à la ligne centrale. Dans le

plan axial, on remarque que la rotation des vertèbres est accompagnée d'une gibbosité visible lors de l'examen clinique.

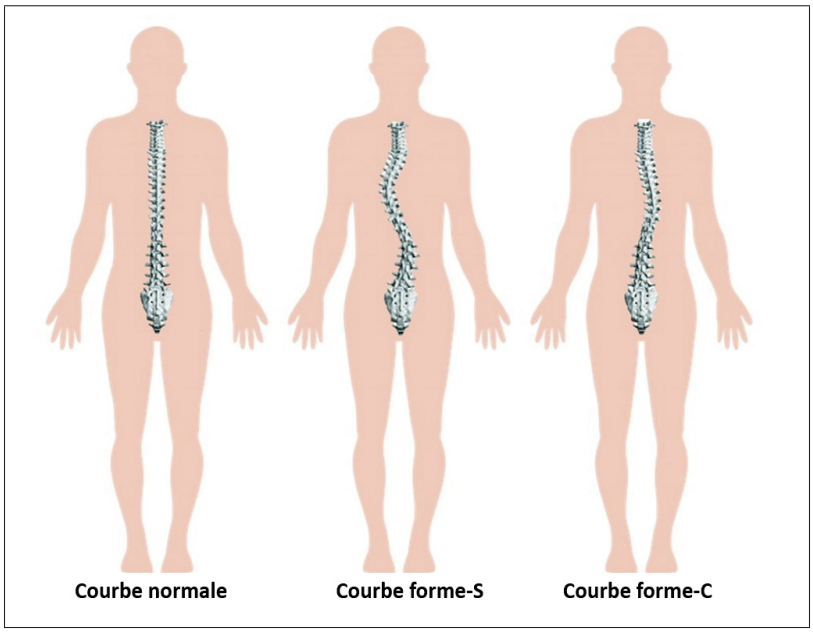


Figure 1.4 Représentation d'une colonne vertébrale saine (gauche) et deux formes de courbure scoliotique (milieu et gauche). Inspirée de Erika (2018)

Il existe différentes formes de scoliose mais la plus fréquente est la scoliose idiopathique qui représente environ 75% des cas. Les scolioses secondaires comme les scolioses dues à des malformations congénitales des vertèbres, les scolioses d'origine neuromusculaire, les scolioses dues à des maladies du tissu conjonctif ou des dystrophies osseuses sont bien plus rares.

La scoliose idiopathique peut être classifiée selon deux critères : l'âge d'apparition et le degré de sévérité. Concernant l'âge, elle est dite infantile si elle apparaît entre l'âge de 0 à 3 ans, juvénile entre 4 à 10 ans et de l'adolescent de 11 ans jusqu'à l'âge adulte. La scoliose idiopathique de l'adolescent est d'ailleurs la forme la plus répandue. Pour ce qui est de la classification selon la gravité, la scoliose est mesurée selon l'angle de Cobb (Cobb, 1948) qui est mesuré par l'intersection des lignes perpendiculaires au plateau supérieur de la vertèbre limite supérieure et au plateau inférieur de la vertèbre limite inférieure. Les vertèbres limites sont repérées par l'expert sur une radiographie postéro-antérieure et sont les deux vertèbres les plus

inclinées aux extrémités de la courbure scoliotique. Deux droites parallèles sont tracées à partir du plateau de ces vertèbres, l'angle formé par ces droites est l'angle de Cobb. Ainsi, un angle mesuré entre 10° et 25° est associé à une scoliose légère à modérée, un angle entre 25° et 35° représente une scoliose modérée à sévère et un angle au-dessus de 35° est indicatif d'une scoliose sévère à très sévère. Ainsi, l'angle de Cobb définit la sévérité de la courbe. Toutefois, aucune mesure ne permet actuellement d'estimer de façon fiable le risque de progression de la courbure scoliotique.

1.1.2.1 Traitement de la scoliose

Le traitement de la scoliose dépend de son degré de sévérité. Pour les cas de scoliose légère, aucun traitement n'est requis. Le patient doit effectuer un suivi par radiographie aux 6 mois afin d'observer si la scoliose s'accentue ou non. Pour ce qui est de la scoliose de gravité modérée, on se tourne vers un traitement non chirurgical. Pour cela, un corset orthopédique est très souvent prescrit. Il a pour principal but de stopper l'évolution de la déformation et de la réduire si la scoliose n'est pas avancée (on parle de 20° à 40° de courbure selon l'ongle de Cobb). On constate très souvent de meilleurs résultats si la marge de croissance du patient est encore grande. Si la scoliose est sévère ou le traitement non chirurgical n'a pas donné les résultats espérés, le recours à la chirurgie est envisagé.

La spondylodèse (ou fusion vertébrale) est une opération chirurgicale invasive souvent adoptée pour freiner la courbure et traiter une scoliose sévère. L'opération consiste à fusionner deux ou plusieurs vertèbres entre elles. Ceci se fait en enlevant les DIVs et en y greffant des tissus osseux prélevés du pelvis (bassin). Ensuite une procédure appelée dérotation de la scoliose est employée. Cela consiste à faire des petits trous dans le pédicule de chaque vertèbre au niveau supérieur et inférieur des deux extrémités droites et gauches et d'y poser des vis pour faire passer des tiges. L'effet immédiat de l'opération est idéalement de corriger et de stopper la progression de la déviation grâce à la fusion. Un suivi post-opératoire est nécessaire pour contrôler le redressement de la colonne grâce à la procédure de dérotation.

Il existe aussi d'autres types de chirurgies qui ne requièrent aucune fusion de vertèbres. La scoliose est caractérisée par une hypercyphose qui est une courbure dorsale due à une surcroissance au niveau antérieur qui cause une difformité de la colonne. Cette déformation peut être corrigée par une réduction antérieure comme le recours à une discectomie qui consiste à réaliser une résection de certains DIVs pour relâcher la pression créée par la torsion. Le grand intérêt de la discectomie est qu'elle peut être réalisée par thoracoscopie, une modalité chirurgicale minimalement invasive. Là encore, un suivi post-opératoire est requis pour contrôler et suivre le redressement de la colonne vertébrale.

1.1.2.2 Système de navigation par ordinateur

Les systèmes de navigation par ordinateur dans le domaine médical offrent de nouvelles possibilités aux experts durant la chirurgie pour avoir une meilleure visualisation de la structure anatomique et acquérir des informations quantitatives en temps réel. Durant une discectomie, le chirurgien est amené à réduire une certaine quantité de tissus des disques selon le degré de difformité et il serait important d'avoir en temps réel des informations quantitatives sur les DIVs et les CVs comme le pourcentage de tissus restant, taille, position et orientation des deux structures ... etc. Dans ce contexte, la reconstruction 3D de la colonne vertébrale est un excellent moyen de visualisation.

La pratique des chirurgies minimalement invasives amène des défis pour le développement de systèmes d'assistance. Les incisions faites sont très petites par rapport à une approche totalement invasive. Souvent, le chirurgien ne peut faire passer qu'une seule caméra endoscopique ne permettant pas une visualisation complète. Aussi, en orthopédie, des objets métalliques (les vis et les tiges) sont souvent requis mais limitent encore plus l'accès aux structures anatomiques. De ce fait, la visualisation 3D aide à maximiser les chances de réussite en limitant l'erreur humaine.

Dans le cas de traitement de la scoliose, à l'égard des structures anatomiques sur lesquelles le chirurgien se focalise, un système de navigation chirurgical se doit d'acquérir un modèle 3D

pré-opératoire des DIVs et des CVs. Pour obtenir ce modèle 3D, il est primordial d'extraire d'abord les structures d'intérêt et de délimiter leurs contours dans l'image, ce qui revient à la segmentation des DIVs et CVs. Une méthode de segmentation automatique, rapide, exacte et précise joue donc un rôle très important dans la qualité du système de navigation. Dans ce contexte, l'IRM constitue une modalité de choix pour obtenir l'information nécessaire à la reconstruction du modèle pré-opératoire 3D. En raison de la posture de la colonne vertébrale lors de la prise de l'IRM qui n'est pas semblable à celle en position allongé sur la table d'opération, une radiographie de la colonne vertébrale est acquise juste avant le début de la discectomie et recalée sur le modèle 3D pré-opératoire. Pendant l'opération, ce même modèle se mettra à jour en temps réel à l'aide des images qui proviennent de la caméra endoscopique grâce à un recalage 3D/2D pour obtenir une visualisation de la colonne vertébrale lors de la chirurgie.

1.1.3 Imagerie par résonance magnétique

L'IRM est une modalité d'imagerie médicale utilisée le plus souvent à des fins de diagnostique. Elle n'expose pas le patient à des radiations nocives comme la tomodensitométrie. Le patient est placé dans un puissant champ magnétique qui va stimuler les noyaux d'hydrogène trouvés dans les molécules d'eau du corps humain et orienter leurs aimantations propres (ou spins) dans le même sens. Ensuite, des impulsions electro-magnétiques appelées *ondes radio-fréquence* sont émises puis relâchées dans le but de détourner les spins de leur alignement afin de générer un signal de *résonance magnétique* qui sera capté par l'appareil. Grâce à des algorithmes de reconstruction d'images, il est alors possible de former des images 3D représentant les propriétés spécifiques des tissus relativement à ces stimuli électro-magnétiques. L'IRM est souvent utilisée pour la planification des chirurgies ainsi que dans plusieurs recherches cliniques dans l'étude de la scoliose. Par exemple, Guo *et al.* (2003) ont étudié et comparé la croissance différentielle anormale des composants antérieurs et postérieurs des vertèbres thoraciques chez des patients scoliotiques. Birchall *et al.* (2005) ont utilisé l'IRM pour étudier le degré de torsion des vertèbres chez des patients scoliotiques.

1.2 Revue des connaissances en apprentissage profond

L'apprentissage automatique est un domaine de l'intelligence artificielle dont l'objectif est d'entraîner des modèles à résoudre un ou plusieurs problèmes à partir d'un ensemble de données dit d'apprentissage présentées sous forme d'exemples. L'apprentissage peut se faire de manière *non supervisée*, *supervisée* ou parfois d'une combinaison des deux. On parle d'apprentissage supervisé lorsqu'un exemple représente un couple d'une donnée x_i et de son étiquette y_i. Le but est d'entraîner un modèle à apprendre une fonction de prédiction à partir d'un ensemble d'observations assez représentatif pour lui permettre de prédire l'étiquette réelle d'une donnée de test non observée appartenant à la même population. On parle de résoudre un problème de régression si le but est d'estimer une valeur continue ou de classification si l'ensemble des valeurs en sortie est fini. Par exemple dans ce mémoire, le problème de segmentation peut être considéré comme une classification supervisée parce que nous disposons de trois classes (Arrière plan (AP), CV, DIV) et un voxel d'une image ne peut appartenir qu'à une seule d'entre elles. Un classifieur est évalué par sa capacité à prédire correctement qu'une donnée x_i soit associée à sa classe y mais surtout à généraliser sa solution à toute la population. La généralisation est estimée par le *biais* et la *variance* du classifieur. Le biais est défini par la différence entre la prédiction moyenne faite par le classifieur sur les données de validation et la valeur réelle. La variance est définie par la sensibilité du classifieur aux fluctuations et à varier sa prédiction pour une observation lors de l'entraînement. Une variance élevée est synonyme d'un modèle qui n'est pas capable de trouver une réelle relation entre les données. Dans le meilleur des cas, un classifieur bien entraîné a un rapport biais-variance bas, on pourrait donc parler d'un modèle qui se généralise bien. Dans le cadre de notre projet, le scénario idéal serait qu'on arrive à entraîner un classifieur qui soit capable de reconnaître les DIVs et CVs avec exactitude quelque soit la modalité IRM et quelque soit le degré de sévérité de la scoliose. Cependant, les classifieurs sont sujets à deux phénomènes à éviter : le sur-apprentissage et le sous-apprentissage. Le premier se caractérise par l'incapacité à être performant face à des données non observées pendant l'entraînement. Dans ce cas, le biais est donc élevé. Quant au sous-apprentissage, c'est tout simplement l'incapacité du classifieur à cerner les données d'apprentissage. Dans ce cas,

c'est la variance qui est élevée. Ceci peut se manifester par un manque de données pour l'entraînement ou bien par un classifieur complexe. Celle-ci est liée au nombre de paramètres qu'il comporte.

L'apprentissage automatique se décline en une grande palette d'algorithmes. Ces algorithmes se situent sur un spectre dont les extrêmes correspondent à deux catégories : *paramétrique* et *non paramétrique*. Les algorithmes paramétriques sont plus traditionnels. Deux des méthodes les plus connues de cette catégorie sont les machines à vecteurs de support (SVMs) et le perceptron. Ces algorithmes sont connus pour être limités dans leur capacité d'apprentissage car le nombre de paramètres est à déterminer manuellement au préalable. Les méthodes non-paramétriques sont plus sophistiquées. Le nombre de paramètres s'adapte par rapport au nombre de données d'entraînement. Plus celui-ci augmente et plus la capacité du modèle augmente aussi. Théoriquement, le nombre de paramètres peut aller jusqu'à l'infini. Ceci donne à ces méthodes une meilleure flexibilité et leur permet d'aboutir à des modèles qui ont une meilleure portée de généralisation. Notre projet est consacré à l'apprentissage statistique-probabiliste, plus particulièrement aux réseaux de neurones dans leur version non paramétrique.

Le but des algorithmes d'apprentissage statistique-probabilistes est de modéliser la distribution des classes Y conditionnée aux données X. On note cette distribution $p(y_i|x_i;w)$ où w correspond aux paramètres du modèle. Concrètement, l'algorithme a comme objectif de maximiser l'espérance de la distribution de sortie du modèle $p_{modele}(y|x;w)$ sur l'ensemble des données d'entraînement X et leurs étiquetages Y. C'est un processus itératif qui vise à maximiser la vraisemblance des données d'apprentissage. Cette maximisation revient à décrire une fonction objectif J sur ces données (Goodfellow *et al.*, 2016). On note :

$$J(w) = E_{X,Y\sim P_{donnes}} L(X,Y,w) \tag{1.1}$$

La fonction de coût L est employée pour déterminer le taux d'erreur de prédiction que le modèle réalise sur les données d'entraînement. w est le vecteur des paramètres. Le processus

d'apprentissage tend à minimiser l'erreur en ajustant les paramètres :

$$w^* = \underset{w}{\operatorname{argmin}} J(w) \tag{1.2}$$

L'optimisation des paramètres d'un modèle peut se faire de plusieurs façons. En apprentissage automatique, la méthode la plus populaire est l'algorithme de la descente du gradient introduit par Cauchy (1847). C'est un processus itératif qui calcule à l'instant t le gradient de la fonction J par rapport au vecteur des paramètres w_t et avance dans sa direction opposée jusqu'à atteindre un minimum mais sans garantie qu'il soit global :

$$\Delta = -\nabla_w J(f(x), y) \tag{1.3}$$

où $f(x)$ est la prédiction du modèle, Δ est la dérivée de la fonction f qui informe sur la direction et l'amplitude de son gradient. Dans le développement présenté jusqu'ici, aucun élément n'indique la distance à parcourir. Pour se faire, un hyperparamètre est introduit à la descente du gradient qui va représenter le taux d'apprentissage α par itération. Le calcul du nouveau vecteur de paramètres se fait donc comme ceci :

$$w_{t+1} = w_t + \alpha\Delta \tag{1.4}$$

En réalité, le choix de α est crucial. Celui-ci détermine la vitesse de convergence. Une valeur trop grande conduit à des changements trop brusques où le gradient passe de convergence à divergence sans atteindre des optima. Une valeur trop petite permet théoriquement d'atteindre un optimum mais prendrait un temps non négligeable.

L'algorithme de la descente du gradient a inspiré plusieurs méthodes d'optimisation. D'un point de vue pratique, sa version simpliste est très limitée. Celle-ci est appliquée seulement après avoir passé l'ensemble des données d'apprentissage au modèle, ce qui conduit à un seul calcul du gradient par itération. L'entraînement est donc très lent, voire impossible si l'ensemble des données ne peut être retenu en mémoire. L'algorithme de descente du gradient stochastique a contourné cela en optimisant les paramètres après l'entraînement de chaque

exemple. La convergence est ainsi plus rapide et permet d'explorer d'autres régions pour espérer trouver un meilleur optimum local. Cependant la convergence vers le minimum global se complique. L'algorithme de descente du gradient par mini-batch a été proposé pour bénéficier des avantages des deux approches précédemment citées en réalisant l'optimisation après l'entraînement sur n données d'apprentissage. Toutefois, la descente du gradient par mini-batch ne résout pas toutes les limitations. Le choix du taux d'apprentissage reste empirique et une seule valeur ne peut pas convenir à toutes les étapes d'optimisation des paramètres où le comportement du gradient dans les zones creuses reste imprévisible. On remarque donc des oscillations des paramètres sans convergence efficace qui tendent vers une stagnation.

Pour réduire ce problème, le *momentum* (Qian, 1999) est introduit dans les algorithmes de la descente du gradient, ce qui permet au modèle de garder un historique des directions précédentes du gradient. Le momentum améliore la qualité de l'optimisation car le calcul de la dérivée de la fonction de coût n'est pas tout à fait exact. Il est plutôt estimé sur un batch de données. La direction du gradient n'est donc pas optimale. Ajouter un terme qui indique le calcul de la dérivée sur les anciennes décisions permet une meilleure estimation.

L'évolution de l'apprentissage automatique a mené vers des modèles dotés d'une très large capacité mais dont l'entraînement est difficile. De nombreuses études se sont alors penchées sur une optimisation mieux guidée notamment grâce au taux d'apprentissage adaptatif. Logiquement, ce dernier devrait augmenter lorsque le gradient est grand ce qui signifie qu'il prend la direction vers une zone d'optimum et inversement dans le cas contraire pour éviter la divergence. Dans cette optique, Kingma & Ba (2015) ont proposé Adam, un algorithme d'optimisation qui nous intéresse tout particulièrement car c'est celui que nous utilisons dans notre projet. Adam estime un taux d'apprentissage α_i différent pour chaque paramètre w_i en estimant la moyenne et la variance des gradients précédents. Les auteurs ont aussi proposé une façon de limiter l'influence des premières dérivées calculées lors des premières itérations. (Voir plus de détails en Annexe I).

1.2.1 Réseaux de neurones artificiels

Il y a déjà de cela de nombreuses années, les processeurs ont surpassé le cerveau humain pour résoudre des calculs complexes et en des temps records qui ne font que diminuer. Cependant, la perception innée de l'humain est extrêmement difficile à reproduire numériquement. Les réseaux de neurones sont des modèles qui s'inspirent du fonctionnement du cerveau humain pour arriver à doter un processeur d'une telle perception.

1.2.1.1 Du neurone au réseau neuronal

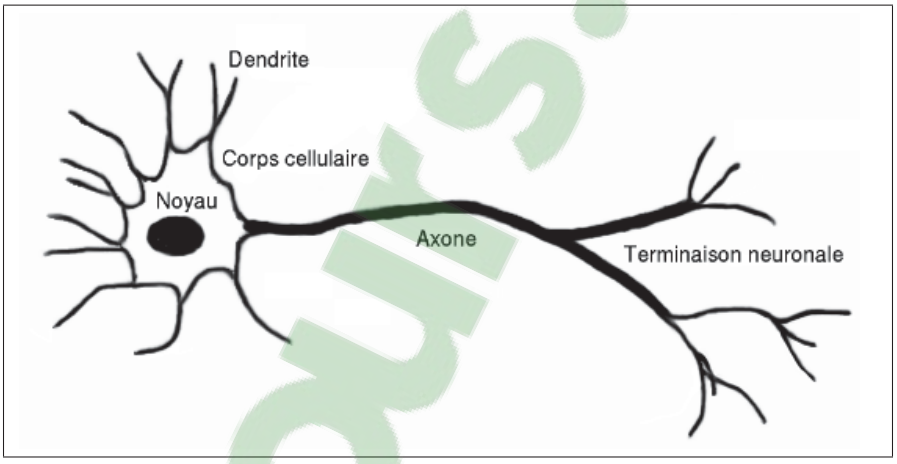


Figure 1.5 Neurone biologique. Inspirée de Chevalier (2017)

Un neurone biologique (figure 1.5) est constitué d'un corps cellulaire doté d'un noyau et des prolongements appelés *dendrites*, d'une fibre nerveuse appelée *axone* et d'une zone synaptique constituée de *synapses*. Un neurone génère un signal électrique et le transmet à un autre neurone via les synapses. Si le signal est assez stimulant, les dendrites du noyau récepteur vont s'activer à le recevoir et le retransmettre de la même manière à un autre neurone en lui faisant parcourir son axone. Les neurones artificiels (figure 1.6) sont connectés entre eux et se transmettent l'information de la même manière. La sortie d'un neurone n'est que la somme pondérée de ses entrées. Une fonction d'activation est appliquée à cette sortie pour estimer si le signal est assez stimulant et nécessite d'être propagé. Mathématiquement la sortie d'un

neurone artificiel est :

$$y = f(\sum_{i}^{N} x_i \cdot w_i + b) \tag{1.5}$$

où x_i est une donnée numérique en entrée du réseau pondérée par le poids associé w_i. b est un terme appelé *biais* dont le rôle est de translater la fonction d'activation f.

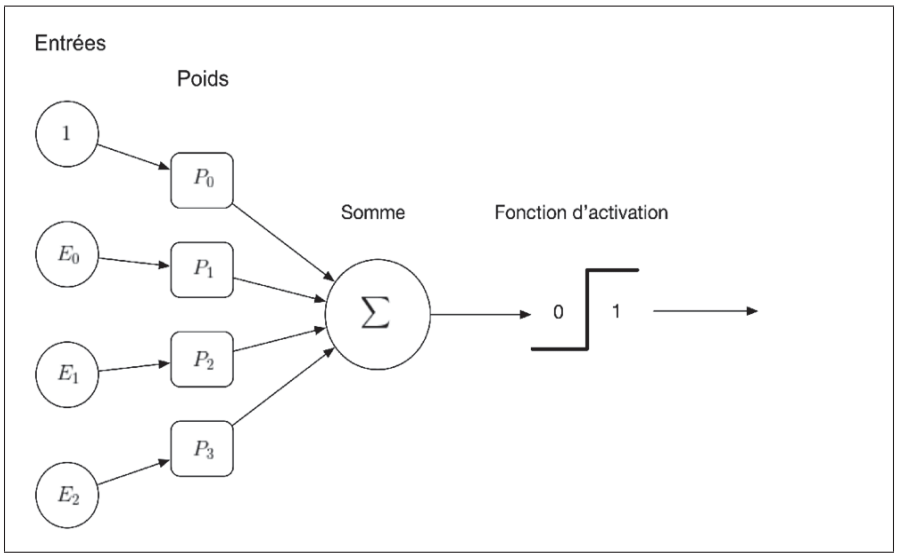


Figure 1.6 Neurone artificiel

Le neurone artificiel à lui seul n'est pas d'une grande utilité. Par contre, l'idée a fait émerger l'apprentissage automatique et une grande panoplie d'algorithmes ont été développés avec comme idée de passer d'un neurone à un réseau de neurones. Celui-ci s'illustre dans la famille des algorithmes d'apprentissage statistique paramétrique. On peut définir mathématiquement le réseau de neurones comme une fonction f paramétrée par des poids w tel que : $f_w(x) = y$. Il est composé d'une couche de données en entrée connectée à une couche cachée qui contient un certain nombre de neurones qui est connectée à son tour à une couche de sortie. Ceci a évolué vers un Perceptron multi-couche (MLP). On peut l'exprimer comme une succession de fonctions non linéaires mises en cascade $f_w(x) = g_{w_n}(x_n) \circ g_{w_{n-1}}(x_{n-1}) \circ ... \circ g_{w_0}(x)$, où $g_{w_j}(x_j)$ correspond à la j^e couche cachée. Nous pouvons aussi voir un réseau de neurones comme un extracteur de *caractéristiques* qui ne sont que les résultats des transformations subies par les données en entrées x. Cependant, plus un réseau de neurones a de couches cachées, plus il y a

de paramètres à optimiser et plus l'entraînement est difficile. Les réseaux de neurones étaient juste dotés d'une plus grande capacité que leurs prédécesseurs, mais sans connaissance sur la manière de les entraîner efficacement. Il aura fallu attendre de meilleurs processeurs mais surtout l'algorithme de rétropropagation proposé par Rumelhart *et al.* (1988), une élégante manière de propager la fonction de coût (Équation 1.1) sur l'ensemble des poids et les optimiser pour que les réseaux de neurones puissent dévoiler leur potentiel. Les recherches menées par Cybenko (1989) ont démontré la capacité d'un réseau de neurones à résoudre n'importe quelle fonction continue dans un espace fermé à condition que le modèle soit doté d'une grande capacité. Ceci revient à une de leurs propriétés qui leur permet d'extraire des caractéristiques par eux-mêmes et ouvre la voie à leur utilisation dans divers champs d'applications. En traitement d'images, Lecun *et al.* (1998) ont introduit ce qui nous intéresse particulièrement dans nos travaux : un premier réseau de neurones convolutif (CNN) dont l'architecture a été baptisée LeNet5, un algorithme permettant de classifieur avec une grande précision des images de chiffres manuscrites. Par la suite, les avancées apportées par Hinton *et al.* (2006) ont permis d'entraîner efficacement des réseaux de neurones profonds où plusieurs couches cachées se succèdent en formant une représentation non linéaire dont la complexité augmente au fur et à mesure qu'on avance en profondeur.

1.2.1.2 Réseaux de neurones convolutifs et apprentissage profond

Les CNNs partagent le mécanisme global d'un MLP. Les CNNs sont définis par un ensemble de couches formant une hiérarchie de caractéristiques (Figure 1.7). L'opération principale dans un CNN est la *convolution*. Celle-ci a pour rôle d'extraire les caractéristiques de la couche qui la précède et produit le résultat sous forme d'une *carte de caractéristiques*. Le paragraphe qui suit explique ce mécanisme.

Une image 2D est perçue par l'ordinateur comme une matrice où chaque élément représente une intensité de pixels. Cette matrice peut être en 3 dimensions dans le cas où l'image est en couleur. Chaque dimension représente un canal qui correspond à soit du rouge, du bleu ou du vert. En prenant une image d'entraînement en entrée, le réseau produit des opérations

de convolution en utilisant un certain nombre de filtres sur chaque canal avec un certain *pas*. Chaque filtre est une matrice dont les dimensions sont fixes dans chaque couche. Ils parcourent l'image ou la carte en entrée et produisent chacun un canal en sortie où le résultat de chaque convolution est enregistré dans l'emplacement qui lui correspond.

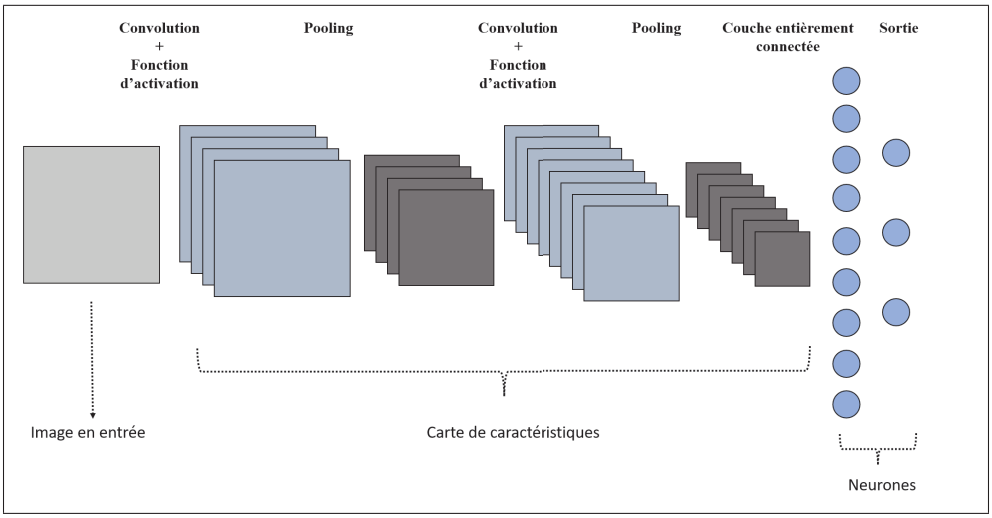


Figure 1.7 Exemple d'une architecture CNN

La carte de caractéristiques correspond donc à toutes les sorties produites par les filtres et aura comme dimensions $W_{i+1}H_{i+1}C_{i+1}$, avec :

$$W_{i+1} = \frac{W_i - F + 2P}{S} + 1 \tag{1.6}$$

$$H_{i+1} = \frac{H_i - F + 2P}{S} + 1 \tag{1.7}$$

$$C_{i+1} = K \tag{1.8}$$

où F est la taille du filtre, S est le pas de la convolution, P est le *Padding* qui permet de contrôler la dimension spatiale de la carte en sortie, K est le nombre de filtres utilisés. À partir de ces informations, nous pouvons reconnaître que les CNNs ont certaines propriétés :

- *Connectivité peu dense :* Chaque neurone est connecté à un sous-ensemble de l'image en entrée ou carte de caractéristiques.

- *Partage des poids :* Malgré que les neurones ne soient pas connectés entre eux, ils partagent tout de même des poids.
- *Invariance à la translation :* Ceci est une conséquence de l'opération de convolution.

En soi, l'opération de convolution est une opération linéaire. Pour obtenir des caractéristiques qui représentent une transformation non-linéaire des données en entrées, une fonction d'activation non linéaire est appliquée sur la carte des caractéristiques après chaque opération de convolution. Le choix de la fonction d'activation est critique. Un mauvais choix conduit au problème de disparition du gradient (Goodfellow *et al.*, 2016) ce qui rend le réseau incapable d'optimiser ses paramètres. Le choix de la fonction d'activation doit donc être stratégique pour éviter cet effet de stagnation. En ce qui concerne les CNNs, l'utilisation de la fonction d'activation *Rectified Linear Units* (*ReLU*) introduite par Nair & Hinton (2010) dans les couches cachées est très fréquente. La fonction *ReLU* et sa dérivée sont définies comme suit :

$$ReLU(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases} \qquad ReLU'(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$$

La popularité de la *ReLu* est dû à sa simplicité. D'abord d'un point de vue calcul, la dérivée nécessite très peu d'opérations. De plus, la fonction désactive les paramètres ayant une valeur nulle ou négative. Ceci rend le modèle parcimonieux ce qui met en relief les informations pertinentes et réduit le sur-apprentissage. Par contre, le fait d'avoir des sorties à 1 pour chaque valeur positive implique une moyenne d'activation plus grande que zéro. Il a été prouvé que quand il s'agit de réseaux de neurones, centrer la moyenne à zéro accélérerait l'apprentissage (LeCun *et al.*, 1991). Aussi, le problème qui se pose est que les paramètres mis à zéro ne seront plus mis à jour lors d'itérations subséquentes de l'optimiseur. Pour remédier aux problèmes du *ReLU*, des fonctions d'activation qui en sont inspirées ont été proposées. Nous détaillerons dans cette partie celle que nous avons utilisée : *Exponential Linear Units* (*ELU*) introduite par Clevert *et al.* (2015).

$$ELU(x) = \begin{cases} \zeta(e^x - 1) & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases} \qquad ELU'(x) = \begin{cases} ELU(x) + \zeta & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$$

ELU est similaire à la *ReLU* si la valeur est positive. Par contre, si la valeur est négative, celle-ci n'est pas mise systématiquement à zéro. *ELU* peut prendre des valeurs négatives ce qui permet de pousser la fonction vers une moyenne de zéro. Le paramètre ζ permet de contrôler l'échelle de la partie négative. Par défaut sa valeur est mise à 1.

Pour ce qui est de la couche de sortie, le choix de la fonction d'activation doit correspondre à la fonction de coût qui est choisie en fonction de la tâche que le modèle est entraîné à réaliser. En classification binaire, la fonction d'activation *sigmoïde* notée σ associée à *l'entropie croisée* notée L_{CE} comme fonction de coût sont souvent adoptées.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{1.9}$$

Elle prend des valeurs sur l'intervalle $[0,1]$ ce qui est idéal dans le cas où le modèle doit prédire la sortie comme une loi probabiliste. La fonction sigmoïde peut avoir une forte confiance en sa prédiction. Plus x tend vers un grand nombre positif et plus la fonction *sigmoïde* tendra vers 1. Cela est paradoxalement vrai pour le cas négatif où la prédiction tend vers 0. Ceci engendre un problème majeur que l'on peut visualiser à partir de la figure 1.8 qui représente la fonction et sa dérivée qui est définie comme suit :

$$\sigma'(x) = \sigma(x)(1-\sigma(x)) \tag{1.10}$$

Plus la fonction sigmoïde est confiante en sa prédiction plus le gradient est faible ce qui implique une faible mise à jour. En tout début d'entraînement, l'initialisation aléatoire des poids peut causer un tel scénario ce qui ralentit la convergence. En classification binaire, l'utilisation de la fonction sigmoïde dans la couche en sortie accompagnée de la fonction de l'entropie croisée comme fonction de coût est grandement adopté. Cela nous informe directement sur la distance entre la prédiction et l'étiquette (voir preuve en annexe I).

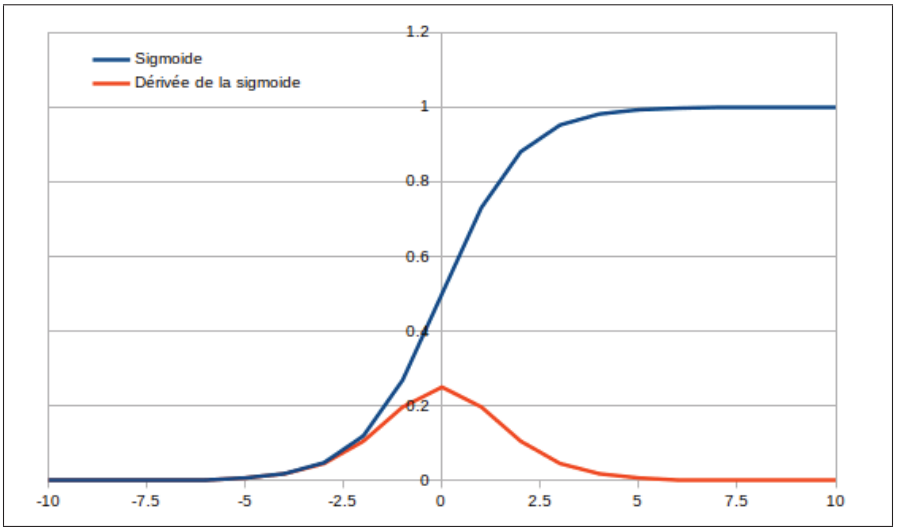


Figure 1.8 Graphe de la fonction sigmoïde et sa dérivée

Dans le cas multi-classes, on se réfère par contre à la fonction d'activation $softmax$ qui est une généralisation de la fonction *sigmoïde* définie comme suit :

$$softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}} \tag{1.11}$$

L'intérêt ici est que le calcul de la probabilité d'appartenance à chaque classe pour une donnée x_i ne soit pas indépendant et que la somme des probabilités soit égale à 1. Le sous-échantillonnage de la carte de caractéristiques appelé *pooling* est aussi une opération de base très importante dans les CNNs. Traditionnellement, la méthode du *maxpooling* est employée par défaut. L'opération consiste à sous-échantillonner la carte en entrée par un facteur en parcourant une fenêtre de taille fixe avec un pas défini en prenant dans chaque région la valeur maximale.

1.2.1.3 Régularisation

Les réseaux de neurones sont facilement sujets au problème de sur-apprentissage. Des stratégies existent pour limiter ce phénomène. Nous nous intéressons ici particulièrement à la régularisation-L2 et au dropout, méthodes auxquelles nous avons eu recours dans notre projet.

La régularisation en général est une technique qui est censée réduire la complexité du modèle en pénalisant la fonction de coût. La régularisation-L2 ajoute un terme à la fonction de coût qui représente la somme pondérée de tous les paramètres au carré. Celle-ci favorise des paramètres de faible valeur sans pour autant leur donner une valeur nulle. :

$$L' = L + \lambda \sum_{i=1}^{n} w_i^2. \tag{1.12}$$

Le dropout a été introduit par Hinton *et al.* (2012). C'est une méthode très simple et efficace de régularisation spécialement dédiée aux réseaux de neurones profonds. L'idée est d'ignorer selon une probabilité aléatoire des neurones durant l'entraînement, ce qui veut dire qu'ils ne contribuent ni à la propagation avant ni à la rétropropagation. L'effet du dropout est de rendre le réseau moins sensible à des paramètres spécifiques et de l'aider à avoir une meilleure capacité de généralisation. Le dropout est souvent comparé aux algorithmes de boosting (Freund *et al.*, 1999) parce qu'au file des itérations, le réseau se voit attribuer des neurones différents ce qui lui donne une certaine représentation de plusieurs architectures en un seul modèle.

1.2.1.4 Transfert d'apprentissage

Le transfert d'apprentissage est une technique d'apprentissage automatique dans laquelle un modèle entraîné à effectuer une tâche est réutilisé comme point de départ d'un second entraînement dans le but d'effectuer une deuxième tâche connexe. En apprentissage profond, cette technique est souvent adoptée pour réduire la nécessité d'avoir une grande base d'entraînement car les paramètres sont déjà pré-entraînés à condition que les caractéristiques extraites durant les deux entraînements soient partagées. (Goodfellow *et al.*, 2016).

1.2.1.5 Réseaux de neurones convolutifs pour la classification d'images

Les CNN ont dévoilé leur potentiel grâce à Krizhevsky *et al.* (2017) qui ont introduit le réseau AlexNet. Celui-ci englobe les fondements de base d'un CNN tel qu'introduit par LeNet5 Lecun *et al.* (1998). Cependant, l'architecture est plus profonde et les images en entrée sont 8 fois

plus grandes. Le réseau est aussi doté des avancées en matière de fonction d'activation (*ReLU*) et de régularisation (*Dropout*). De plus, grâce aux calculs des réseaux de neurones devenus possibles sur des processeurs graphiques GPUs, le réseau a été entraîné avec une immense base de données appelée *imageNet* qui contient des millions d'images étiquetées. L'impacte de AlexNet a conduit à une importante motivation de la communauté académique à se focaliser sur les réseaux de neurones.

Les CNNs se sont montrés très robustes pour résoudre des problèmes de classification d'images et plusieurs architectures ont été proposées pour accroître leur performance. Zeiler & Fergus (2013) ont proposé le *ZFNet* qui n'est qu'un réseau AlexNet mais avec un paramétrage différent. Ils ont ainsi démontré l'importance du choix des valeurs des hyper-paramètres. Simonyan & Zisserman (2014) ont proposé le *VGG-net* qui était à ce moment-là l'une des architectures les plus profondes. Ils ont démontré que la profondeur d'un réseau est critique pour atteindre de meilleures performances à condition d'avoir une importante base de données d'entraînement. Plusieurs études se sont basées sur VGG-net pour extraire des caractéristiques d'une image (Badrinarayanan *et al.*, 2015). Szegedy *et al.* (2015) ont introduit le module *inception* via l'architecture *GoogleNet* dont le principe est de réaliser en parallèle plusieurs opérations de convolution avec des tailles de filtres variables et de les fusionner dans la prochaine couche. L'objectif de GoogleNet est de réduire la complexité de calcul en comparaison aux CNNs traditionnels en limitant significativement le nombre de paramètres. Présenté la même année lors de la compétition ILSVRC 2014, GoogleNet a remporté le premier prix en dépassant VGG-Net et en s'approchant de très près des performances d'un expert. He *et al.* (2016) ont proposé *ResNet* dans le but de diminuer le problème de disparition du gradient. Les architectures précédemment citées ont réussi à atteindre les meilleures performances dans l'état de l'art mais leur utilisation a un prix : la quantité de données d'entraînement. Des modèles avec une si grande capacité ont besoin de constamment apprendre. Dans le cas contraire, le gradient diminue à chaque fois jusqu'à causer sa disparition dans le réseau ce qui pénalisera l'optimisation des paramètres. Le ResNet est doté de *connexions résiduelles*. La sortie d'un bloc résiduel

peut être définie comme :

$$x_{l+1} = R(x_l) + x_1 \tag{1.13}$$

où $R(x_l)$ est la sortie d'un ensemble varié d'opérations et x_l est la sortie de la couche cachée précédente. Comme dans la figure 1.9, on peut voir que la fonction R est la sortie de deux convolutions et une ReLu. Ces connexions résiduelles permettent de créer des raccourcis au gradient, transportant des informations d'une couche à une autre. Aussi surprenant que cela puisse paraître, une opération aussi simple permet d'entraîner des très profonds réseaux, réduire le problème de disparition du gradient et gagner en capacité de généralisation. Au vu de cela, des études s'en sont inspirées en ayant comme idée de créer des connexions courtes dans leur réseau.

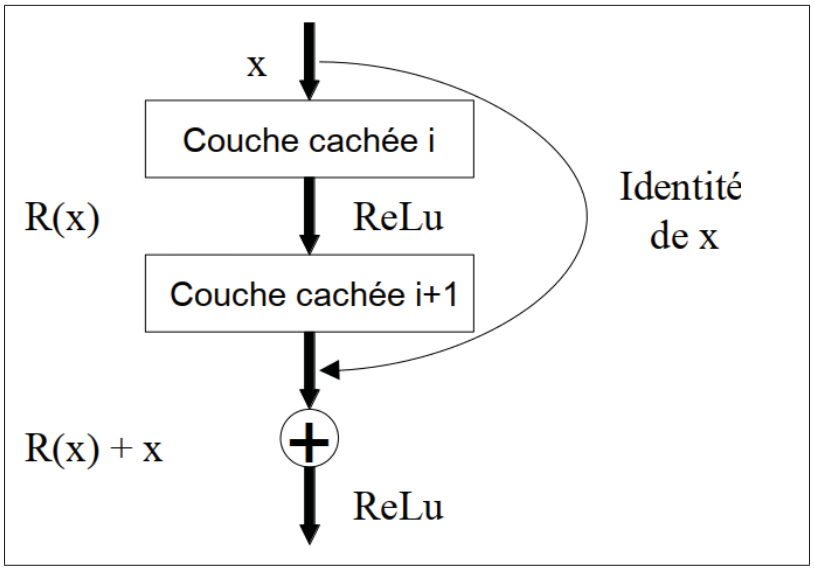


Figure 1.9 Schéma d'un bloc résiduel inspiré de He *et al.* (2016)

Huang *et al.* (2017) ont proposé le *DenseNet*. Cette architecture a comme particularité que toutes les couches soient connectées entre elles (i.e pour un réseau à L couches, il existe L(L+1)/2 connexions). Contrairement au ResNet qui somme la sortie d'une couche avec celle qui la précède, le DenseNet concatène la carte de caractéristiques de la couche cachée i avec toutes les cartes de caractéristiques des $i-1$ couches cachées précédentes. Le DenseNet s'est

montré extrêmement robuste face au problème de disparition du gradient d'erreur dépassant même le ResNet.

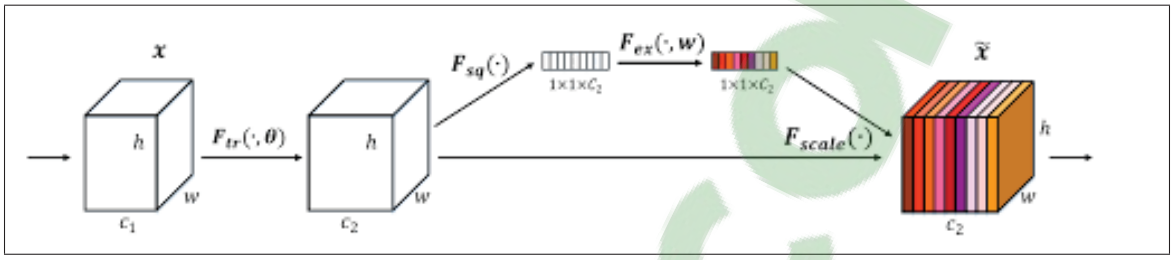


Figure 1.10 Schéma du bloc squeeze and excitation. Tiré de Hu *et al.* (2018)

Hu *et al.* (2018) ont proposé le réseau *squeeze and excitation*. Celui-ci introduit un bloc (Figure 1.10) qui permet de mieux encoder l'interdépendance entre les canaux et leur importance relative pour le problème à résoudre. Le principe général consiste à pondérer chaque canal au vu de son importance et ainsi attribuer aux caractéristiques les plus discriminantes de plus importantes valeurs et inversement pour les moins significatives. Comme son nom l'indique, une opération de compression (*squeezing*) est d'abord réalisée en transformant la carte de caractéristiques u_c de taille $H \times W \times C$ en descripteur de canaux représenté par un vecteur z de taille C, i.e une valeur numérique par canal en utilisant un *pooling moyen global* F_{sq} :

$$z = F_{sq}(u_c) = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j). \tag{1.14}$$

Ensuite vient l'opération d'excitation qui a pour but de capturer la dépendance entre les canaux. Pour cela, z passe par deux couches entièrement connectées dont la première est dotée d'une fonction ReLu γ et la deuxième d'une fonction sigmoïde σ comme fonction d'activation :

$$s = F_{ex}(z,W) = \sigma(W_2 \gamma(W_1 z)). \tag{1.15}$$

W_1 et W_2 sont les paramètres employés durant les deux opérations de convolutions. s est alors un vecteur de taille C qui représente le vecteur de pondération de la carte de caractéristiques u_c. C'est à dire que l'élément i du vecteur s représente le poids des caractéristiques de $u_{c,i}$.

D'ailleurs nous pouvons mathématiquement formuler cela comme le produit de la carte de caractéristiques u_c et le vecteur s_c

$$\hat{u}_c = u_c.s_c \tag{1.16}$$

Globalement, SE améliore grandement la représentation du réseau par une recalibration dynamique des canaux et rehausser ainsi les caractéristiques les plus discriminantes. Ceci a grandement contribué aux performances obtenues par le réseau sur différentes bases de données de validation.

1.3 Conclusion

À travers cette revue des connaissances, nous avons identifié dans un premier temps les deux structures anatomiques essentielles à segmenter pour concevoir un système de navigation par ordinateur lors de la discectomie, à savoir : Les DIVs et les CVs. Ensuite, nous nous sommes intéressés à la théorie de l'apprentissage profond et particulièrement au recours des CNNs, leur mécanisme d'apprentissage et l'évolution des architectures proposées pour réaliser une classification d'images. Dans le chapitre suivant, nous verrons plus en détail des approches basées sur les CNNs pour réaliser la segmentation des DIVs et CVs en IRM.

CHAPITRE 2

REVUE DE LA LITTÉRATURE

L'objectif général de ce projet est de réaliser la segmentation simultanée des DIVs et CVs à partir d'IRM sur le plan sagittal de patients atteints de scoliose en ayant recours à l'apprentissage profond. À cet égard, nous présentons dans ce chapitre une revue de littérature approfondie des méthodes de segmentations des IVDs et CVs à partir d'IRM (Section 2.1) en introduisant à la fois des méthodes classiques de segmentation (Section 2.1.1) et des méthodes basées sur l'apprentissage profond (Section 2.1.2)

2.1 Revue des méthodes de segmentation de la colonne vertébrale

La segmentation est la tâche d'identifier un ensemble de pixels comme appartenant à un objet dans une image. On parle de segmentation sémantique dans le cas où les objets d'intérêt sont multiples et il est nécessaire d'identifier leur classe. En imagerie médicale, la segmentation est utilisée pour retrouver des organes et autres structures anatomiques, tumeurs, ect. dans le but d'acquérir des informations quantitatives utiles pour la prise de décisions dans des études cliniques, réalisation de diagnostics ou en chirurgie. La segmentation est une étape majeure dans tout système d'assistance par ordinateur et les performances de ces systèmes y sont étroitement liées. Bien plus importante qu'une classification d'image ou une simple détection d'objet, la frontière entre des structures anatomiques est une information cruciale. D'ailleurs, la première difficulté à laquelle on fait face en segmentation est la définition de cette frontière qui peut s'avérer difficile à retrouver à l'oeil nu.

2.1.1 Méthodes classiques de détection et segmentation de la colonne vertébrale

Plusieurs méthodes de détection et segmentation des structures osseuses de la colonne vertébrale à partir d'IRM ont été proposées dans la littérature. Les CVs et les DIVs sont principalement au coeur de ces études. Cependant, la validation sur des images de patients scoliotiques

est quasiment inexistante. La difficulté d'accéder à des données d'une pathologie bien spécifique explique en grande partie ce vide. Des méthodes qui segmentent les deux structures à la fois sont très rares. Les premières approches se focalisent principalement sur une segmentation 2D sur le plan sagittal en raison de la forme rectangulaire des deux structures plus facilement repérable sur ce plan-là. La segmentation en 3D est survenue avec l'évolution des méthodes classiques et surtout l'apprentissage profond.

2.1.1.1 Disques intervertébraux

Nous retrouvons dans la littérature sur la segmentation des DIVs des méthodes automatiques et semi-automatiques. Parmi les méthodes semi-automatiques, certaines sont basées sur une segmentation par graphe. Ben Ayed *et al.* (2011) ont introduit une méthode interactive pour segmenter les DIVs en se basant sur l'algorithme graph cut (Boykov & Funka-Lea (2006)). L'algorithme tente d'optimiser une fonction de coût en se basant sur une information a priori sur l'intensité et la géométrie des DIVs dans l'image. Dans leurs expérimentations, les auteurs ont mis l'accent sur l'apport des informations a priori dans leur approche. Cette information est en réalité une interaction humaine où l'utilisateur doit tracer une ellipse pour chacun des DIVs. Les résultats ont démontré une forte dépendance de l'approche à l'objet a priori.

Nous retrouvons aussi des méthodes basées sur un atlas. Michopoulou *et al.* (2009) ont proposé une méthode de segmentation semi-automatique des DIVs sains et dégénérés au niveau lombaire. L'approche est basée sur un atlas d'un DIV créé en utilisant une technique de recalage de repères à partir d'un ensemble de régions d'intérêt centrées sur des DIVs. Les auteurs ont ensuite utilisé trois méthodes de segmentation en utilisant l'atlas qui apporterait une connaissance a priori sur l'anatomie des DIVs. Pour les segmenter, un regroupement flou est utilisé pour classifieur les pixels en se basant sur l'intensité. Dans la première méthode, l'algorithme Bezdek's FCM (Hathaway & Bezdek, 1988) est utilisé pour calculer le taux de chaque type de tissu (os, DIV, fluide cérébrospinal) dans chaque voxel. Dans la deuxième méthode, l'algorithme Bezdek's FCM est remplacé par l'algorithme Pharm-RFCM (Cannon *et al.*, 1986). La troisième méthode est semblable à la deuxième sauf qu'une déformation élastique est ajoutée

lors du recalage de l'atlas dans l'image. Les auteurs ont mené une étude comparative entre les trois approches et il s'est avéré que la troisième méthode est plus robuste suivie de la deuxième. Toutefois, celle-ci prend un temps non négligeable à s'exécuter. En plus qu'elle ne soit pas automatique, cette approche souffre d'une complexité de calcul élevée et qui force l'auteur à se tourner vers la deuxième méthode en sacrifiant la performance de la segmentation pour un temps d'exécution raisonnable. De plus, la qualité de segmentation est fortement dépendante de la formation de l'atlas qui est une étape manuelle. De ce fait, la mise en pratique de cette approche en clinique n'est pas réaliste.

Nous retrouvons aussi les méthodes de segmentation basées sur les régions. Law *et al.* (2013) ont proposé une méthode de segmentation des DIVs avec comme objectif qu'il y ait le minimum d'interaction humaine possible. Le processus consiste à détecter approximativement la région des vertèbres à l'aide d'une représentation de flux orienté anisotrope qui requiert le marquage manuel de la première vertèbre cervicale. Cette première étape est suivie d'une transformée de distance directionnelle maximale pour générer des descripteurs des DIVs dans l'image afin de former une fonction d'énergie à minimiser par un contour actif. Bien que l'interaction humaine dans cette approche soit minime comparée aux autres approches semi-automatiques, elle reste tout de même sensible au bruit présent dans l'IRM. Les résultats ont montré une sous-segmentation considérant les ligaments qui entourent les DIVs comme étant une partie d'eux.

Pour ce qui est des méthodes automatiques, nous retrouvons d'une part des méthodes ayant recours aux classifieurs et les caractéristiques manuellement choisies et extraites. Chevrefils *et al.* (2009) ont proposé une méthode de segmentation automatique des DIVs à partir d'IRM de sujets sains et de patients ayant une scoliose sur les plans sagittal et coronal. L'algorithme watersheds est d'abord employé pour obtenir une première segmentation des DIVs. Le résultat est affiné en ayant recours à un classifieur KNN qui prend en entrée un vecteur de caractéristiques de textures statistiques et spectrales. Dans leur étude expérimentale, les auteurs ont mené une étude comparative en utilisant d'autres types de caractéristiques texturales. Les résultats ont dévoilé une meilleure performance dans le cas où les caractéristiques de textures

statistiques et spectrales sont simultanément présentes dans le vecteur de caractéristiques. Cependant, les auteurs dévoilent un taux de rappel sur les images de validation relativement bas par rapport à la précision qui est plutôt élevée. Les auteurs remettent en cause l'habileté de la méthode à définir correctement les contours des DIVs en lien avec sa tendance à produire une sur-segmentation.

Oktay & Akgul (2013) ont proposé une méthode de détection des DIVs. La détection est réalisée en entraînant un classifieur SVM avec des caractéristiques HOGs suivi d'un modèle graphique probabiliste proposé pour affiner la détection. Les résultats obtenus sont très satisfaisant. Les caractéristiques HOGs décrivent l'apparence globale d'un objet (Les DIVs dans cette étude). Les DIVs sains ont une forme plutôt similaire et pas très complexe. Il est donc tout à fait possible de les décrire simplement avec des caractéristiques HOGs. Toutefois, l'étude n'est pas validée sur des DIVs déformées (Par une scoliose par exemple).

On retrouve aussi des méthodes de segmentation automatiques basées sur les régions. Chen *et al.* (2015) proposent une méthode de localisation et segmentation des DIVs. L'étape de la localisation se fait en entraînant un modèle de régression guidé par les données qui estime un vecteur de déplacement de plusieurs points aléatoires de l'image au centre de chaque DIV. La segmentation est effectuée de la même manière que la détection, à la différence que cette fois-ci le modèle de régression est remplacé par une classification de chaque pixel comme étant un DIV ou faisant partie de l'arrière-plan de l'image. Dans leur étude expérimentale, les auteurs ont d'abord montré que les performances de leur approche à réaliser la segmentation des DIVs dépassent celles obtenues avec un algorithme de forêt d'arbres décisionnels. Toutefois, certaines limitations ont été rapportées. La segmentation est fortement dépendante de l'étape de détection. Cette dernière est robuste uniquement dans le cas des tranches médianes du volume où tout le tronc est clairement visible. L'approche est très sensible aux déformations qui peuvent se rapporter aux structures rachidiennes, que cela soit à cause d'une quelconque pathologie ou dans le cas où la colonne vertébrale est tout simplement moins visible.

2.1.1.2 Corps vertébraux

Pour segmenter les CVs, nous retrouvons des méthodes semi-automatiques et automatiques. Pour la catégorie semi-automatique, nous citons Hille *et al.* (2018) qui ont proposé une méthode avec un minimum d'interaction humaine dont le but est qu'elle soit robuste pour un grand nombre de modalités IRM et différentes pathologies dont la scoliose. Grâce à des points manuellement placés sur chaque vertèbre, l'algorithme calcule les dimensions des CVs et extrait leurs caractéristiques d'intensité. Un modèle de forme d'une vertèbre est placé à l'intérieur de chacune en se fiant aux dimensions précédemment calculées. Une première segmentation est effectuée à l'aide d'un seuillage adaptatif. Ensuite, le résultat est morphologiquement filtré pour obtenir les contours des CVs. Les résultats ont montré une très bonne validité pour ce qui est des patients avec et sans pathologie. De plus, la représentation des CVs à l'aide de caractéristiques basées sur l'intensité ont permis à l'algorithme d'avoir de la flexibilité quant à son application sur diverses modalités IRM. Cependant, les modalités avec peu de contraste dans la région des CVs souffrent de sur-segmentation.

Pour ce qui est des méthodes automatiques, Peng *et al.* (2006) proposent une méthode de détection et segmentation des CVs. La détection se fait grâce à un profil d'intensité qui permet d'obtenir le centre de chaque DIV et CV en se fiant aux maxima locaux du profil. À partir de ces points et en utilisant un filtre Canny et des opérations morphologiques, les contours des CVs sont extraits. Le problème avec cette approche est sa forte sensibilité au bruit dans l'image ainsi que l'hypothèse que les DIVs et les CVs ont systématiquement des intervalles d'intensité éloignés, ce qui n'est pas toujours le cas. Le champ d'application de cette approche est fortement réduit au vu des diverses modalités d'acquisition d'IRM.

Huang *et al.* (2009) ont proposé une approche de détection et segmentation des CV en 3 étapes. La première consiste à entraîner un détecteur de CV en utilisant l'algorithme adaBoost (Freund *et al.*, 1999). Les caractéristiques choisies des images d'entraînement sont une combinaison d'une représentation par ondelettes avec l'intensité de l'image. Pour affiner le résultat du détecteur, une méthode d'ajustement de courbe spécialement conçue pour la colonne vertébrale

et proposée par Vrtovec *et al.* (2005) est employée comme post-traitement. L'idée est d'utiliser cette courbe pour éliminer les faux positifs et retrouver les CVs manquants. Enfin, après avoir détecté tous les CVs, l'algorithme de coupe normalisée (Shi & Malik (1997)) est employé pour les segmenter. L'étude expérimentale a dévoilé l'apport majeur apporté par le post-traitement lors de la détection des CVs.

2.1.1.3 Disques intervertébraux et corps vertébraux

Kelm *et al.* (2013) ont proposé une méthode de détection et segmentation qui peut s'appliquer à la fois sur DIVs et CVs à partir d'IRM ou d'images tomographiques. La détection est réalisée suite à un apprentissage à espace marginal (introduit par Zheng *et al.* (2008)). L'algorithme graph cut est ensuite employé pour retrouver les contours de la structure anatomique. La contribution majeure de cette étude est sa flexibilité par rapport à la structure et à la modalité d'images médicales. De plus, les résultats ont montré que l'approche est robuste à certaines pathologies comme la dégénescence des DIVs. Toutefois, même si l'algorithme graph cut semble robuste pour suivre les contours de la structure, les résultats qualitatifs montrent clairement que plusieurs chevauchements entre les deux structures se produisent. Aussi, l'algorithme est incapable de segmenter les deux structures simultanément. Nous voyons en cette méthode une détection rapprochée (c.à-d. plus précise qu'une fenêtre centrée) de la structure plus qu'une réelle segmentation.

Neubert *et al.* (2012) ont proposé une méthode de segmentation simultanée des DIVs et des CVs. La méthode proposée commence par localiser les centres des vertèbres grâce à un filtre de Canny et un profil d'intensité. Des modèles de forme moyenne sont alors placés sur ces points afin de réaliser la segmentation. Les résultats ont montré de bonnes performances de l'approche. Toutefois, le temps requis pour segmenter une seule vertèbre est estimé à 35 minutes. De plus, la phase de détection des vertèbres n'est pas robuste d'où le fait que les auteurs se limitent à tester leur approche sur uniquement les tranches médianes du volume où les DIVs et CVs sont très apparents. Sur les tranches où la colonne vertébrale commence à être apparente et sur celles où celle-ci disparaît, les DIVs et CVs apparaissent sous une forme différente. Ce

cas de figure ferait échouer la détection et donc la segmentation en l'occurrence. Il est aussi rapporté que la segmentation des DIVs souffre d'une sous-segmentation en se propageant jusqu'aux nerfs de la moelle épinière sur les images pondérées T2.

2.1.2 Méthodes de segmentation d'images médicales par apprentissage profond

À l'image des approches précédemment citées, les méthodes classiques de segmentation entraînent des limitations majeures qui ne favorisent pas leur déploiement en clinique pour effectuer des tâches critiques. Si une approche est performante, elle nécessite un temps non négligeable d'exécution ou une interaction humaine. Si l'approche est automatique et ne nécessite pas un mécanisme d'apprentissage, elle est très sensible au bruit dans l'image. Pour les méthodes nécessitant un mécanisme d'apprentissage, une phase de détection est quasiment obligatoire car les performances de la segmentation y sont étroitement liées. De plus, le choix des caractéristiques à extraire est problématique car même cela restreint le champ d'application de la méthode à un type spécifique d'images où le classifieur apprend une caractéristique de la structure à segmenter et non pas sa représentation globale. Cela limite aussi l'application de ce type de méthodes dans le cas où les structures anatomiques sont déformées.

Le succès des méthodes d'apprentissage profond à réaliser la classification d'images a étendu leur utilisation afin de résoudre des tâches plus complexes dont la segmentation sémantique en l'interprétant comme un problème soit de régression ou de discrimination. Le premier cas de figure est très peu répandu mais tout à fait possible. D'ailleurs, Suzani *et al.* (2015) se sont tournés vers cette voie pour proposer une méthode de détection, localisation et segmentation des CVs. L'approche est divisée en deux phases. La localisation est traitée comme un problème de régression résolu par un réseau de neurones profond. Ce dernier est entraîné à calculer la distance entre un voxel et le centre de tous les CVs de la même façon que cela a été fait par Chen *et al.* (2015). La segmentation se fait via le recalage d'un modèle statistique. Contrairement à Chen *et al.* (2015), les résultats ont montré que l'utilisation des réseaux de neurones a permis une plus grande robustesse à détecter les CVs sans se restreindre à la tranche médiane du

volume. Toutefois, les résultats qualitatifs dévoilent une sur-segmentation où certaines parties des CVs sont considérées comme étant l'arrière-plan de l'image.

Résoudre un problème de segmentation par discrimination est grandement populaire. Les premières approches basées sur les CNNs pour réaliser une segmentation n'étaient en réalité qu'une classification pixel par pixel où un pixel était classifié en se basant sur la valeur des pixels de son voisinage. Lors de l'inférence, un filtre sous la forme d'une fenêtre glissante centrée en un pixel parcours toute l'image. Chaque pixel est classifié individuellement et l'opération est répétée autant de fois que le nombre de pixels présent dans l'image. Ciresan *et al.* (2012) ont utilisé ce mécanisme lors du (*2012 Electron microscopy segmentation challenge in the IEEE International Symposium on Biomedical Imaging (ISBI)*) où ils ont fini premiers. Même si les CNNs ont dépassé les autres approches pour les tâches de segmentation et que cette méthode de fenêtre glissante semble intuitive, elle n'est pas sans faille. Pour des raisons de complexité de calcul et de ressources matérielles, le plus souvent les fenêtres sont de taille très petite par rapport à celle des images en entrée et donc l'information sur laquelle se base le classifieur est limitée. Mais surtout, la méthode souffre de calculs redondants qui retardent considérablement l'inférence. Jianhua *et al.* (2016) ont utilisé cette même stratégie pour segmenter les DIVs à partir d'IRM. Les résultats obtenus sont satisfaisants mais dans cette étude encore, aucune validation n'est effectuée sur des IRMs de sujets dont la colonne vertébrale est déformée.

L'évolution des CNNs pour réaliser la segmentation est basée sur l'hypothèse que les opérations de convolutions et de produits scalaires sont linéaires et interchangeables et qu'il est possible de substituer les couches entièrement connectées par des couches de convolution. Long *et al.* (2015) furent les premiers à poser cette hypothèse en introduisant ce qui marquera la base des plus récentes architectures CNNs pour réaliser la segmentation sémantique : les réseaux complètement convolutifs (FCNs). Jusque là, les CNNs suivaient les grandes lignes posées par Lecun *et al.* (1998) i.e. une architecture pyramidale où la résolution spatiale de la représentation de l'image se réduit au fil des couches et la profondeur s'accroît jusqu'à ce qu'elle soit transmise à des couches complètement connectées qui seront connectées à la couche en

sortie. Les FCNs suppriment les couches complètement connectées et les remplacent par une structure pyramidale mais cette fois-ci en réduisant la profondeur et en augmentant la résolution spatiale par des opérations de sur-échantillonnage. La dimension de la couche en sortie s'accorde à celle de l'image en entrée. De même pour la profondeur et au nombre de classes où chaque canal correspond au masque de segmentation de la classe qui lui est associée. Les couches complètement connectées ne conviennent pas à la segmentation pour deux raisons : elles obligent les images en entrée à être de la même dimension et elles entraînent une perte en informations locales très importantes en segmentation. Korez *et al.* (2016) ont proposé une méthode de segmentation 3D de CVs à partir d'IRM en introduisant un FCN 3D entraîné à générer une carte de probabilité des CVs dans le volume. Un post traitement s'appuie sur la carte de probabilité produite pour guider un modèle déformable vers les contours des CVs. Chen *et al.* (2016) sont parmi les premiers à proposer une approche basée sur l'apprentissage profond pour la segmentation des DIVs en IRM via une architecture FCN 3D. Dans leur étude expérimentale, les auteurs ont répondu à la question : vaut-il mieux faire de la segmentation sur des images 2D ou 3D ? Il s'est avéré que les résultats penchent vers ceux obtenus par le FCN 3D où l'exploitation de l'information contextuelle volumétrique a donné une segmentation plus précise.

Le FCN original introduit par Long *et al.* (2015) a une représentation pyramidale qui se termine par une seule opération de sur-échantillonnage suivie par la couche de sortie. Ceci a apporté une grande précision de localisation et de classification et de réduction de la complexité. Toutefois un problème non négligeable est survenu : la faible résolution des images en sortie. Pour surmonter cela, plusieurs nouveaux types de FCN ont été proposés mais on se concentrera seulement sur les encodeurs-décodeurs en raison de notre choix méthodologique. Ronneberger *et al.* (2015) ont introduit une architecture basée sur les FCN spécialement développée pour la segmentation de données biomédicales qui est devenue un standard dans ce domaine. L'architecture est baptisée le *u-net* en référence à sa forme d'encodeur décodeur. Le u-net a surpassé toutes les autres méthodes au (*2015 Electron microscopy segmentation challenge in the IEEE International Symposium on Biomedical Imaging (ISBI)*). En général, les architectures FCN

de type encodeur-décodeur sont constituées de deux parties. L'encodeur est une représentation pyramidale classique semblable à celle des CNNs. La différence réside dans la deuxième partie où à défaut d'avoir une seule opération de sur-échantillonnage pour retrouver seulement la dimension spatiale comme dans le FCN classique, le décodeur est symétrique à l'encodeur (Figure 2.1) mais remplace le sous-échantillonnage par le sur-échantillonnage. Dans cette partie, le but avant d'arriver à retrouver la dimension spatiale est de récupérer les détails de l'image pour atteindre une résolution complète en associant caractéristiques globales et locales via des sauts de connexions *longues* en provenance de l'encodeur. Comme suite logique au succès que le u-net a suscité, diverses approches de segmentation d'images médicales y ont eu recours et certaines ont apporté des modifications à son architecture pour contourner ses limites.

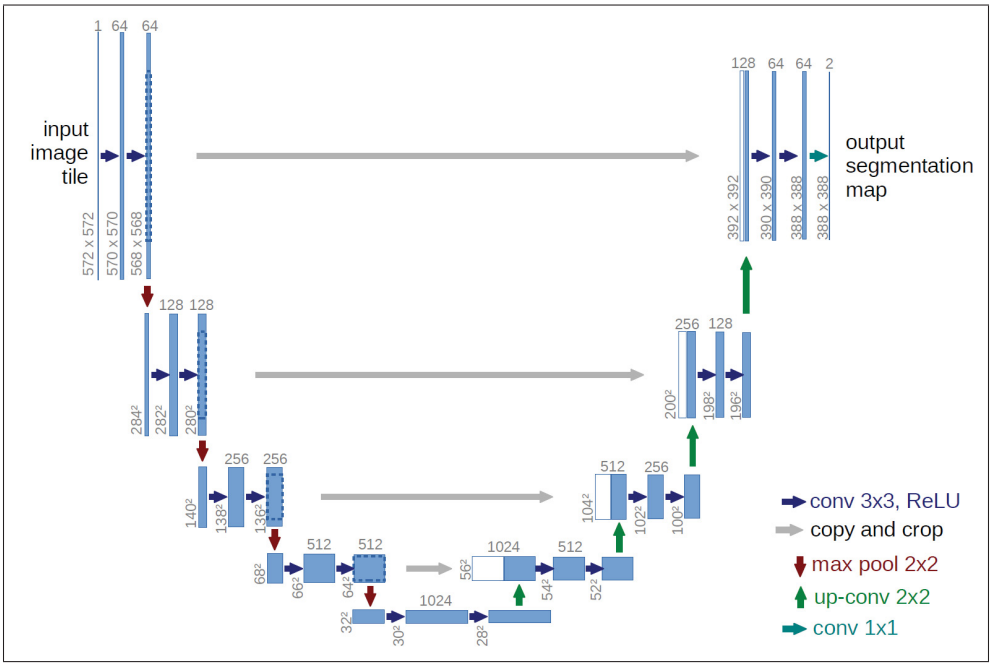


Figure 2.1 Architecture u-net. Tirée de Ronneberger *et al.* (2015)

Jen-Tang *et al.* (2018) ont utilisé un u-net en ajoutant seulement une opération de normalisation par batch après chaque convolution pour segmenter des CVs. La segmentation serait envisagée à être utilisée dans un outil d'aide au diagnostic pour évaluer la sévérité de dégénération de la sténose lombaire. Toutefois, dans leur étude expérimentale, l'approche a été testée

sur certains patients scoliotiques. Les résultats ont été présentés sur uniquement les cas où la scoliose est légère. Les auteurs ont indiqué que les cas de scolioses sévères seraient un défi pour leur approche. Kim *et al.* (2018) ont proposé une méthode de segmentation des DIVs en introduisant l'architecture BSu-net dotée d'une stratégie d'apprentissage en cascade. Celle-ci est constituée de deux réseaux. Le premier a une forme u-net où les opérations de convolution sont remplacées par des blocs résiduels et les opérations de sous-échantillonnage sont une concaténation de la sortie d'un maxpooling avec la sortie de deux convolutions avec un pas de 2. Cela est jugé moins agressif qu'utiliser simplement un maxpooling qui a tendance à garder un seul pixel dans un voisinage et peut supprimer d'importantes informations. Le deuxième réseau est une série de blocs résiduels où chacun est formé par la concaténation du dernier bloc résiduel du premier réseau avec la segmentation qu'il a prédit. L'apprentissage en cascade et le recours à une meilleure stratégie de sous échantillonnage ont dévoilé une segmentation plus exacte que celle produite par un simple u-net, ce qui souligne tout l'intérêt de la propagation des caractéristiques à travers le réseau. Dolz *et al.* (2018) ont entraîné un u-net qui prend en compte l'information provenant de différentes modalités dans le but de segmenter des DIVs. L'architecture comprend une entrée pour chaque modalité. L'information des différentes modalités est transmise depuis les différentes couches de la partie encodeur. Pour cela, les auteurs se sont inspiré du DenseNet pour connecter les couches entre elles. Dans la partie compressée, les sorties des différents chemins sont concaténées avant de passer par la partie décodeur. Des opérations de convolutions dilatées de différentes échelles ont été utilisées pour inclure de l'information multi échelle. L'étude a démontré que l'appui sur une simple stratégie de fusion des caractéristiques est insuffisant pour complètement exploiter l'information qui provient des différentes modalités et qu'une fusion de multiples sources permettrait de capturer des caractéristiques plus discriminantes. Plus encore, la propagation des caractéristiques inter chemins dans la partie encodeur via les connexions du DenseNet ont permis d'obtenir de meilleurs résultats que si la fusion se faisait uniquement dans la partie compressée.

Toutefois, il n'est pas juste de croire que le u-net obtient systématiquement les meilleures performances. Li *et al.* (2018) ont fini premiers au (*2018 Intervertebral disk segmentation*

challenge in the IEEE International Symposium on Biomedical Imaging (ISBI)) en proposant une approche basée sur un 3D FCN multi échelle et multi modalité. L'architecture proposée est de 3 entrées respectivement pour 3 fenêtres de taille différente. Pour pallier le problème de déséquilibre de classe, une fonction d'entropie croisée pondérée est utilisée pour donner plus d'importance aux voxels des DIVs. Un dropout multi modalité employé pour réduire la co-adaptation des caractéristiques lors de l'entraînement. Le principe est qu'à chaque itération une des modalités est aléatoirement choisie à laquelle un dropout est appliqué. L'intérêt est de ne pas générer des caractéristiques redondantes ce qui cause le sur-apprentissage, mais plutôt viser des caractéristiques plus discriminantes. Une étude comparative avec un u-net a dévoilé de meilleurs résultats pour le FCN proposé. Cela reviendrait à une meilleure stratégie d'extraction des caractéristiques entre les différentes modalités ainsi que l'incorporation de l'information multi-échelle.

Récemment, Roy *et al.* (2018) se sont inspirés du *squeeze and excitation* pour proposer un bloc qui conviendrait plus à une segmentation. En effet, le bloc proposé par Hu *et al.* (2018) ne tient pas en compte l'information locale qui est crutiale en segmentation parce qu'elle informe sur le voisinage d'un pixel et pondère selon l'importance du canal uniquement. Pour remédier à cela, les auteurs proposent d'ajouter un bloc nommé *sSE* parallèlement à celui qui existe qui a été renommé *cSE*. Le bloc *sSE* compresse la carte de caractéristiques le long des canaux avec une convolution ($1 \times 1 \times C$) et excite sa sortie avec une fonction *Sigmoïde*. Les deux blocs sont combinés dans le bloc *scSE* (Fiure 3.5). Le bloc est introduit dans un DenseNet pour réaliser la segmentation de multiples structures dans le cerveau à partir d'IRM et la segmentation de multiples organes à partir d'image CT. Dans leur étude expérimentale, les auteurs ont mis l'accent sur l'importance de l'information locale en segmentation. Quatre réseaux ont été entraînés dont le premier contenait le bloc *scSE* et le dernier en était dépourvu. La partie intéressante de cette étude est de voir que le bloc *sSE* a dépassé le bloc *cSE* prouvant que leur impact est significatif en segmentation mais aussi d'un point de vue général, le recours à une recalibration des canaux est une stratégie payante pour augmenter les performances d'un modèle à réaliser la segmentation d'images médicales.

2.2 Conclusion

À travers la revue de littérature, nous constatons que la segmentation des structures anatomiques de la colonne vertébrale suscite beaucoup d'intérêt au sein la communauté scientifique. Les approches classiques ont apporté des premières solutions de segmentation des DIVs et CVs mais en entraînant les lacunes liées à chacune des méthodes employées. Les approches orientées vers l'apprentissage profond ont éliminé la plupart de ces limites mais il existe toujours très peu de méthodes permettant de segmenter simultanément les DIVs et les CVs, et très peu d'études ont été réalisées dans le contexte particulier de la scoliose. De plus, les architectures proposées jusqu'à présent n'ont pas eu recours à un mécanisme de recalibration des canaux de caractéristiques qui vise à sélectionner des caractéristiques discriminantes qui apporteraient une meilleure compréhension des structures DIVs et CVs et ainsi l'adaptation aux déformations occasionnées par des pathologies comme la scoliose. Dans le chapitre suivant, nous détaillerons notre méthodologie de segmentation simultanée des DIVs et des CVs dont les hypothèses ont été posées pour apporter des solutions aux limites observées dans la revue de littérature.

CHAPITRE 3

MÉTHODOLOGIE

Dans ce travail, nous proposons une méthode basée sur l'apprentissage profond pour segmenter des DIVs et des CVs sur le plan sagittal d'IRM de sujets atteints ou non de scoliose idiopathique (Figure 3.1). La segmentation 2D est préférée à la segmentation 3D à cause du manque de volumes de patients. De plus, l'étude de la scoliose est réalisée sur le plan sagittal en raison de la visibilité des deux structures simultanément sur la même tranche ainsi que leur forme qui se distingue plus aisément des structures anatomiques voisines en comparaison aux deux plans orthogonaux. De ce fait, nous avons préféré nous tourner vers un apprentissage avec un nombre conséquent de données au détriment de l'information volumique que d'obtenir un sur-apprentissage pour cause de manque de données. D'ailleurs, notre base de données d'entraînement est constituée d'images de sujets non scoliotiques et d'images de patients scoliotiques (Section 3.1.3). Un pré-traitement (Section 3.1) est appliqué sur l'ensemble de ces images pour trois raisons : réduire le bruit généré dans les IRMs, apparier les images des deux bases de données et augmenter le nombre d'images via des stratégies d'augmentation géométrique. Pour ce qui est du modèle, nous nous sommes orientés vers une architecture FCN de type encodeur-décodeur inspirée du u-net et traitée comme un problème de multi-classification à 3 classes (DIV, CV et arrière-plan de l'image) (Section 3.2). Un mécanisme de recalibration des cartes de caractéristiques est introduit dans l'architecture via un bloc nommé *scSE* qui pondère les canaux des cartes de caractéristiques selon leur importance. Le réseau prend en entrée une tranche 2D sagittale dans sa taille originale et calcule une carte de probabilité pour chacune des classes avant de produire le masque de segmentation. Dans un premier temps, le réseau est entraîné avec l'ensemble de la base de données de sujets non scoliotiques. Cet entraînement servira à paramétrer le modèle et à se familiariser avec les IRM du tronc afin de reconnaître les caractéristiques globales des structures de la colonne appartenant ou non aux régions de DIVs et CVs. Ensuite, nous ré-entraînons le réseau avec un petit nombre d'images de la base de données de patients scoliotiques pour ainsi l'amener à tenir compte des déformations dues

à la scoliose (Section 3.4.6). Notre modèle est évalué par la suite selon des mesures basées sur la précision et le rappel (Section 3.5).

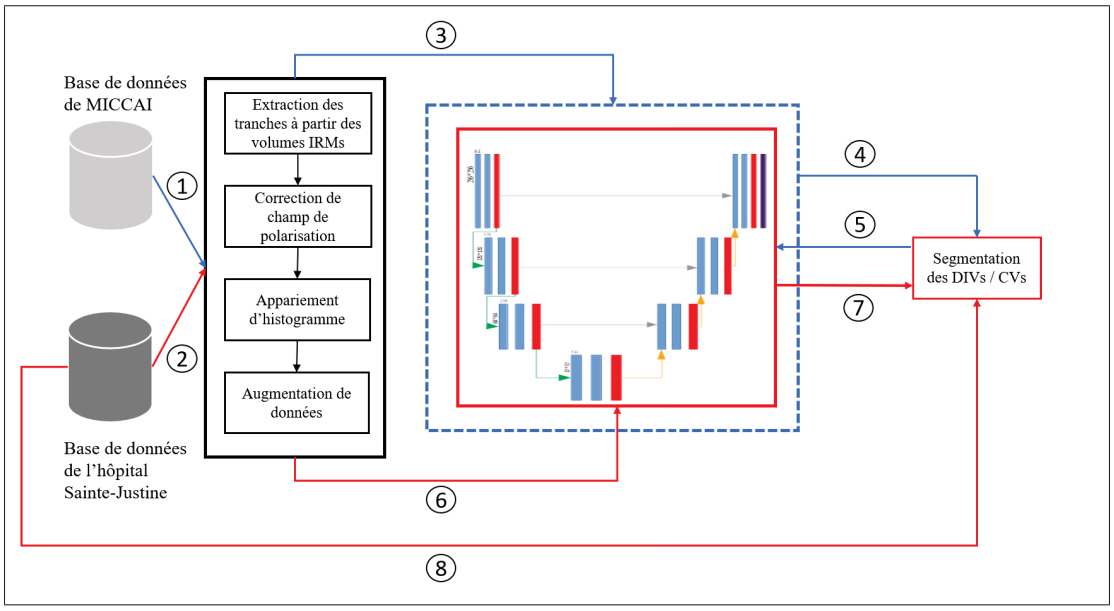


Figure 3.1 Vue d'ensemble de la méthode de segmentation des DIVs et CVs proposée. (1) et (2) pré-traitements appliqués sur les IRMs. (3) et (4) entraînement du modèle à segmenter les DIVs / CVs avec les images de sujets non scoliotiques. (5) transfert d'apprentissage. (6) et (7) ré-entraînement du modèle avec les images de patients scoliotiques. (8) segmentation des DIVs / CVs de patients scoliotiques à partir d'images non observées.

3.1 Pré-traitement

Le pré-traitement est une étape nécessaire dans la majorité des travaux réalisés en ayant recours à l'apprentissage automatique. Cela consiste à traiter les données dans leur état brut pour en faire un ensemble de données nettoyées et normalisées pour favoriser leur exploitation et ainsi maximiser les performances obtenues par le modèle. Dans le cadre de notre projet, ayant uniquement recours aux tranches sur la coupe sagittale extraites des volumes IRMs pour entraîner notre modèle, nous présentons dans cette section les problèmes liés à ces images et les méthodes utilisées pour les préparer à la phase d'apprentissage.

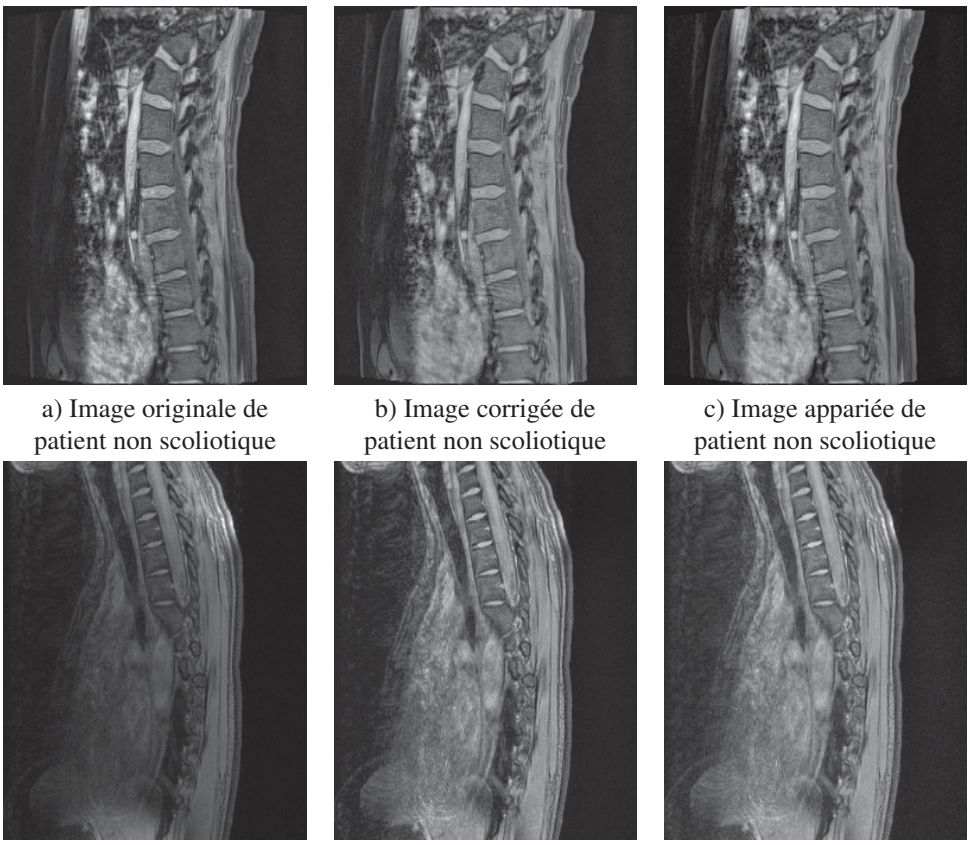


Figure 3.2 Résultat du pré-traitement sur des images issues des deux bases de données. Les images du milieu (b-e) sont le résultat de l'algorithme N4ITK appliqué sur les images originales de gauche (a-d). Les images de droite (c-f) sont le résultat de l'appariement d'histogramme appliqué sur les images du milieu (b-e).

3.1.1 Correction de champ de polarisation

En raison des signaux de basses fréquences indésirables causés par la non homogénéité du champ magnétique, les données d'IRM sont bruitées et cela engendre une perte d'information et un contraste variable. Les algorithmes de traitement d'images se basent sur l'hypothèse que l'information spatiale est invariante. Si par exemple on considère une vertèbre C1 et une autre L5 qui se trouvent dans deux régions éloignées de l'image et que l'intensité n'est pas homogène, un tel phénomène peut affecter la performance du modèle en réduisant sa capacité à se généraliser et par la même occasion accroître le risque de sur-apprentissage. Donc un pré-

traitement est nécessaire pour corriger les IRM. Nous avons décidé d'utiliser la méthode de correction de champ de polarisation N4ITK proposée par Tustison *et al.* (2010) dont l'implémentation est disponible dans la librairie ITK. C'est une version améliorée de la correction de champ de polarisation N3 proposée par Sled *et al.* (1998) devenue une méthode standard de pré-traitement d'IRM. Cette dernière corrige de façon itérative le contraste de l'image et ne nécessite aucune information a priori.

Le principe est d'estimer à chaque itération le champ de polarisation représenté par une combinaison de B-Splines, ainsi que la distribution de l'intensité des pixels de l'image. En pratique, l'algorithme pose l'hypothèse que l'image I est corrompue par un signal bruité n et que son rôle est de débruiter l'image itérativement via des opérations de déconvolution. Nous illustrons cela par la notation suivante :

$$v(x) = u(x)f(x) + n(x) \tag{3.1}$$

où v est l'image, u est l'image non corrompue, f est le champ de polarisation et n est un bruit gaussien indépendant. En supposant un scénario dépourvu de bruit et en définissant la notation $\hat{v} = log\ v$.

$$\hat{v}(x) = \hat{u}(x) + \hat{f}(x) \tag{3.2}$$

À partir de cela, l'image non corrompue est calculée itérativement. L'image à la n ème itération est :

$$\hat{u}^n = \hat{v} - \hat{f}^n \tag{3.3}$$

$$\hat{u}^n = \hat{v} - S\left\{\hat{v} - E[\hat{u}|\hat{u}^{n-1}]\right\} \tag{3.4}$$

Où $\hat{u}^0 = v$ et $S\{.\}$ représente le processus de lissage qui vise à approximer la valeur de la B-spline. $E[\hat{u}|\hat{u}^{n-1}]$ est l'estimé courant de l'image corrigée.
L'amélioration apportée par la N4 est la mise à jour continue durant la correction de l'image.

L'équations 3.3 et 3.4 ont été remplacées par :

$$\hat{u}^n = \hat{u}^{n-1} - \hat{f}_r^n \tag{3.5}$$

$$\hat{u}^n = \hat{u}^{n-1} - S^* \left\{ \hat{u}^{n-1} - E[\hat{u}|\hat{u}^{n-1}] \right\}_f \tag{3.6}$$

Le fait de remplacer $S\{.\}$ par $S^*\{.\}$ permet d'obtenir un espacement entre les points de contrôle plus petit ce qui alloue l'accommodation d'un champ plus intense sans accroître le risque que l'algorithme échoue, ce qui a aussi pour avantage d'accélérer le processus.

Les résultats obtenus en appliquant la N4ITK sur une image de la base de données de MICCAI (Figure 3.2a) et une image de la base de données du CHU Sainte-Justine (Figure 3.2d) sont représentés dans les figures 3.2b et 3.2e. À partir des images d'origine, nous pouvons apercevoir le problème d'hétérogénéité de l'intensité des IRMs. Celui-ci est plus accentué sur l'image 3.2d en raison de l'ancienneté du modèle d'acquisition où l'image obtenue est plus bruitée que les modèles IRMs récents. Toutefois, nous remarquons que le ton de gris des images corrigées 3.2b et 3.2e est nettement mieux réparti et les structures anatomiques sont homogénéisées en termes d'intensité.

3.1.2 Appariement d'histogramme

La correction de champ de polarisation permet de réduire le bruit dans l'IRM et homogénéise l'intensité pour l'ensemble des images appartenant à la même base de données. Cependant, dans notre cas, nous disposons de deux bases de données de sources différentes. Pour réduire la variation entre les bases de données, nous proposons d'utiliser un appariement d'histogramme en utilisant une image de la base de données des patients atteints de scoliose comme image référence et d'y apparier les histogrammes de toutes les images dont nous disposons. Ceci a pour but d'ajuster l'intensité et rehausser le contraste d'une image en appliquant une transformation sur l'ensemble de ses pixels. Concrètement le processus calcule une bijection M d'intensité à intensité. Le processus est défini comme suit :

- Soient une image référence S et une image cible C. Nous calculons leurs histogrammes $h(S)$ et $h(C)$.
- Nous calculons la fonction de répartition à partir de l'histogramme des deux images $Fr(S)$ et $Fr(C)$.
- Pour chaque ton de gris c : Chercher un s tel que : $Fr(c) = Fr(s)$.
- Appliquer à chaque pixel d'intensité c dans l'image cible, la transformation $M(c) = s$.

Un résultat de l'appariement d'histogramme sur les images 3.2a et 3.2d est présenté à partir des images 3.2c et 3.2f.

3.1.3 Augmentation de données

Pour combler en partie la nécessité d'alimenter le réseau avec un grand nombre de données, nous avons eu recours à une pratique courante en apprentissage profond : l'augmentation de données. Cette pratique a pour effet de présenter une information sous différents aspects dans l'image ce qui est favorable à l'entraînement des paramètres et d'éviter le sur-apprentissage. Notre stratégie d'augmentation est purement géométrique et est appliquée *en ligne* (i.e durant l'entraînement). Nous avons donc appliqué pour chaque image en entrée une combinaison aléatoire d'opérations afin de fournir au réseau des variations des informations présentes dans les images d'origine à chaque itération. Les opérations employées (Figure 3.3) sont : rotation, translation, transvection (*shear mapping*), retournement horizontal et déformation élastique dont les paramètres sont aléatoires et choisis selon un intervalle fixe. Les valeurs ont été expérimentalement choisies (Tableau 3.1) de telle sorte à obtenir des images transformées sans occasionner d'importantes déformations à la colonne vertébrale. Le recours à la déformation élastique et la transvection permettent d'obtenir des variations de CVs et des DIVs étirés ou rétractés qui ressemblent localement à celles qui pourraient se produire si le sujet est atteint de scoliose. La rotation et la translation permettent au modèle de récupérer l'information dans différentes régions de l'image.

Opération	Intervalle / valeur	Unité
Rotation	[-25 , 25]	Degré
Translation	[-50 , 50]	Pixel
Transvection	[-30 , 30]	Radian
Déformation élastique	[0 , 50]	Pixel (Noyau du filtre gaussien)
	[0 , 300]	Pixel (Facteur de multiplication du filtre gaussien)

Tableau 3.1 Intervalle des valeurs des opérations géométriques d'augmentation de données

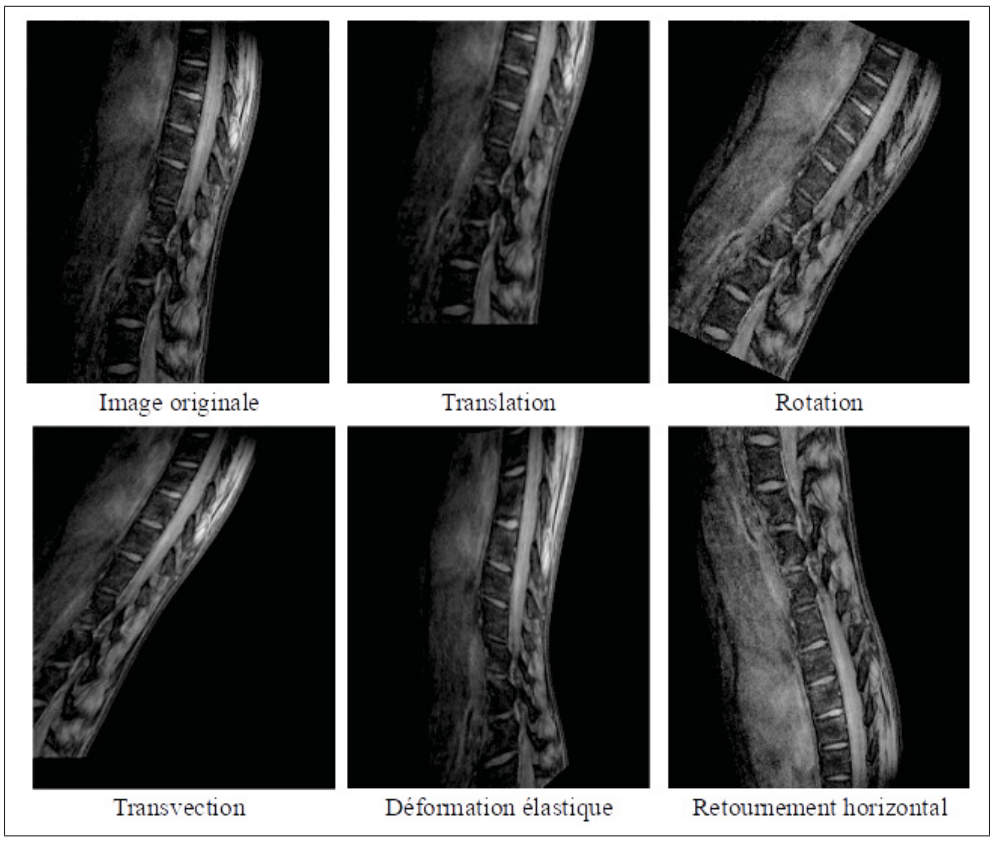


Figure 3.3 Exemples d'images transformées via les stratégies d'augmentation de données employées

3.2 Architecture du modèle proposé

Notre méthode se base sur l'architecture FCN de type encodeur-décodeur du u-net (Figure 3.4). L'encodeur consiste en une structure pyramidale d'un CNN tout à fait classique. Au fil

de la profondeur, la dimension de l'image réduit et sa profondeur s'accroît. Le décodeur est complètement symétrique à l'encodeur. En revanche, la structure pyramidale est inversée. Le but est de retrouver la dimension et la résolution de départ. Chacune des parties du réseau est constituée de 3 niveaux de même dimension. Chaque niveau est doté de deux couches de convolution avec *ElU* comme fonction d'activation et des filtres de taille 3×3 avec un pas de 1 dont le nombre augmente dans la partie encodeur et décroît dans la partie décodeur. Elles sont suivies par un bloc scSE (Figure 3.5) et d'une couche d'échantillonnage : *maxpooling* dans la partie encodeur et *déconvolution* dans la partie du décodeur. La déconvolution a été préférée à un simple redimensionnement comme l'interpolation pour continuer de profiter de l'extraction de caractéristiques. Les deux opérations d'échantillonnage utilisent un filtre de taille 2×2 avec un pas de 1. Celles-ci disposent du même paramétrage pour obtenir systématiquement les mêmes dimensions dans les deux parties de chaque niveau. La couche en sortie utilise une fonction *softmax* comme fonction d'activation étant donné que nous entraînons le modèle à effectuer une multi-classification.

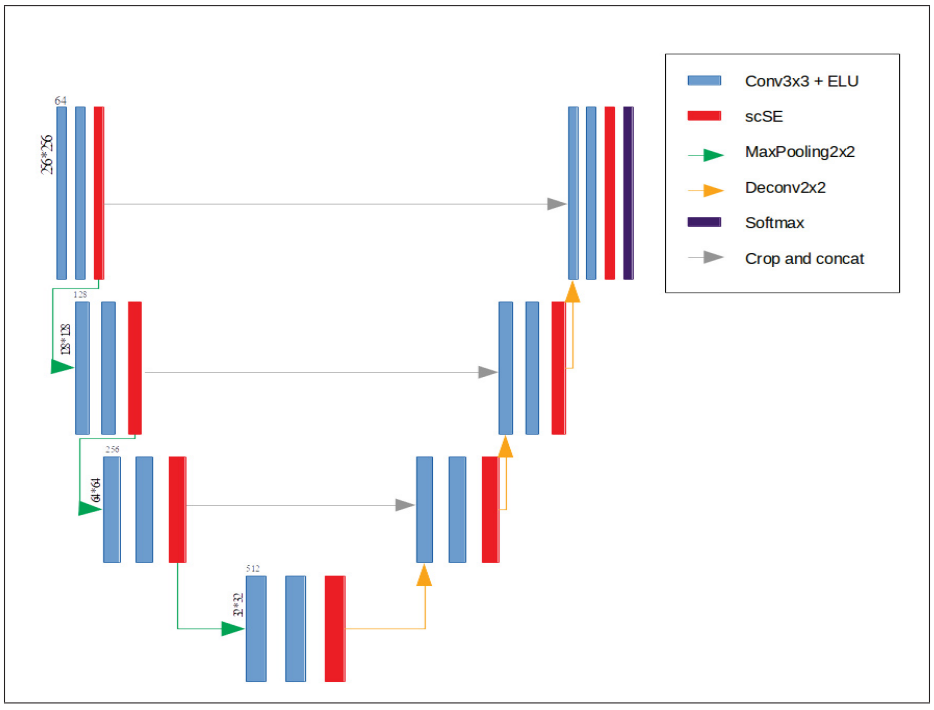


Figure 3.4 Vue d'ensemble de l'architecture encodeur-décodeur utilisée

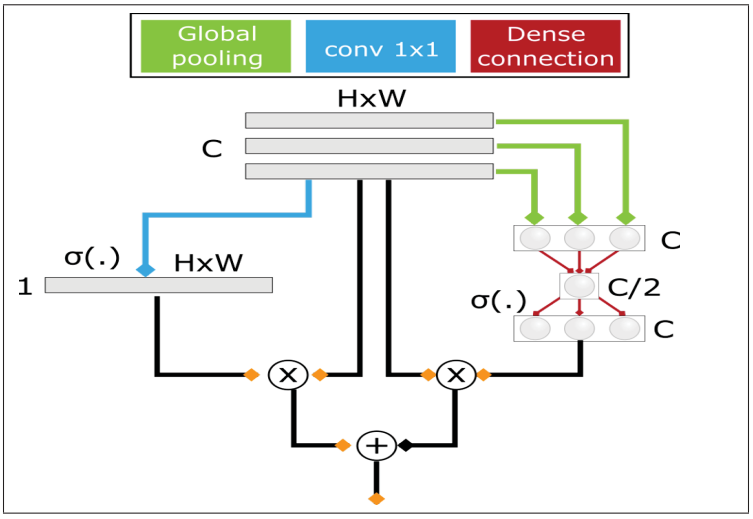


Figure 3.5 Vue d'ensemble du bloc scSE

3.3 Fonction objectif

Dans notre réseau de neurones, nous nous sommes tournés vers la combinaison du coefficient de kappa de Cohen (κ) (Cohen, 1960) et de l'entropie croisée (Rubinstein & Kroese, 2004) qui est basée sur la divergence de Kullback-Leibler (Kullback & Leibler, 1951) pour représenter notre fonction objectif.

Le coefficient Kappa Cohen compare le niveau d'accord entre deux observations (vérité terrain et prédiction du modèle) en considérant que l'accord (vérité terrain = prédiction modèle) peut se faire fortuitement. Le kappa est défini comme suit :

$$\kappa = \frac{P_{precision} - P_{chance}}{1 - p_{chance}} \tag{3.7}$$

On note comme propriété :

- $\kappa = 1$: indique un parfait accord.
- $\kappa = 0$: indique un accord fortuit.
- $\kappa = -1$: indique qu'il n'y a pas d'accord.

$P_{precision}$ est l'accord observé (ou simplement la précision). Quant à P_{chance} est la probabilité d'un accord fortuit. Tout deux sont définis comme suit :

$$P_{precision} = \frac{|VP| + |VN|}{|VP| + |VN| + |FP| + |FN|} = exactitude \tag{3.8}$$

$$P_{chance} = \frac{|VN| + |FP|}{|VP| + |VN| + |FP| + |FN|} . \frac{|VN| + |FN|}{|VP| + |VN| + |FP| + |FN|} \tag{3.9}$$

où :

- Vrai positif (VP) = Nombre de cas où prédiction et vérité terrain sont positives.
- Vrai négatif (VN) = Nombre de cas où prédiction et vérité terrain sont négatives.
- Faux positif (FP) = Nombre de cas où prédiction positive et vérité terrain négative.
- Faux négatif (FN) = Nombre de cas où prédiction négative et vérité terrain positive.

Ces 4 mesures forment la matrice de confusion qui a pour but de mesurer la qualité d'un système de classification. Dans notre problème de multi-classification, la matrice de confusion est définie selon le tableau 3.2.

	DIV_Prédit	**CV_Prédit**	**AP_Prédit**
Réel_DIV	VP	FN	FN
Réel_CV	FN	VP	FN
Réel_AP	FP	FP	VN

Tableau 3.2 Représentation de la matrice de confusion de la classification DIV / CV / AP

En apprentissage automatique, κ est avantageusement employé pour assister la convergence en présence du problème de déséquilibre des classes. Playout (2018) a démontré cela en comparant le meilleur et pire cas à savoir : un équilibre et déséquilibre de classes. Il en a conclu que dans un cas équilibré, κ se ramène à l'exactitude car la performance se résume à l'accord relatif entre les observateurs. Dans un cas de déséquilibre, il devient équivalent à la précision ce qui est intéressant car la précision ne prend pas en compte le taux de prédictions négatives correctes qui pourrait biaiser la mesure.

L'entropie croisée est utilisée pour résoudre des problèmes de classification parce qu'elle permet de réduire la distance entre deux distributions $P_{reelles}$ et $P_{predits}$ qui représentent respectivement la distribution des données réelles et des données prédites par le modèle. En réalité, la fonction de l'entropie croisée utilise la divergence de Kullback-Leibler pour mesurer la similitude entre les deux distributions. Shlens (2014) a détaillé cette mesure ainsi que son lien avec la théorie de vraisemblance. Il explique qu'en réalité, elle mesure la vraisemblance qu'une donnée de la distribution $P_{reelles}$ soit générée par une distribution $P_{predits}$. Elle est définie comme suit :

$$D_{\mathrm{KL}}(P_{reelles}|P_{predits}) = \sum_{i}^{N} P_{reelles}(x_i) \log \frac{P_{reelles}(x_i)}{P_{predits}(x_i)}. \tag{3.10}$$

On dit que $P_{reelles}$ est générée par $P_{predits}$ si et seulement si : $D_{KL}(P_{reelles}|P_{predits}) = 0$. Et que plus $D_{KL}(P_{reelles}|P_{predits})$ est élevé moins il est probable que $P_{reelles}$ soit générée par $P_{predits}$.
Le fait qu'on s'intéresse à $D_{KL}(P_{reelles}|P_{predits})$ n'est pas anodin. Cette mesure débouche sur l'entropie croisée (Voir Annexe I pour démonstration). D'ailleurs cette dernière est définie sur l'ensemble des données N et l'ensemble des classes K comme suit :

$$L_{EC}(P_{reelles}, P_{predits}) = \sum_{i}^{N} \sum_{j}^{K} -P^{(j)}_{reelles}(x_i) log(P^{(j)}_{predits}(x_i)). \tag{3.11}$$

Comme nous l'avons vu précédemment, l'entropie croisée est efficace dans un cas de classification si la fonction d'activation de la couche en sortie est une fonction *sigmoïde* ou une fonction $softmax$ pour calculer le gradient d'erreur. Aussi, étant confrontés à un problème de déséquilibre de classes où l'arrière-plan est nettement plus dominant que les CVs et encore plus les DIVs, nous choisissons comme fonction objectif F à minimiser :

$$F = 1 - \kappa + L_{EC} \tag{3.12}$$

cela permet de prendre en considération le déséquilibre des classes tout en favorisant la minimisation de la distance entre distribution des données prédites par notre modèle et la distribution des données réelles.

3.4 Stratégie d'entraînement

Le choix de recourir à l'apprentissage profond pour réaliser la segmentation des DIVs et CVs n'est pas une évidence. En effet, bien que les approches basées sur l'apprentissage profond aient démontré de meilleures performances dans de nombreux problèmes complexes de reconnaissance de formes. Ces performances sont étroitement liées au nombre de données d'entraînement ainsi qu'à leur diversité. Les réseaux de neurones sont dotés d'un très grand nombre de paramètres qui nécessitent un flux constant de nouvelles informations pour les optimiser. Les performances seraient limitées autrement pour cause de sur-apprentissage. Dans notre projet, la collecte d'une telle base de données représente un défi majeur. Pour des raisons éthiques, les données biomédicales sont souvent privées et dans le cas où celles-ci sont publiques, l'étiquetage n'est que rarement disponible. De ce fait, nous avons été contraints à réaliser la vérité terrain nous mêmes après avoir étudié l'anatomie de la colonne vertébrale à partir d'IRM afin de valider la faisabilité de notre approche.

En apprentissage profond, les techniques auxquelles nous avons eu recours comme le transfert d'apprentissage ou l'augmentation de données permettent de réduire cette nécessité d'avoir un grand nombre de données. Aussi, il est utile de recourir à des mécanismes comme le bloc *scSE* pour définir la pertinence des caractéristiques extraites et favoriser les plus discriminantes.

3.4.1 Base de données

Lors de la création de la base de données d'apprentissage, l'enjeu était de produire une segmentation des CVs et des DIVs en IRM chez des patients scoliotiques en ayant recours à l'apprentissage profond sans pour autant avoir beaucoup de données provenant de patients scoliotiques pour y parvenir. De ce fait, nous avons décidé de contourner le problème en composant notre base de données d'entraînement essentiellement d'images acquises sur des sujets non scoliotiques auxquelles s'ajoutent quelques images de patients scoliotiques. Pour les images non scoliotiques, nous avons pu compter sur un jeu de données rendu publiquement disponible [1]

[1] https ://ivdm3seg.weebly.com/

pour un défi de segmentation des DIVs organisé pour la conférence MICCAI en 2018. Quant aux images scoliotiques, nous avons utilisé des images acquises au CHU Sainte-Justine de Montréal que Chevrefils *et al.* (2007) ont utilisé dans leur étude de segmentation des DIVs.
Les données de MICCAI sont fournies avec un étiquetage complet des DIVs. Celles du CHU Sainte-Justine ont un étiquetage partiel (i.e un seul disque par volume de patient est segmenté). De ce fait, nous avons procédé à un étiquetage manuel des deux bases de données avec le logiciel 3D Slicer [2].

3.4.2 Base de données de MICCAI

Quatre volumes IRM pour 16 sujets adultes non scoliotiques ont été acquis avec un système à 1.5 Tesla magnetom Avanto (Siemens, Erlangen, Allemagne). Pour chaque sujet, quatre modalités différentes sont acquises : phase opposée, eau, phase interne et gras dans le même espace et sont donc parfaitement alignées. Ces images ont été acquises dans l'étude menée par Belavy *et al.* (2010) dont le but est d'étudier l'effet de l'inactivité du corps humain et simuler les effets de la micro gravité sur le corps humain. Le protocole d'acquisition est une séquence Dixon avec un contraste T1 sur le plan sagittal avec les paramètres suivants : Temps de répétition TR de 10,6 ms et temps d'echo TE de 4,76 ms. L'épaisseur des tranches est de 2 mm et celles-ci comptent chacune de 256×256 pixels avec un espace entre les pixels de 1,25 mm. L'espace entre les voxels est donc de 2*1,25*1,25 mm^3. Chaque volume contient 36 tranches. Le tableau 3.3 résume les statistiques démographiques liées aux patients.

Dans notre projet, nous avons décidé de nous restreindre aux protocoles phase opposée et eau pour leur similitude en termes d'intensité avec les images de test et ainsi de ne pas considérer plusieurs représentations d'une même caractéristique comme des informations différentes.

[2] https ://www.slicer.org

	Min	Max
Age (années)	21	45
Poids (Kg)	59	81,8
Taille (cm)	169	190

Tableau 3.3 Statistiques démographiques des sujets de la base de données MICCAI

3.4.3 Base de données du CHU Sainte-Justine

Les informations concernant la base de données du CHU Sainte-Justine sont tirées de la thèse doctorale de Chevrefils (2010) qui les a acquises pour mener son étude avec l'aide de deux experts.

La base de données comprend 1 volume IRM pour 11 adolescents scoliotiques dont 3 sont considérés comme étant de faible sévérité (angle de Cobb allant de 12° à 24°), 4 sont considérés de sévérité moyenne (angle de Cobb allant de 28° à 35°) et 4 sont considérés de haute sévérité (angle de Cobb allant de 43° à 60°). Le découpage des données d'entraînement et de tests est résumé dans le tableau 3.4. Pour les volumes utilisés dans la phase de test, nous avons gardé uniquement les tranches sagittales où nous apercevons au moins un DIV ou un CV.

Les images ont été acquises avec un système à 1.5 Tesla magnetom Avanto (Siemens, Erlangen, Allemagne). Les unités émettrices et réceptrices de radiofréquences (RF) sont constituées d'une bobine. Le protocole d'acquisition est une séquence 3D MEDIC (*Multi Echo Data Image Combination*) avec un contraste T1 sur le plan sagittal avec des paramètres : TR= 23 ms et TE= 12 ms. L'épaisseur des tranches est de 1 mm et chacune compte 256×256 pixels menant à une taille de voxel de 1 mm^3.

Patient ID	Nombre de tranches	Sévérité de la scoliose	Phase
Patient 1	89	Élevée	Entraînement
Patient 2	67	Modérée	Entraînement
Patient 3	40	modérée	Test
Patient 4	55	Élevée	Test
Patient 5	47	Légère	Test
Patient 6	40	Légère	Test
Patient 8	68	Élevée	Test
Patient 9	39	Légère	Test
Patient 10	42	Modérée	Test
Patient 11	60	Modérée	Entraînement
Patient 12	53	Élevée	Test

Tableau 3.4 Découpage de la base de données du CHU Sainte-Justine en données d'entraînement et de test avec la classification du degré de sévérité et le nombre de tranches pour chaque volume

3.4.4 Étiquetage manuel

Pour étiqueter les CVs et les DIVs, nous avons utilisé le logiciel 3D Slicer qui est un logiciel libre d'accès de visualisation et de traitement d'images médicales. Nous avons choisi de l'utiliser pour sa simplicité d'utilisation, sa grande palette d'outils de segmentation et son support des extensions DICOM, NIFTI et NPY. L'étiquetage s'est fait semi-manuellement pour la base de données MICCAI et manuellement pour la base de données du CHU Sainte-Justine, sur le plan sagittal dans les deux cas.

3.4.4.1 Étiquetage de la base de données de MICCAI :

Le processus d'étiquetage (Figure 3.6) est simple à réaliser mais nécessite une grande précision pour minimiser l'erreur humaine. D'ailleurs, c'est un procédé que nous avons ré-itéré pour la plupart des volumes. Voici les étapes que nous avons suivies pour réaliser l'étiquetage :

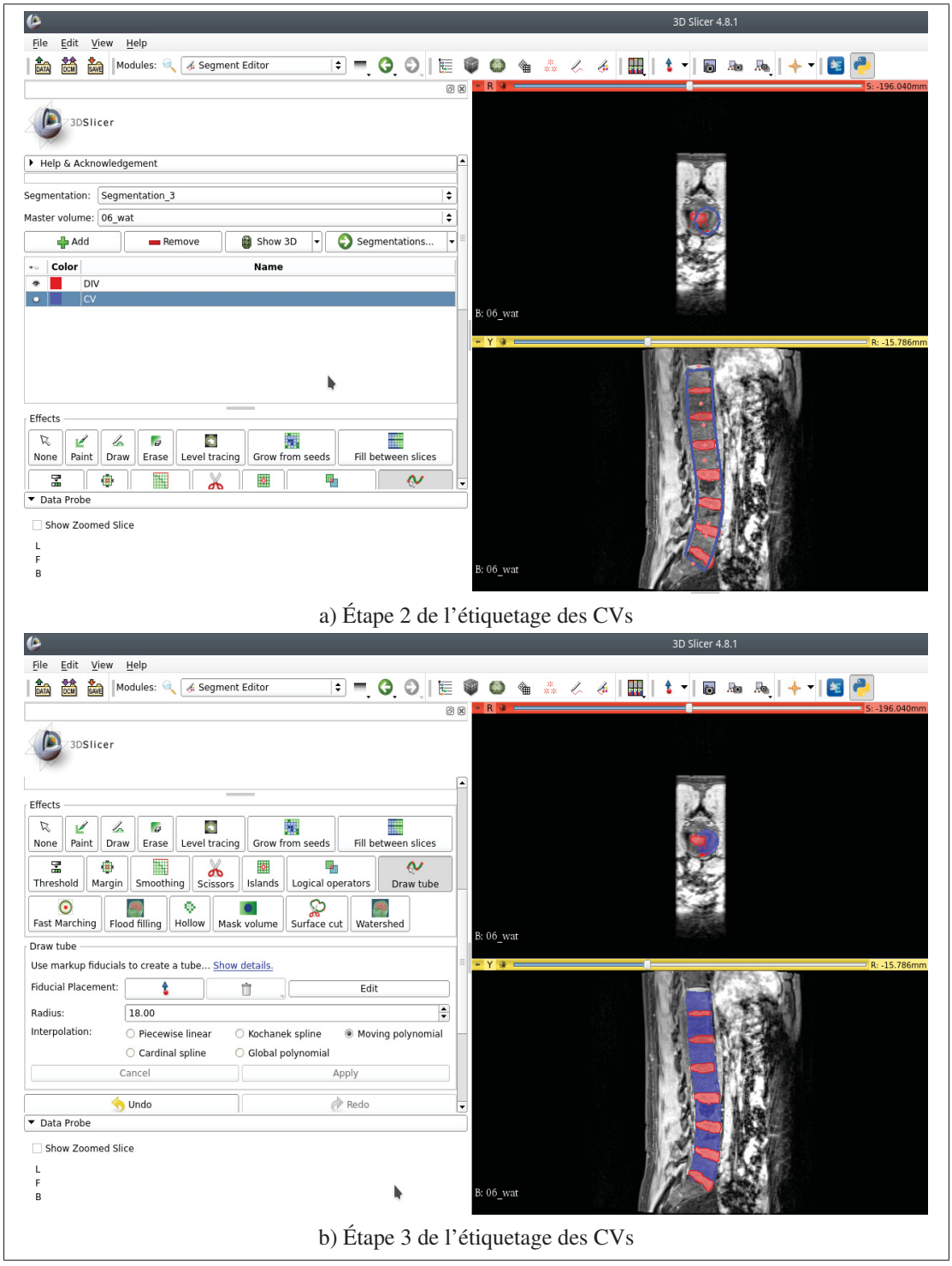
- **Étape 1 :** Importer l'étiquetage des DIVs.
- **Étape 2 :** Utiliser *drawTube*, un outil de segmentation qui relie des points manuellement posés sur l'image en produisant un tube pour segmenter l'ensemble de la colonne. Son utilisation consiste à placer sur la tranche médiane des points approximativement au centre

des DIVs ou des CVs, ensuite régler le rayon du tube pour qu'il soit ajusté aux frontières latérales des CVs et des DIVs ce qui va suivre la courbure naturelle de la colonne vertébrale. L'avantage de *drawTube* est que le marquage se propage sur plusieurs tranches du volume en avant et en arrière d'où l'intérêt de se placer sur la tranche médiane.

- **Étape 3 :** Soustraire du marquage produit par le tube la segmentation des DIVs fournie avec la base de données publique pour obtenir une segmentation grossière des CVs.
- **Étape 4 :** Affiner manuellement la segmentation des CVs.

3.4.4.2 Étiquetage de la base de données du CHU Sainte-Justine :

En raison de la déformation liée à la scoliose, nous ne pouvons appliquer exactement le même processus que celui réalisé sur les images de MICCAI. La variation d'une tranche à une autre est plus importante dans cette base de données et nous ne pouvons donc pas propager l'étiquetage en profondeur. De ce fait, nous segmentons manuellement les DIVs et CVs de chacune des tranches individuellement en commençant par tracer le contour des deux structures pour éviter tout chevauchement entre les deux.



a) Étape 2 de l'étiquetage des CVs

b) Étape 3 de l'étiquetage des CVs

Figure 3.6 Exemple d'étiquetage des CVs d'un volume de la base de données de MICCAI

3.4.5 Phase expérimentale

Dans notre projet, nous entraînons 6 variantes différentes de notre modèle que nous évaluons durant la phase expérimentale selon qu'ils incluent ou non des blocs scSE et/ou qu'ils utilisent ou non un transfert d'apprentissage des données non scoliotiques vers les données scoliotiques et/ou qu'ils segmentent simultanément ou séparément les DIVs et les CVs. Les informations relatives aux variantes du modèle sont résumées dans le tableau 3.5. Les variantes incluant le bloc scSE entraînées avec les données de patients scoliotiques sont initialisées par le meilleur modèle incluant le bloc scSE entraîné sur 50 epoch (nombre de fois où nous faisons passer les données d'entraînement à travers le réseau) avec les données de la base de données de MICCAI (respectivement vrai dans le cas sans bloc scSE). La raison pour laquelle les modèles n'ayant pas subi un transfert d'apprentissage sont entraînés sur 100 epoch est de préserver l'équité avec ceux qui ont subit un transfert d'apprentissage. Nous voulons éviter de comparer deux modèles ayant été choisis sur un intervalle d'epoch différent.

L'intérêt de cette phase expérimentale est d'étudier l'impact des décisions de conception relatives à notre méthode, à savoir : introduction du bloc scSE, transfert d'apprentissage et segmentation multi-classes des deux structures DIVs et CVs ce qui reviendrait à réaliser une classification *une contre toutes* signifiant une classe à prédire parmi l'ensemble des classes ou bien la réalisation de 2 classifications binaires (aussi appelée classification *une contre une*). D'ailleurs ce dernier point de l'étude expérimentale est une des premières stratégies à discuter dans tout problème de classification incluant un nombre supérieur à deux classes. Intuitivement, on pourrait croire qu'entraîner un modèle pour chacune des classes à réaliser une classification binaire et joindre les différentes prédictions lors de l'inférence serait un moyen plus sûr d'aboutir à un bon résultat que d'entraîner un seul modèle à effectuer la tâche en raison de la facilité d'entraînement. En effet, plus des classes partagent des caractéristiques et plus celles-ci nécessitent d'être complétées par d'autres informations, d'où l'intérêt de cibler des caractéristiques les plus discriminantes pour chacune des classes. Cependant, cette réflexion fait abstraction d'autres facteurs. Une classification *une contre une* favoriserait une meilleure compréhension des classes et réduirait le chevauchement inter-classes dans un cas de segmentation.

Nom modèle	scSE	Transfert d'apprentissage	Nombre de classes	Epoch
se_scol	Oui	Oui	3	50
noSe_scol	Non	Oui	3	50
se_miccai	Oui	Non	3	100
noSe_miccai	Non	Non	3	100
bi_disc	Oui	Oui	2	50
bi3_vertebre	Oui	Oui	2	50

Tableau 3.5 Information sur les modèles entraînés et utilisés dans la phase expérimentale

3.4.6 Algorithme d'entraînement

L'entraînement du modèle se fait en deux étapes totalement supervisées. Nous avons opté pour l'algorithme Adam pour entraîner le réseau avec un taux d'apprentissage α initialisé à $0,0005$ et un paramètre de régularisation L2 initialisé à 0,001. Étant donné que nous proposons un modèle de segmentation 2D, nous utilisons les tranches sagittales des volumes IRM comme entrée du réseau. Ce dernier produit en sortie une carte de probabilité à 3 canaux (un canal pour chacune des classes) et permet ainsi de déduire le masque de segmentation. Nous commençons par entraîner le réseau avec l'ensemble de la base de données de MICCAI, et nous utilisons des images de patients scoliotiques pour réaliser la validation du modèle durant l'entraînement. Nous initialisons les poids du réseau selon l'initialisation de Glorot & Bengio (2010) qui s'adapte aux données en entrée et évite ainsi une initialisation aléatoire. Le réseau est entraîné sur 50 *epoch* tout en sauvegardant les modèles à un intervalle fixe d'itérations par *epoch*. Le meilleur modèle est sélectionné selon ses performances sur les données de validation.

La deuxième étape de l'entraînement est d'incorporer des informations liées aux déformations occasionnées par la scoliose. Le modèle ayant déjà appris la structure globale des images ainsi qu'à reconnaître les DIVs et les CVs, nous avons décidé d'utiliser le minimum d'images scoliotiques pour en laisser le maximum pour le test. Alors, nous avons ré-entraîné le réseau en transférant les paramètres appris par le modèle précédemment sélectionné. Nous avons utilisé pour l'entraînement des images provenant de 3 volumes de patients différents seulement et en ré-utilisant les mêmes données de validation qu'à l'étape précédente. Le reste des images sco-

liotiques est réservé pour la phase de test. Encore une fois, le réseau est entraîné sur 50 *epoch* et le meilleur modèle est sélectionné.

3.5 Mesures d'évaluation

Pour évaluer la performance d'un algorithme de segmentation, on peut se référer à diverses mesures de similitude (ou dissimilitude). Dans la littérature, nous pouvons trouver des mesures supervisées et non supervisées (Zhang *et al.*, 1996). Toutefois, nous nous consacrons seulement aux mesures de similitude de la première catégorie.

Les mesures supervisées évaluent la performance de l'algorithme en le comparant à une vérité terrain. Dans notre projet nous avons évalué notre approche en nous basant sur le coefficient de Dice et la courbe précision-rappel, deux mesures qui prennent en compte la précision (Pr) et le rappel (Ra) :

$$Pr = \frac{|VP|}{|VP| + |FP|} \tag{3.13}$$

$$Ra = \frac{|VP|}{|VP| + |FN|} \tag{3.14}$$

On ne peut se référer à une seule de ces deux mesures individuellement. Dans le cas de segmentation d'images, le rappel (appelé aussi sensibilité dans la littérature) renseigne sur le pourcentage de pixels considérés comme étant positifs par rapport à tous les pixels appartenant à cette classe. Quant à la précision, elle nous informe sur le pourcentage de pixels considérés comme réellement positifs par rapport à tous les pixels prédit par le modèle comme étant positifs. De ce fait, plus le modèle retourne des prédictions positives et plus le rappel est élevé ce qui ne garantit en rien une précision élevée si toutes ces prédictions sont erronées. À l'opposé, si la précision est élevée cela n'indique pas que tous les cas positifs ont été retrouvés. Cependant, à cause du problème de déséquilibre de classes, la précision a tout de même des propriétés très intéressantes comparé à la spécificité, une autre mesure supervisée fréquemment utilisée qui

prend en compte le taux de prédictions négatives correctes qui pourrait biaiser la mesure :

$$Specificite = \frac{|VN|}{|VN| + |FP|} \tag{3.15}$$

Dans notre cas, si notre modèle est capable de prédire correctement l'arrière-plan de l'image, cela ne nous indique pas qu'il est capable de faire de même pour les DIVs et les CVs. Généralement, évaluer un modèle de classification revient à évaluer le compromis entre fausses prédictions et détections manquées. Van Rijsbergen (1979) ont combiné le rappel et la précision en proposant la F_β-mesure :

$$F_\beta = \frac{(\beta^2 + 1)Pr.Ra}{\beta^2.Pr + Ra} \tag{3.16}$$

β est un paramètre qui permet d'ajuster le compromis entre précision et rappel. Sa valeur varie selon le problème mais par défaut $\beta = 1$. D'ailleurs, cette métrique F_1 est surtout connue en imagerie pour être une mesure statistique géométrique des plus utilisées. Elle est introduite par Sørensen (1948) et elle est connue sous le nom de coefficient de Sørensen-Dice ou coefficient de Dice qui mesure l'indice de chevauchement spatial entre la vérité terrain et la prédiction. Nous avons utilisé cette mesure pour évaluer la segmentation produite par notre modèle. Elle est notée comme suit :

$$Dice = \frac{2|VP|}{2|VP| + |FP| + |FN|} \tag{3.17}$$

De plus, nous nous référons à la courbe précision-rappel qui est une représentation graphique de la mesure F_1 et permet de visualiser le compromis entre le rappel représenté en abscisse et la précision représentée en ordonnée. Cette courbe permet d'évaluer le compromis entre ces deux mesures, tout comme nous pouvons la voir comme un compromis entre le taux de faux positifs et de taux négatifs (Exemple d'une courbe précision-rappel dans la figure 3.7). La qualité de cette courbe peut être résumée par une mesure appelée *espace sous la courbe : AUC* qui est calculée par la méthode des trapèzes (Atkinson, 1989). Plus cette valeur est élevée et plus la précision et le rappel le sont simultanément et donc garants de taux de faux positifs et de faux négatifs bas. Donc un AUC élevé signifierait que le modèle retourne un résultat exacte. À titre d'exemple, la figure 3.7 dévoile une courbe bleue obtenue par le modèle A et

une autre courbe orange obtenue par le modèle B. Nous voyons que la courbe bleue se trouve au dessus de la courbe orange. Alors, le modèle A obtient de meilleures performances. Toutefois, si visuellement l'analyse n'est pas évidente, on se réfère à l'AUC pour avoir une comparaison numérique.

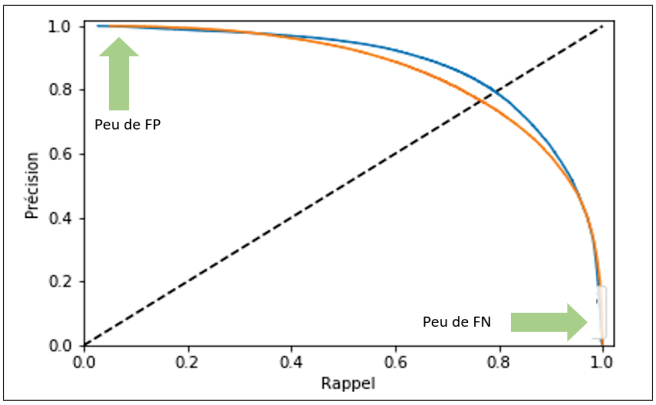


Figure 3.7 Exemple d'une courbe précision-rappel

3.6 Conclusion

Dans ce chapitre, nous avons détaillé la méthodologie de notre travail et argumenté les choix que nous avons pris pour proposer un modèle de segmentation des DIVs et CVs à partir des tranches IRMs de patients scoliotiques sur le plan sagittal. Dans le prochain chapitre, nous analyserons et discuterons les résultats obtenus par notre modèle de segmentation en le comparant à plusieurs variantes pour confirmer nos hypothèses : Recourir au bloc scSE pour doter le modèle d'une capacité de se généraliser, transfert d'apprentissage pour réduire la nécessité d'avoir un grand nombre d'images de patients scoliotiques et entraîner un unique modèle de segmentation des DIVs et CVs au lieu d'entraîner deux modèles à segmenter une classe chacun.

CHAPITRE 4

RÉSULTATS ET DISCUSSION

Nous présentons dans ce chapitre les résultats de segmentation des DIVs et CVs obtenus lors de la phase expérimentale élaborée et détaillée dans la section 3.4.5. Une analyse des performances des six modèles entraînés (se_scol, noSe_scol, se_miccai, noSe_miccai, bi_disc, bi_vertebre) sera faite à travers les études comparatives suivantes :

1. Segmentation avec le modèle incluant ou non le bloc scSE, avec et sans transfert d'apprentissage. (Section 4.1)
2. Segmentation des DIVs avec le modèle entraîné à réaliser la segmentation des deux structures, avec le modèle entraîné à réaliser uniquement la segmentation des DIVs. (Section 4.2)
3. Segmentation des CVs avec le modèle entraîné à réaliser la segmentation des deux structures, avec le modèle entraîné à réaliser uniquement la segmentation des CVs. (Section 4.2)

La validation de l'approche est réalisée sur 198 tranches sur le plan sagittal issues de 7 volumes de patients scoliotiques avec différents degrés de sévérité (Table 3.4). La performance de segmentation est exprimée en termes du coefficient de Dice (Équation 3.5). Nous dévoilons aussi les performances obtenues par les modèles lors de l'entraînement sur les données de validation en termes du coefficient de Kappa de Cohen (Équation 3.7), rappel (Équation 3.14) et précision (Équation 3.13).

4.1 Bloc scSE et transfert d'apprentissage

Le bloc scSE ainsi que le transfert d'apprentissage représentent deux composants essentiels de notre méthodologie. Il est donc important d'étudier leur apport en termes de performances réalisées par le modèle de segmentation simultanée des DIVs et des CVs. Le tableau 4.1 résume les résultats obtenus par les 4 modèles étudiés : avec et sans bloc scSE en incluant le transfert

d'apprentissage (respectivement *se_scol* et *noSe_scol*), avec et sans bloc scSE sans transfert d'apprentissage (respectivement *se_miccai* et *noSe_miccai*) lors de l'entraînement sur les données de validation exprimées en terme de coefficient de Kappa, rappel et précision. Clairement, les modèles ayant bénéficié d'un transfert d'apprentissage obtiennent de meilleurs résultats que les deux autres, quelle que soit la mesure de similitude. À travers le rappel, nous constatons que le modèle *noSe_scol* retourne autant de prédictions positives que le modèle *se_scol* mais que ce dernier obtient un κ plus élevé, ce qui indique qu'il retourne plus de résultats correctement prédits. Nous remarquons aussi que les modèles entraînés avec le bloc scSE obtiennent de meilleurs résultats en comparaison avec leurs équivalents qui en sont dépourvus. Cette validation lors de l'entraînement nous a donc permis d'appuyer notre hypothèse de départ en constatant que le bloc scSE a un impact sur la qualité de la segmentation et nous reconnaissons donc que celui-ci accompagné d'un transfert d'apprentissage est une combinaison clé dans notre approche.

Modèle	**Kappa (κ)**	**Rappel**	**précision**
noSe_miccai	0,744	0,773	0,809
noSe_scol	0,819	0,891	0,82
se_miccai	0,781	0,829	0,806
se_scol	**0,844**	**0,888**	**0,857**

Tableau 4.1 Performance des modèles sur les données de validation

Pour mieux visualiser la distinction entre l'apport du transfert d'apprentissage et celui du bloc scSE, nous présentons des résultats quantitatifs qui se trouvent dans les figures 4.1, 4.2 et 4.3, ainsi que des résultats qualitatifs dans les figures 4.4, 4.5 et 4.6. La figure 4.1 dévoile les performances obtenues par les 4 modèles à réaliser la segmentation des DIVs et les CVs au niveau du pixel dans l'image pour chaque volume de patient scoliotique de la base de données de test. Le coefficient de Dice de chaque tranche du volume est calculé et la moyenne par volume est représentée sur le graphique. Nous confirmons à partir des deux graphiques que le modèle *se_scol* surpasse les 3 autres quelque soit le volume de test et que le modèle *noSe_miccai* obtient les moins bonnes performances à chaque fois.

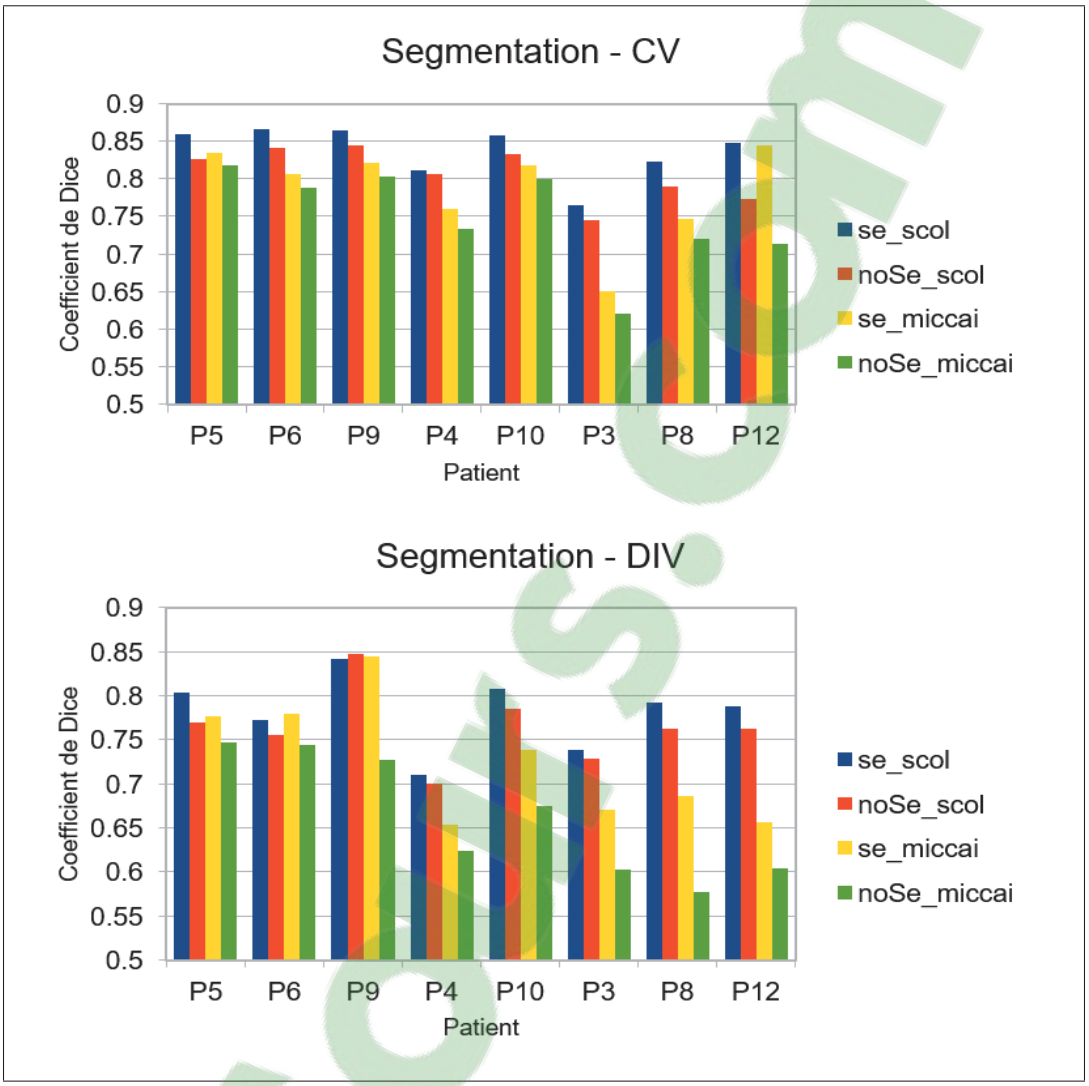


Figure 4.1 Performances des 4 modèles à segmenter les CVs et les DIVs par volume de patient exprimées en terme de coefficient de Dice. Les patients sur l'axe des X sont ordonnés selon le degré croissant de sévérité de la scoliose

Nous remarquons que la moyenne des coefficients de Dice par volume baisse en fonction de la gravité de la scoliose. Toutefois, cette régression est plus importante dans le cas des modèles n'ayant pas bénéficié de transfert d'apprentissage. Dans le cas des DIVs, le coefficient de Dice associé aux modèles *se_scol* et *noSe_scol* a un écart-type évalué sur l'ensemble des volumes de 0.04 tandis que celui des modèles *se_miccai* et *noSe_miccai* a un écart-type de 0.07. Pour ce qui est des performances en tant que telles, le modèle *se_scol* atteint une moyenne de coefficient

de Dice 80% au niveau pixel dans les cas de scoliose légère et 75% dans les cas de scoliose modérée et sévère. Dans le cas des CVs, ce même modèle atteint une moyenne de 86% de bonnes prédictions dans le cas de scoliose légère, 83% dans le cas de scoliose modérée et 80% dans le cas de scoliose sévère. Les résultats sont meilleurs dans le cas des CVs parce qu'ils couvrent un plus grand nombre de pixels dans l'image que les DIVs et que les modèles n'ont pas beaucoup de difficulté à classifier correctement l'intérieur d'un CV. Nous supposons donc que la différence entre les modèles se joue principalement au niveau des contours et de l'habileté à localiser les DIVs et les CVs déformés ou qui n'apparaissent pas totalement dans l'image.

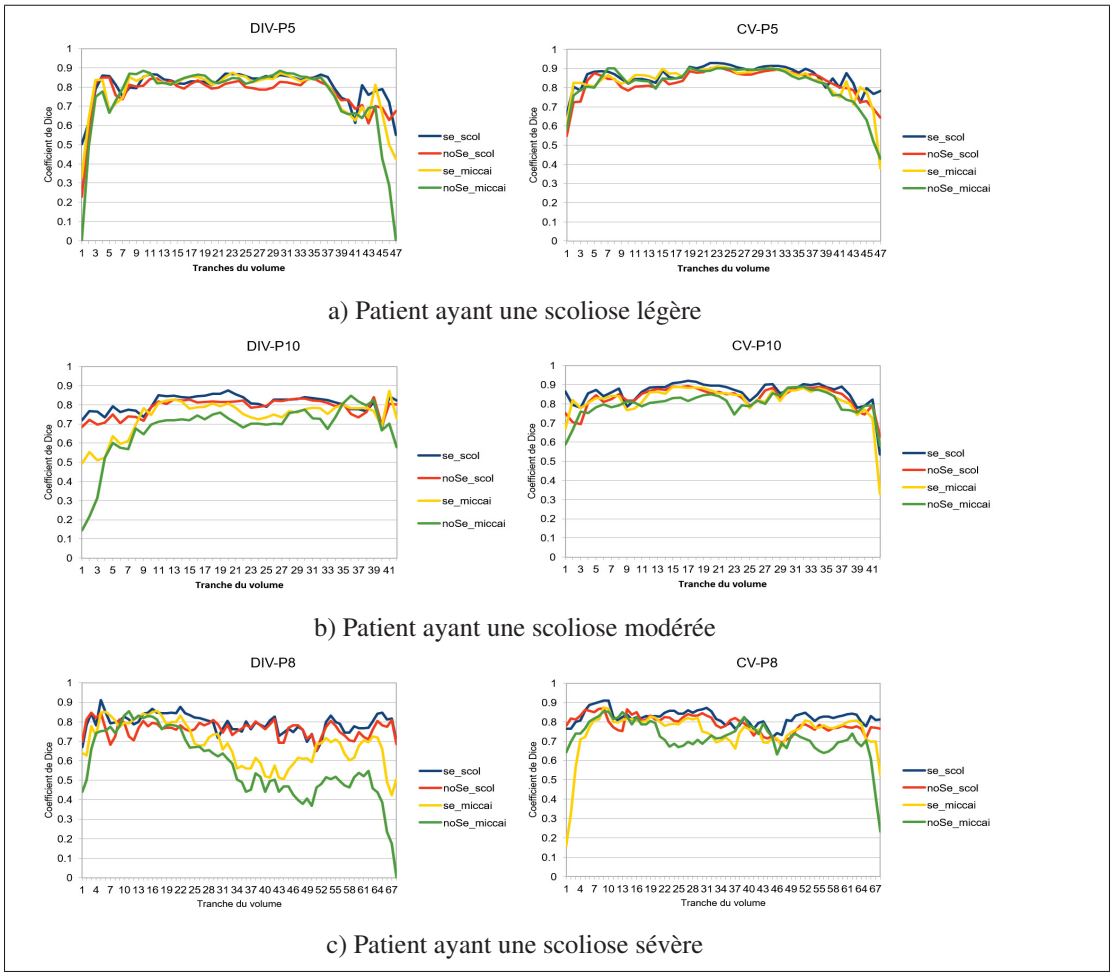
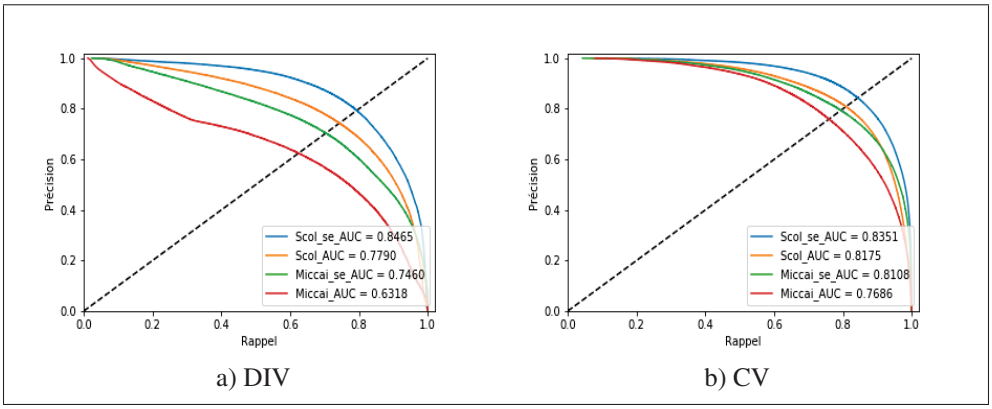


Figure 4.2 Résultats de segmentation tranche par tranche sur des volumes de patients ayant un degré de sévérité différent

Pour mieux visualiser le comportement des modèles face aux différentes images du volume (extrémités, scoliose non apparente, scoliose apparente), la figure 4.2 présente la performance des 4 modèles à segmenter tranche par tranche les DIVs et les CVs dans les 3 niveaux de sévérité de la scoliose. Nous remarquons que dans les premières tranches des trois volumes pour les deux structures anatomiques, le coefficient de Dice est plus bas par rapport à la moyenne du volume ce qui pourrait s'expliquer par le fait que ce sont des images où la forme des deux structures n'est pas tout à fait complète et que la prédiction est donc moins évidente. Par ailleurs, ce type d'images est peu présent dans nos données d'entraînement donc les modèles n'arrivent pas à complètement généraliser les caractéristiques liées aux deux structures. Dans le cas de scoliose légère (Figure 4.2a), nous remarquons que les quatre modèles obtiennent des performances proches mais que celles-ci baissent dans les 10 dernières tranches où la scoliose est apparente. Dans les cas de scoliose modérée et sévère (Figures 4.2b et 4.2c), les performances des différents modèles sont un peu plus éloignées les uns des autres et on peut mieux s'apercevoir de la supériorité du modèle *se_scol*. Cependant, ce qui est plus intéressant à observer est l'apport du transfert d'apprentissage notamment dans le cas de la scoliose sévère au niveau de la segmentation des DIVs où on remarque un contraste entre les modèles *se_scol* et *noSe_scol* d'un côté qui témoignent une bonne stabilité de tranche en tranche et les modèles *se_miccai* et *noSe_miccai* de l'autre. Toutefois, le modèle *se_miccai* montre une capacité à rehausser sa prédiction en atteignant même les performances du modèle *noSe_scol* dans la région où la scoliose est apparente dans le cas des CVs. Nous supposons alors que le modèle *noSe_scol* étant plus spécifique que le modèle *se_miccai* a une meilleure habileté à distinguer les différentes structures présentes dans les images tests car leur apparence lui est plus familière. Mais le modèle *se_miccai* est doté d'une meilleure capacité de généralisation et dispose d'une meilleure compréhension des structures DIVs et CVs malgré l'‘absence dans la base de données d'apprentissage d'images issues de la même base de données que les images tests.



a) DIV b) CV

Figure 4.3 Comparaison des courbes précision-rappel obtenues suite à la segmentation des DIVs et CVs à l'aide des 4 modèles

Pour aller plus loin dans notre étude comparative, nous portons cette fois-ci notre attention uniquement sur les tranches où la scoliose est visible afin d'observer l'impact du bloc scSE et le transfert d'apprentissage sur la scoliose uniquement. La figure 4.3 représente deux ensembles de courbes précision-rappel propres aux DIVs et aux CVs et résume les résultats des 4 modèles. À partir des deux graphiques, nous aboutissons au même constat qu'avec le coefficient de Dice. Le modèle *se_scol* propose le meilleur compromis entre le nombre de prédictions positives et le taux de prédictions positives correctes et inversement pour le modèle *noSe_miccai* qui lui semble retourner beaucoup de prédictions positives incorrectes. Il est toutefois intéressant de voir que pour les deux structures anatomiques, dans le cas où le rappel est élevé et la précision est faible, le modèle *se_miccai* a un compromis entre taux de faux positifs et faux négatifs égal à, voire même plus favorable que celui du modèle *noSe_scol* qui semble produire plus de fausses prédictions que le modèle *se_scol*. Cela appuierait notre analyse de l'effet du bloc scSE qui semble jouer un rôle déterminant dans le choix des caractéristiques qui décrivent les structures DIVs, CVs.

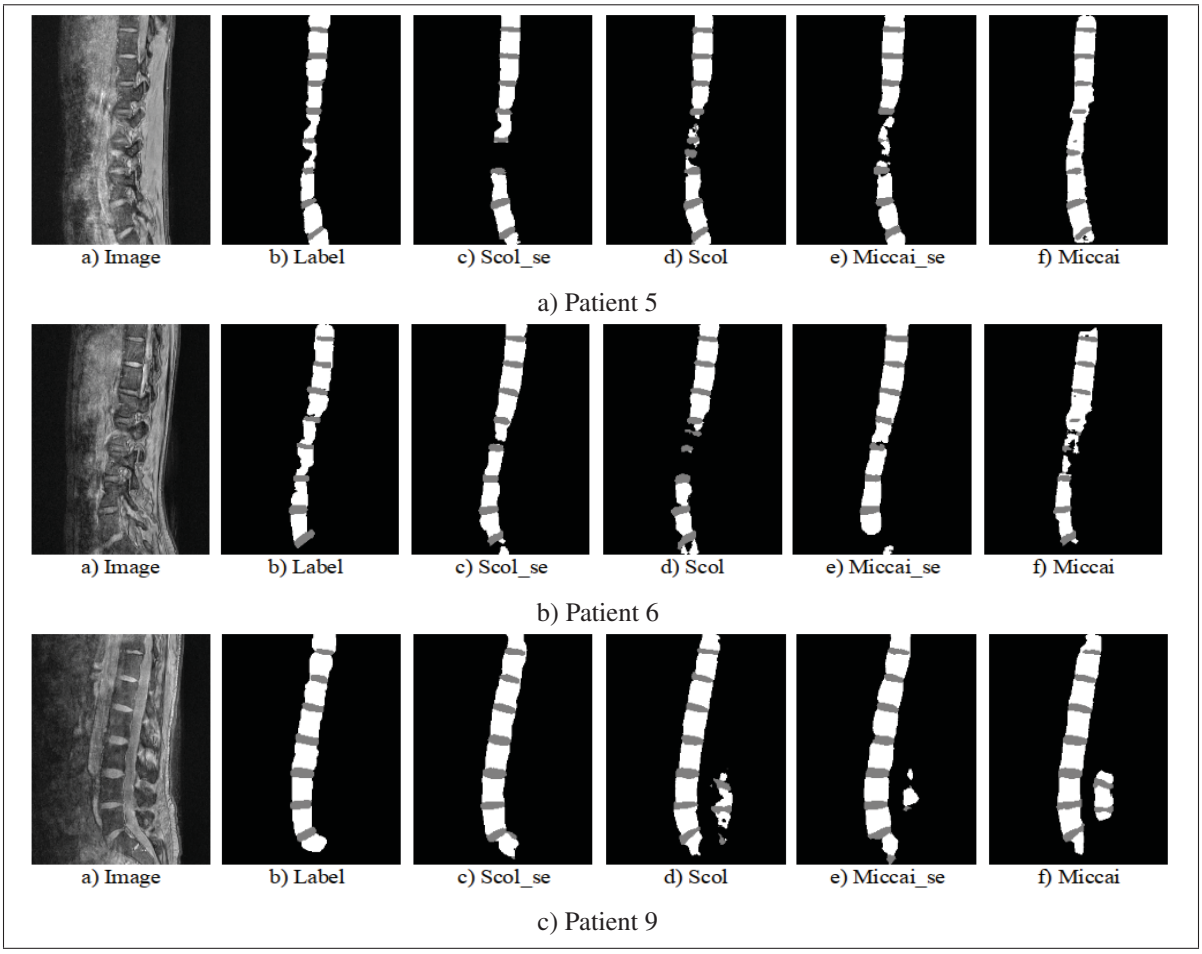


Figure 4.4 Segmentation des DIVs et CVs à partir d'une tranche du volume des patients atteints de scoliose légère avec les 4 modèles

Pour appuyer l'analyse faite à partir des résultats quantitatifs, nous nous référons à des résultats qualitatifs où nous segmentons une tranche de chaque volume de patient où la scoliose est apparente avec les 4 modèles. Nous pouvons apercevoir à partir des figures 4.5a, 4.5b et 4.6c que les modèles *se_scol* et *se_miccai* sont capables d'identifier les CVs cervical et thoracique se trouvant respectivement aux extrémités supérieure et inférieure de la colonne vertébrale que nous avons même omis de segmenter dans notre étiquetage manuel sur certaines tranches. Ces vertèbres sont difficiles à prédire parce qu'elles sont souvent mal définies dans les images où seulement une partie peut être aperçue. Les modèles sont entraînés sur des images de taille réelle où la plupart des vertèbres ont une forme globale et une composition assez similaire.

Donc, ce genre d'exception est moins facile à reconnaître et nous constatons que les architectures des modèles dotées du bloc scSE ont une meilleure capacité à cet égard. À partir des figures 4.6a et 4.4b nous constatons que le modèle *se_miccai* arrive à retrouver un CV déformé alors que le modèle *noSe_scol* l'a considéré comme faisant partie de l'arrière-plan de l'image. Et inversement, dans la figure 4.6c, alors qu'une partie de la colonne vertébrale est cachée par d'autres structures anatomiques, le modèle *noSe_scol* a fait une mauvaise prédiction et a considéré une partie de cette région comme étant un CV. Le même phénomène est observé avec les DIVs dans la figure 4.4a où le modèle *noSe_scol* a considéré un nerf spinal de la moelle épinière rendu visible à cause de la déformation comme étant des DIVs. La figure 4.4c est un exemple parfait pour montrer l'apport du bloc scSE. Un artefact d'acquisition du volume IRM de ce patient cause une distorsion de l'image où la colonne vertébrale est visible en double sur la partie droite. Le modèle *se_scol* est le seul à avoir résisté à cette anomalie contrairement aux trois autres qui ont fait de fausses prédictions. Toutefois, le modèle *se_miccai* en a fait moins que le modèle *noSe_scol* qui ne semble pas être complètement apte à distinguer les vrais CVs et DIVs des fausses structures.

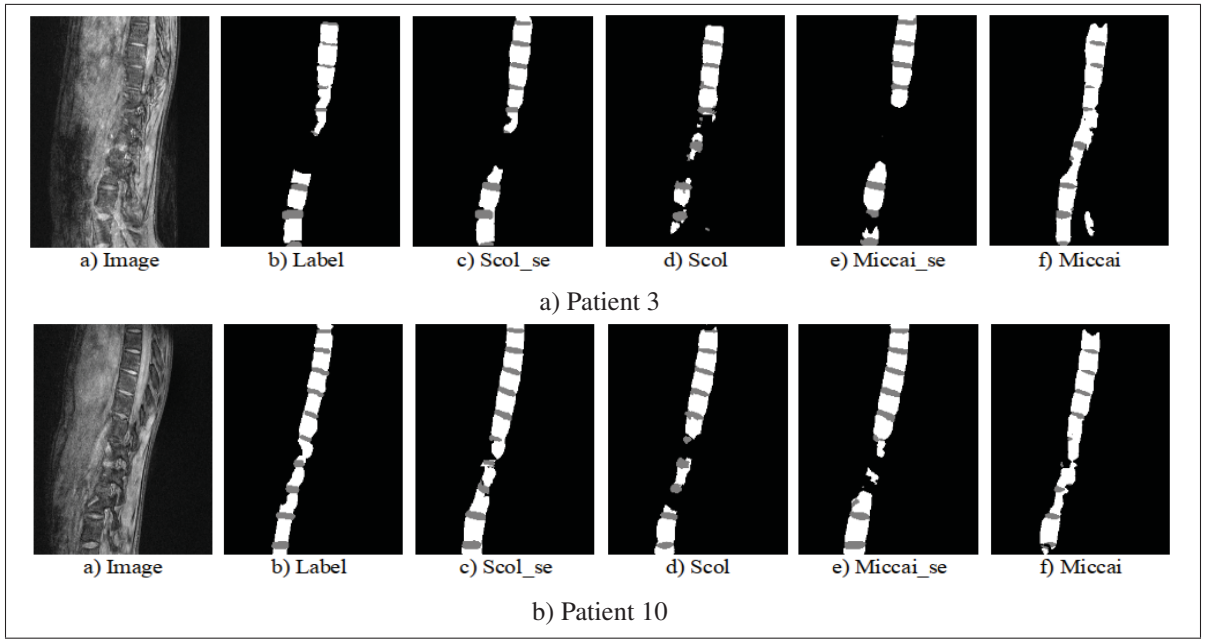


Figure 4.5 Segmentation des DIVs et CVs à partir d'une tranche du volume des patients atteints de scoliose modérée avec les 4 modèles

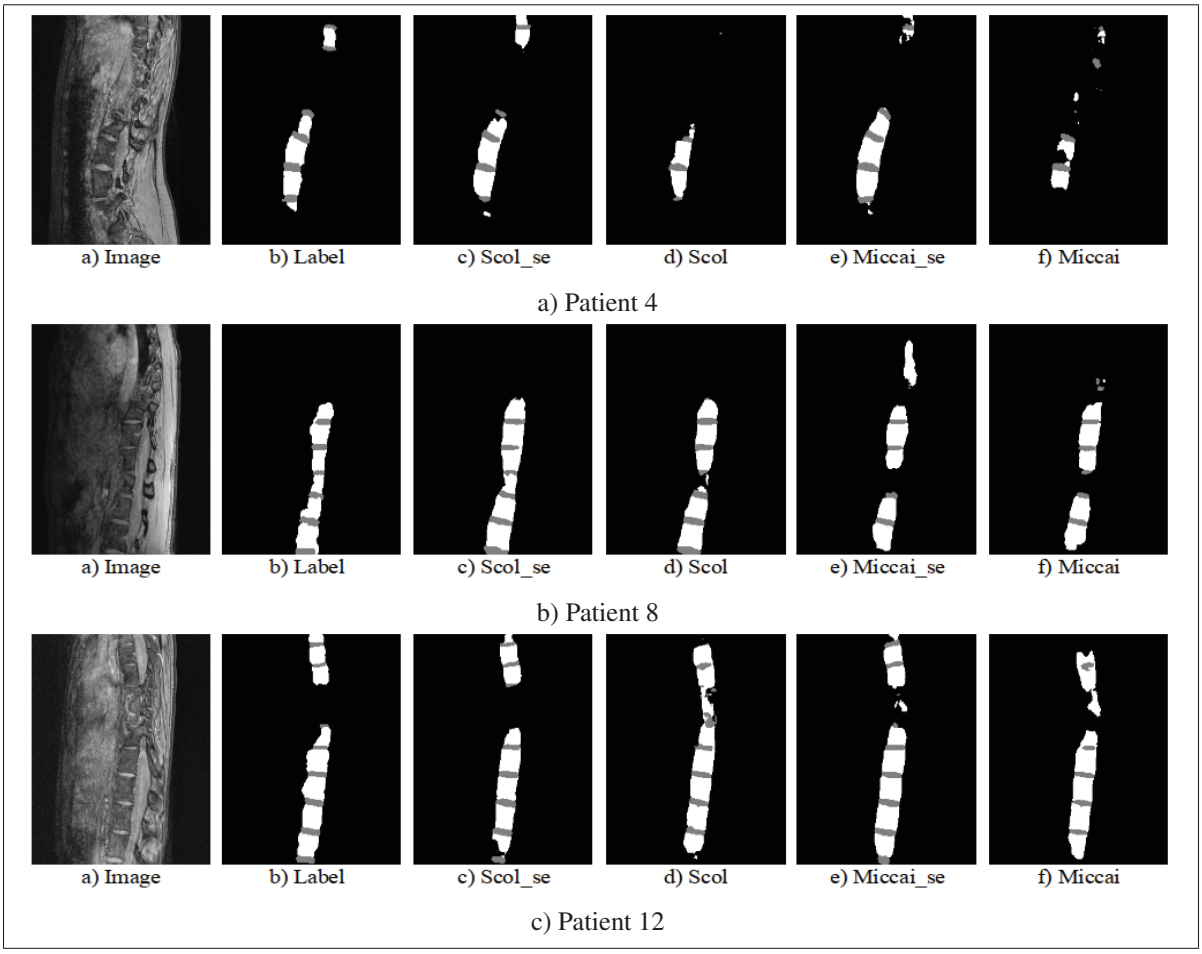


Figure 4.6 Segmentation des DIVs et CVs à partir d'une tranche du volume des patients atteints de scoliose sévère avec les 4 modèles

4.2 Comparaison entre segmentation multi-classes et segmentation bi-classes

Nous nous penchons à présent sur la question de la stratégie de classification via un unique modèle *une contre toutes* ou bien 2 modèles *une contre une*. Pour répondre à cette question, nous avons entraîné deux autres modèles similaires à *se_scol* que nous appelons *bi_disc* et *bi_-Vertèbre* qui ont pour tâche de segmenter uniquement les DIVs pour le premier et les CVs pour le second. Cette fois-ci encore, nous présentons les résultats selon la moyenne des coefficients de Dice par volume (Figure 4.7) et une courbe précision-rappel (Figure 4.8) en comparant

les performances de chacun des deux nouveaux modèles avec le modèle *se_scol* utilisé pour segmenter uniquement la structure respectivement ciblée par chacun des nouveaux modèles.

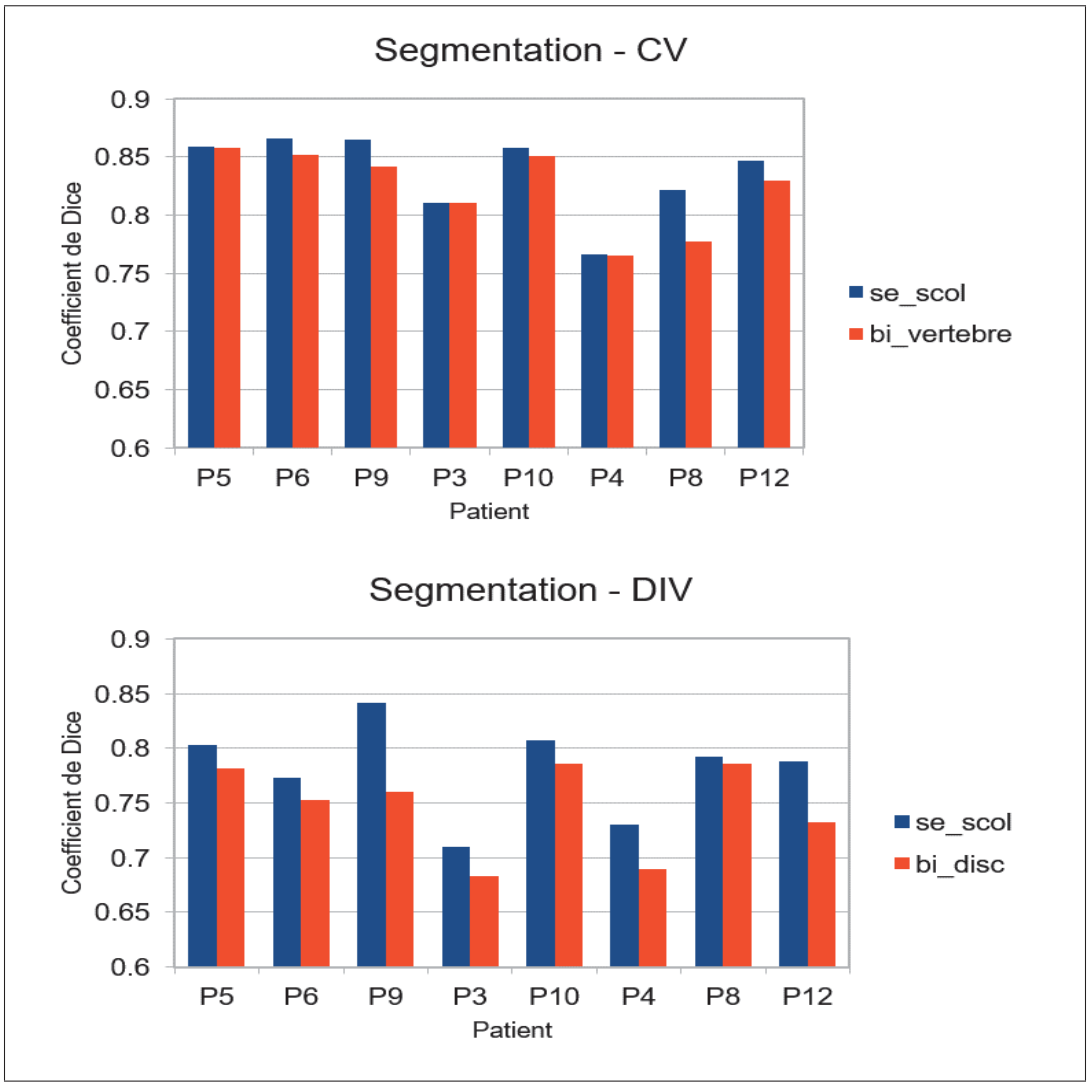
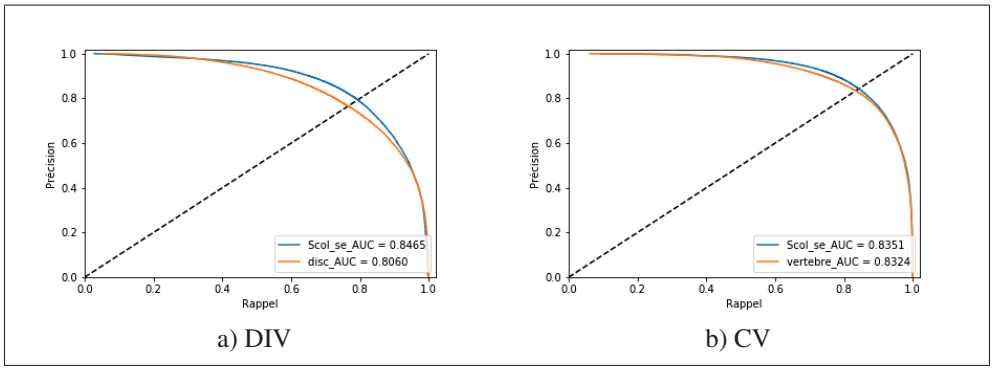


Figure 4.7 Performances du modèle *se_scol* à réaliser une segmentation des CVs et DIVs séparément en comparaison aux modèles *bi_vertebre* et *bi_disc*

À partir de la figure 4.7 nous pouvons constater qu'en termes de coefficient de Dice, le modèle *se_scol* offre des performances supérieures pour les deux structures. Cela s'explique probablement par la forte corrélation qu'il y a entre les CVs et les DIVs. Ces derniers sont connectés ce qui aide à définir les bordures supérieures et inférieures et une segmentation simultanée des

deux structures limiterait considérablement les risques de chevauchement. Cela est appuyé par les AUCs présentés dans la figure 4.8 qui sont à l'avantage du modèle *se_scol*, particulièrement dans le cas des DIVs. En effet, la déformation liée aux DIVs est plus subtile que celle des CVs parce qu'ils occupent moins de pixels, ont tendance à rétrécir dans l'image et deviennent très difficiles à distinguer des tissus mous entourant la vertèbre. L'absence d'information sur la présence de CVs dans le cas du modèle *bi_disc* accentue le risque de fausses prédictions dans ces circonstances.



a) DIV b) CV

Figure 4.8 Comparaison des courbes précision-rappel obtenues suites à la segmentation des DIVs à l'aide des modèles *se_scol* et *bi_disc* et des CVs à l'aide des modèles *se_scol* et *bi_vertebre*

Nous appuyons l'analyse des résultats quantitatifs présentés dans les figures 4.7 et 4.8 par des résultats qualitatifs obtenus en segmentant les DIVs et CVs séparément (Figures 4.9 et 4.10). Nous remarquons à partir de la figure 4.9b que le modèle *bi_disc* est moins robuste que le modèle *scol_se* dans la région où la scoliose est apparente et ne semble pas sûr de sa prédiction face à un chevauchement de tissus au tour de la colonne. On remarque aussi un cas de faux positif où le modèle *bi_disc* a prédit une apophyse épineuse comme étant un DIV. L'unique différence entre ce modèle et le modèle *scol_se* est l'information de délimitation du DIV par du CV présente dans le deuxième qui est fortement bénéfique pour réduire le taux de faux positifs. Nous pouvons confirmer cette déduction en observant la figure 4.10b où les contours des CVs semblent mal définis et chevauchent la région des DIVs.

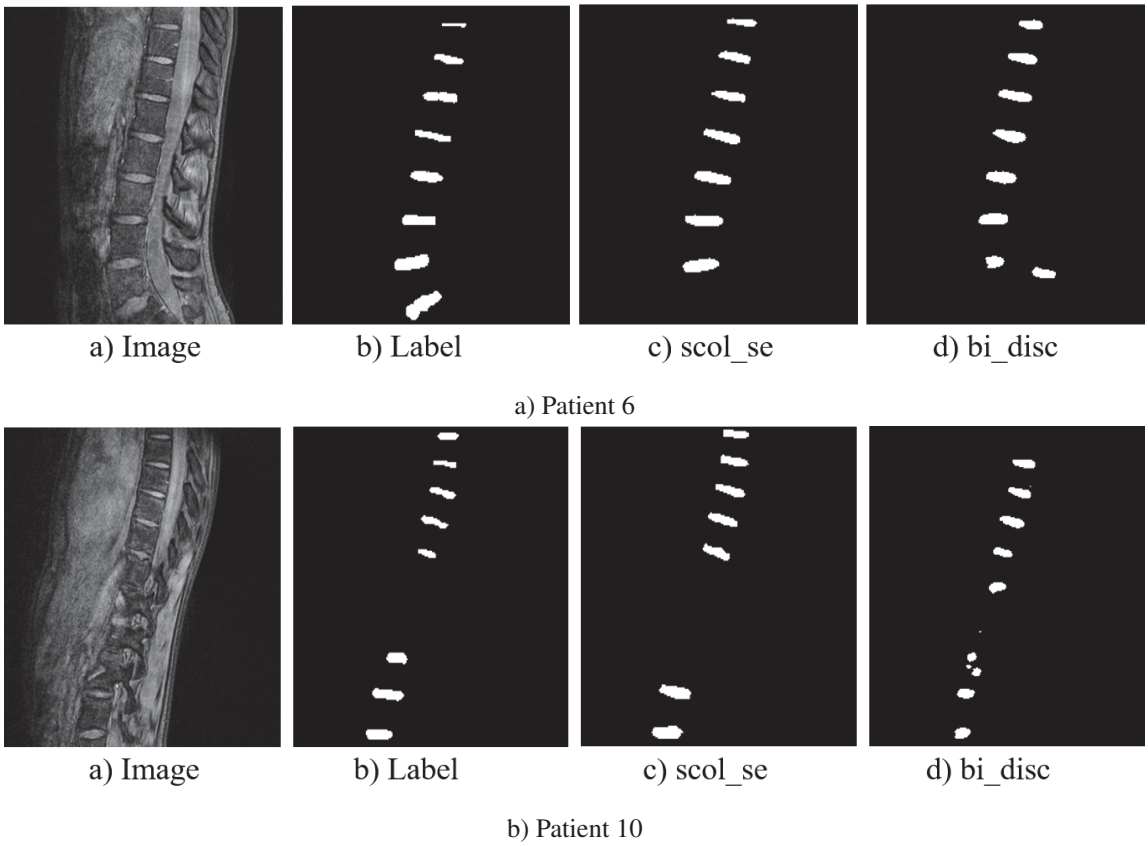


Figure 4.9 Segmentation des DIVs à partir d'une tranche du volume de patients scoliotiques avec les deux modèles scol_se et bi_disc

4.3 Discussion et conclusion

À travers ce chapitre, nous avons dévoilé les résultats obtenus par notre approche de segmentation simultanée des DIVs et CVs à partir d'IRM de patients scoliotiques. Malgré le nombre très limité d'images de patients scoliotiques disponibles lors du ré-entraînement du modèle, notre approche a atteint une segmentation qui est jugée selon les mesures de similarité employées et les résultats qualitatifs observés comme réussie et très encourageante au vu de la difficulté accrue par les déformations de la scoliose. Par la même occasion, ceci nous a permis de confirmer nos hypothèses de départ. Le transfert d'apprentissage s'est avéré une stratégie efficace pour adapter un modèle plus générique à la base de données des images de patients scoliotiques et ainsi être plus robuste face aux déformations anatomiques de la colonne vertébrale.

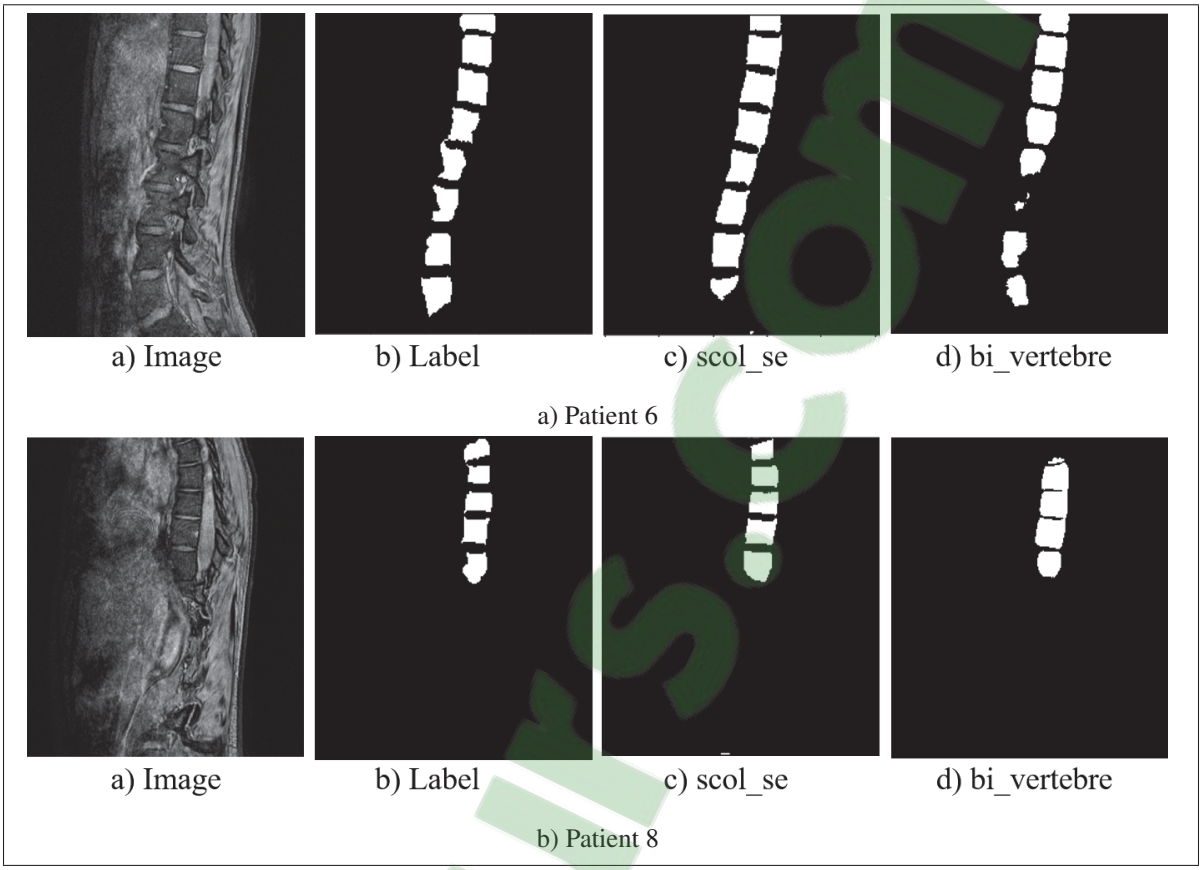


Figure 4.10 Segmentation des CVs à partir d'une tranche du volume de patients scoliotiques avec les deux modèles scol_se et bi_vertebre

Le bloc scSE a augmenté la capacité du modèle à généraliser les caractéristiques discriminantes et ainsi élargir la représentation des structures DIVs et CVs. Nous avons aussi montré que réaliser une segmentation des structures DIVs et CVs est plus intéressante que de réaliser une segmentation binaire pour chaque structure. Nous pensons que la forte corrélation qui existe entre les deux structures est un facteur déterminant de ce succès. Toutefois, nous avons pu constater les limites de ce que nous proposons. Tout d'abord, l'élaboration de notre vérité terrain est réalisée par une personne unique qui n'est pas experte dans le domaine médicale. Par conséquent, la réduction de l'erreur humaine via un consensus de plusieurs observations n'est pas réalisée. Également, la validation d'un expert dans les cas les plus douteux n'a pas été faite, ce qui nous a conduit à écarter certaines images. Nous avons pu observer que le premier point est problématique dans le sens où certains résultats obtenus auraient pu être meilleurs comme

dans les cas observés où notre modèle retrouve des CVs qui n'ont pas été pris en compte dans l'étiquetage manuel. Ceci a comme répercussion de ne pas mesurer avec une précision absolue l'impact de nos choix méthodologiques. Pour ce qui est des limites de l'approche même, nous en avons constaté quelques unes :

1. le modèle de segmentation n'est pas assez robuste pour reconnaître n'importe quelles formes que les structures DIVs et CVs peuvent prendre dans l'image. Cela est facilement visible à partir des premières tranches du volume où les DIVs et CVs commencent à être aperçues et ne sont pas complètement formés c.à.d., les parois ne sont pas toutes visibles. Ce cas de figure est aussi visible à partir des tranches où la scoliose est apparente et que le muscle dorsal recouvre une partie de la colonne vertébrale. Ceci est problématique car les caractéristiques liées aux contours qui sont d'une grande importance dans la localisation des DIVs et CVs ne sont pas présents. Le modèle apprend qu'un DIV et un CV est une structure rectangulaire sujette à une déformation. Plus leur forme s'éloigne de cette représentation et plus il est difficile au modèle de les segmenter correctement. Toutefois, dans la littérature, cette difficulté n'a jamais été traitée et les validations d'approches de segmentation 2D se font sur les tranches médianes du volume.
2. Les performances obtenues n'ont pas été comparées aux approches présentes dans la littérature. Ceci revient principalement à l'unicité de la base de données et à son étiquetage. Aussi, nous n'avons pas pu exploiter l'étiquetage partiel des DIVs fourni avec les données provenant du CHU Sainte-Justine car la majorité est réalisée sur une autre modalité de ces images dont la dimension et la résolution sont différentes et que nous n'avons pas exploité. Ce point-ci amène à une autre limite de notre approche qui est la restriction du champ de vision du réseau de neurones aux dimensions de nos images d'entraînement et qui n'est donc pas robuste face à une importante variation à l'échelle. Ceci nous a pénalisé pour appliquer et valider notre modèle de segmentation sur cette deuxième modalité et ainsi comparer nos performances avec l'unique étude qui a utilisée la même base de données pour valider sa méthode de segmentation des DIVs (Chevrefils *et al.*, 2009). De plus, ce qui nous intéresse à travers ce projet est la segmentation simultanée de l'ensemble des

DIVs et CVs dans une IRM sur le plan sagittal. Or, l'étiquetage partiel fourni consiste en un étiquetage d'un seul DIV par volume.

3. L'approche proposée n'a pas été testée sur d'autres modalités d'IRM et nous ne pouvons valider sa robustesse face aux variations qui existent d'une modalité à une autre. Mais encore, l'unicité de notre base de validation et notre intérêt pour la segmentation simultanée des DIVs et des CVs font que nous ne pouvons situer notre travail dans ce qui est proposé dans la littérature.

CHAPITRE 5

CONCLUSION

À travers ce projet, nous nous sommes penchés sur la segmentation simultanée des DIVs et CVs à partir d'IRMs de patients scoliotiques. L'intérêt de développer une approche robuste et automatique pour réaliser cela est nourri par la nécessité d'y recourir pour fournir un système de navigation par ordinateur lors du traitement de la scoliose idiopathique au cours d'une discectomie minimalement invasive qui a pour objectif de freiner la courbure de la colonne vertébrale et la redresser. Notre approche est basée sur un réseau neuronal convolutif de type encodeur décodeur dans lequel nous avons introduit le bloc scSE pour recalibrer les caractéristiques extraites au fil de l'apprentissage et mettre en avant les plus pertinentes. Nous avons contourné le problème de déséquilibre de classes présent dans les images où l'arrière plan est plus dominant que les DIVs et les CVs en ajoutant le coefficient de Kappa de Cohen à l'entropie croisée pour former la fonction objectif à minimiser. Vu la petite taille de notre jeu de données de patients scoliotiques, le modèle est entraîné dans un premier temps avec des images de sujets non scoliotiques et ré-entraîné sur un petit jeu de données de patients scoliotiques. Plusieurs variantes du modèle ont été entraînées et comparées entre elles selon le calcul du coefficient de Dice et une courbe précision-rappel pour évaluer l'impact de nos choix méthodologiques. Les résultats montrent que le modèle ayant bénéficié d'un transfert d'apprentissage et doté du bloc scSE est le plus performant. Le transfert d'apprentissage a permis au modèle d'étendre sa représentation des deux structures aux déformations qu'ils peuvent subir. Quant au bloc scSE, il lui a permis d'avoir la capacité de se généraliser et de cibler les caractéristiques discriminantes des DIVs et des CVs.

L'étude effectuée à travers ce projet de maîtrise a abouti à une publication d'un article revu par les pairs dans les actes de la conférence IEEE International Symposium on Biomedical Imaging (Guerroumi *et al.*, 2019) où une nouvelle méthode de segmentation des DIVs et CVs à partir d'IRM de patients scoliotiques est introduite dans la littérature. Cette approche se distingue de l'existant par l'attention qu'elle porte à la scoliose. L'objectif principal de cette étude

étant accompli, les hypothèses posées étant confirmées, la méthode est maintenant sujette à diverses améliorations qui peuvent la porter vers un déploiement dans un environnement clinique pour apporter une importante assistance dans l'étude, diagnostique ou traitement de la scoliose. Comme idées d'améliorations futures, nous proposons en premier lieu de s'assurer de la totale intégrité des données en variant l'observation de l'étiquetage et valider cela auprès d'un expert. Pour ce qui est de la base de données, nous suggérons d'accroître le nombre d'images scoliotiques et varier les modalités d'acquisitions. Ce dernier point peut être pris en compte dans l'architecture en introduisant un chemin propre pour chaque modalité et les combiner par la suite. Nous proposons aussi d'accroître le champ de vision du réseau en lui introduisant des entrées de différentes échelles d'une même région de l'image ou d'utiliser des opérations de convolutions dilatées. S'assurer de l'invariance à l'échelle permettrait de valider notre approche sur des images qui ne disposent pas de la même dimension que celle de nos images d'entraînement et par la même occasion effectuer la comparaison avec l'unique étude qui a eu recours à la même base de données d'images scoliotiques pour tester son approche (Chevrefils *et al.*, 2009), ce qui consisterait dans notre projet une étape de validation importante à effectuer. Une segmentation 3D est aussi une idée d'amélioration qui permettrait de prendre en compte l'information volumique. Enfin, pour pallier le manque crucial de données de patients scoliotiques, nous proposons de recourir à un auto-encodeur variationnel dans le but d'apprendre une représentation latente de ces données scoliotiques, qui permettrait de générer de nouvelles images qui pourraient être utilisées durant l'entraînement. Le modèle existant aujourd'hui pourrait être utilisé comme point de comparaison de futures études qui visent à aller au delà de ce que nous avons apporté et ceci est d'une importance capitale. Étant les seuls à s'être confrontés à cette problématique en ayant recours à l'apprentissage profond, notre étude dispose par défaut d'un statut de référence. Mais aussi d'un point de vue pratique, notre méthode peut être utilisée comme outil de base pour annoter des DIVs et CVs chez des sujets atteints de scoliose.

ANNEXE I

1. Fonctions optimisation

Sutskever *et al.* (2013) ont proposé le *Nosterov gradient accéléré* (NAG) qui est une modification de la méthode basée sur le momentum en influençant cette fois-ci le calcul du gradient de la dérivé de la fonction de coût en lui ajoutant le vecteur de vitesse v_t :

$$v_{t+1} = \gamma v_t - \alpha \nabla_w J(w_t + \gamma v_t) \tag{A I-1}$$

$$w_{t+1} = w_t + v_{t+1} \tag{A I-2}$$

Une telle approche permet de donner une meilleure approximation de la prochaine position des paramètres en évitant de dévier la trajectoire.

Les améliorations apportées par le momentum et NAG ont nettement amélioré la qualité et la rapidité de la convergence. Mais toujours est-il que le calcul d'un taux d'apprentissage α adaptatif selon l'importance de l'optimisation n'est pas toujours pris en compte.

Un des premiers algorithmes proposé pour réaliser cela est l'algorithme Adagrad introduit par Duchi *et al.* (2011). Le principe est d'utiliser un α différent pour calculer le gradient de chaque paramètre w_i à chaque instant t. On notera ce gradient $g_{i,t}$:

$$g_{i,t} = \nabla_w J(w_{i,t}). \tag{A I-3}$$

Pour réaliser maintenant l'optimisation, l'algorithme va calculer un α pour chaque paramètre $w_{i,t}$ en se basant sur les précédents gradients calculés pour w_i :

$$G_{ii,t} = \sum_{t=1}^{t} g_{i,t}^2 \tag{A I-4}$$

$$w_{i,t+1} = w_{i,t} - \frac{\alpha}{\sqrt{G_{ii,t} + \varepsilon}} . g_{i,t} \tag{A I-5}$$

Où $G_{ii,t}$ est une matrice diagonale où chaque élément (i,i) est égal à la somme des gradients au carré du paramètre w_i jusqu'à l'instant t. ε est introduit pour éviter la division par zéro. En réalité, l'équation A I-6 peut être substituée par une autre équation qui introduit un produit vectoriel entre G_t et g_t :

$$w_{i,t+1} = w_{i,t} - \frac{\alpha}{\sqrt{G_t + \varepsilon}} \odot g_t \tag{A I-6}$$

Le problème de Adagrad est qu'au fur et à mesure de l'entraînement, la somme des gradients au carré s'accroît au point où α diminue excessivement et tend vers 0, ce qui handicape l'apprentissage du modèle. De plus, l'algorithme original ne prend pas en compte le momentum.

Pour corriger l'inconvénient majeur d'Adagrad, Zeiler (2012) a proposé l'algorithme Adadelta. Le principe est de ne prendre en considération que les p derniers gradients calculés au lieu de l'ensemble. De plus, le stockage des précédents gradients est remplacé par une somme récursive de leur moyenne en décomposition. On note :

$$E[g^2]_{t+1} = \gamma E[g^2]_t + (1-\gamma) g^2_{t+1} \tag{A I-7}$$

Où γ joue le même rôle que le terme du momentum. L'auteur a aussi éliminé l'initialisation du taux d'apprentissage α. Celui ci est calculé comme une somme des moyennes de décompositions (comme avec les gradients) mais cette fois-ci en prenant directement compte des paramètres :

$$E[\Delta w^2]_{t+1} = \gamma E[\Delta w^2]_t + (1-\gamma) \Delta w^2_{t+1} \tag{A I-8}$$

En introduisant la fonction d'erreur quadratique moyenne RMS, la règle d'optimisation d'Adadelta peut s'écrire ainsi :

$$RMS[g]_t = \sqrt{E[g^2]_t + \varepsilon} \tag{A I-9}$$

$$RMS[\Delta\theta]_t = \sqrt{\Delta\theta^2]_t + \varepsilon} \tag{A I-10}$$

$$\Delta\theta_{t+1} = -\frac{RMS[\Delta\theta]_{t-1}}{RMS[g]_t} g_t \tag{A I-11}$$

$$\theta_{t+1} = \theta_t + \Delta\theta_{t+1} \tag{A I-12}$$

1.1 Détail du calcul de la fonciton Adam

Adam estime le taux d'apprentissage α pour chaque paramètre w_i en estimant le moment d'ordre 1 (moyenne) et le moment d'ordre 2 (variance) des précédents gradients. Le calcul de la variance est en réalité la somme de décomposition moyenne des précédents gradients au carré de Adadelta qu'on note v_t. La moyenne est la somme de décomposition moyenne des précédents gradients qu'on note m_t :

$$v_{t+1} = \gamma_1 v_t + (1-\gamma_1) g_{t+1}^2 \tag{A I-13}$$

$$m_{t+1} = \gamma_2 m_t + (1-\gamma_2) g_{t+1} \tag{A I-14}$$

Où γ_1 et γ_2 jouent le rôle d'un terme momentum. L'initialisation de v_t et m_t est mise à zéro. Pour éviter que l'estimation ne soit biaisée et tende dramatiquement vers zéro, les auteurs proposent une correction de l'estimation :

$$\hat{v_t} = \frac{v_t}{1-\gamma_1^t} \tag{A I-15}$$

$$\hat{m_t} = \frac{m^t}{1-\gamma_2^t} \tag{A I-16}$$

En introduisant ces deux équations, la règle d'optimisation d'Adam est :

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v_t}} + \varepsilon} \hat{m_t} \tag{A I-17}$$

On remarque bien que le taux d'apprentissage α est ajusté par un facteur inversement proportionnel à l'amplitude moyenne des gradients précédents. Ce qui est l'effet désiré dans une phase d'apprentissage.

2. La fonction sigmoïde et la fonction de l'entropie croisée dans un cas de classification

Dans le cas d'un apprentissage visant à réaliser une classification, un réseau de neurone quelque soit son type peut être perçu comme un générateur d'une distribution de probabilité $p(w,x)$ qui soit capable de prédire le label d'une observations x_i selon ses paramètres w. En terme statistique, cet apprentissage n'est qu'une maximisation de la vraisemblance L de w étant donné un ensemble d'observation x. On note mathématiquement :

$$L(w|x_1,x_2,...,x_n) = \prod_{i=1}^{n} p(x_i|w) \tag{A I-18}$$

En sachant que maximiser un produit revient à maximiser la somme du logarithme sur toutes les observation, en réseau de neurone on parle plus généralement de minimiser l'opposé du logarithme. Mathématiquement les deux opérations sont similaires. Ce raisonnement est la définition même de la fonction de perte entropie croisée définie comme suit :

$$L_{EC}(p,y) = -\frac{1}{n}\sum_{i=1}^{n}[yln(p) + (1-y)ln(1-p)] \tag{A I-19}$$

L'algorithme de la rétroprogation utilisé dans la descente du gradient s'appuie sur une dérivation en chaîne pour propager les erreurs de la couche de sortie à la première couche cachée. De ce fait le calcul du gradient d'un neurone de la couche cachée revient à :

$$\frac{\partial L_{CE}(\sigma(x),y)}{\partial x} = \frac{\partial L_{CE}(\sigma(x),y)}{\partial \sigma(x)}\frac{\partial \sigma(x)}{x} \tag{A I-20}$$

La dérivée du second terme est 1.10. Celle du premier terme est :

$$\frac{\partial L_{CE}(\sigma(x),y)}{\partial \sigma(x)} = -\left(\frac{y}{\sigma(x)} - \frac{1-y}{1-\sigma(x)}\right) \tag{A I-21}$$

À partir des équations 1.10 et A I-21 :

$$\frac{\partial L_{CE}(\sigma(x),y)}{\partial x} = y - \sigma(x) \tag{A I-22}$$

Cette soustraction est intéressante car plus celle-ci est grande et plus la prédiction est erronée.

3. Relation entre l'entropie croisée et Kullback-Leibler

Nous pouvons montrer le lien qui existe entre les deux en démarrant de la définition même du Kullback-Leibler :

$$D_{\mathrm{KL}}(p\,|q) = \sum_i p_i \log \frac{p_i}{q_i}. \tag{A I-23}$$

$$D_{\mathrm{KL}}(p\,|q) = \sum_i (-p_i \log q_i + p_i \log p_i) \tag{A I-24}$$

$$D_{\mathrm{KL}}(p\,|q) = -\sum_i p_i \log q_i + \sum_i p_i \log p_i \tag{A I-25}$$

$$D_{\mathrm{KL}}(p\,|q) = -\sum_i p_i \log q_i - \sum_i p_i \log \frac{1}{p_i} \tag{A I-26}$$

Nous notons $H(p) = \sum_i p_i \log \frac{1}{p_i}$ qui est l'entropie de la distribution p.

$$D_{\mathrm{KL}}(p\,|q) = -\sum_i p_i \log q_i - H(p) \tag{A I-27}$$

$$D_{\mathrm{KL}}(p\,|q) = \sum_i p_i \log \frac{1}{q_i} - H(p) \tag{A I-28}$$

Nous arrivons donc à la définition de l'entropie croisée qui représente le premier terme de l'équation. Nous notons $H(p,q) = \sum_i p_i \log \frac{1}{q_i}$. Nous pouvons ainsi écrire l'entropie croisée par rapport à la divergence du Kullback-Leibler :

$$H(p,q) = H(p) + D_{KL}(p|q). \tag{A I-29}$$

BIBLIOGRAPHIE

Atkinson, K. (1989). *An Introduction to Numerical Analysis*. Wiley.

Badrinarayanan, V., Kendall, A. & Cipolla, R. (2015). SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *CoRR*, abs/1511.00561.

Belavy, D., Bock, O., Börst, H., Armbrecht, G., Gast, U., Degner, C., Beller, G., Soll, H., Salanova, M., Habazettl, H., Heer, M., Haan, A., Stegeman, D., Cerretelli, P., Blottner, D., Rittweger, J., Gelfi, C., Kornak, U. & Felsenberg, D. (2010). The 2nd Berlin BedRest study : Protocol and implementation. *Journal of Musculoskeletal and Neuronal Interactins*, 10, 207-219.

Ben Ayed, I., Punithakumar, K., Garvin, G. J., Romano, W. & Li, S. (2011). Graph Cuts with Invariant Object-Interaction Priors : Application to Intervertebral Disc Segmentation. *IPMI*, 6801(Lecture Notes in Computer Science), 221-232.

Birchall, D., Hughes, D., Gregson, B. & Williamson, B. (2005). Demonstration of vertebral and disc mechanical torsion in adolescent idiopathic scoliosis using three-dimensional MR imaging. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*, 14(2), 123—129.

Botev, Z. I., Grotowski, J. F., Kroese, D. P. et al. (2010). Kernel density estimation via diffusion. *The annals of Statistics*, 38(5), 2916–2957.

Boykov, Y. & Funka-Lea, G. (2006). Graph cuts and efficient ND image segmentation. *International journal of computer vision*, 70(2), 109–131.

Cannon, R. L., Dave, J. V. & Bezdek, J. C. (1986). Efficient Implementation of the Fuzzy c-Means Clustering Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(2), 248–255. doi : 10.1109/TPAMI.1986.4767778.

Cauchy, A.-L. (1847). Méthode générale pour la résolution des systèmes d'équations simultanées. *Compte Rendu des Séances de L'Académie des Sciences XXV*, Série A(25), 536–538.

Chen, C., Belavy, D., Yu, W., Chu, C., Armbrecht, G., Bansmann, M., Felsenberg, D. & Zheng, G. (2015). Localization and Segmentation of 3D Intervertebral Discs in MR Images by Data Driven Estimation. *IEEE Trans. Med. Imaging*, 34(8), 1719–1729.

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 Augmented Reality*, pp. 375–382.

Chevalier, J. (2017, Juillet, 10). Nos neurones se synchronisent-ils ? [Format]. Repéré à https://images.math.cnrs.fr/Nos-neurones-se-synchronisent-ils.html.

Chevrefils, C. (2010). *Construction d'un modèle pré-opératoire 3D du rachis pour la navigation en thoracoscopie*. Thèse de doctorat, Polytechnique Montréal.

Chevrefils, C., Cheriet, F., Grimard, G. & Aubin, C. (2007). Watershed Segmentation of Intervertebral Disk and Spinal Canal from MRI Images. *ICIAR*, 4633, 1017–1027.

Chevrefils, C., Cheriet, F., Aubin, C. & Grimard, G. (2009). Texture Analysis for Automatic Segmentation of Intervertebral Disks of Scoliotic Spines From MR Images. *IEEE Trans. Information Technology in Biomedicine*, 13(4), 608–620.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. (2016). 3D U-Net : Learning Dense Volumetric Segmentation from Sparse Annotation. 9901, 424–432.

Ciresan, D. C., Giusti, A., Gambardella, L. M. & Schmidhuber, J. (2012). Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. *NIPS*, pp. 2852–2860.

Clevert, D., Unterthiner, T. & Hochreiter, S. (2015). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *CoRR*, abs/1511.07289.

Cobb, J. (1948). Outline for the study of scoliosis. *Instr Course Lect AAOS*, 5, 261–275.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4), 303–314.

Dolz, J., Desrosiers, C. & Ayed, I. B. (2018). IVD-Net : Intervertebral disc localization and segmentation in MRI with a multi-modal UNet. *CoRR*, abs/1811.08305.

Duchi, J., Hazan, E. & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul), 2121–2159.

Erika. (2018, October, 02). S-Curve vs C-Curve Scoliosis Treatment [Format]. Repéré à https://www.scoliosissos.com/news/post/s-curve-vs-c-curve-scoliosis-treatment.

Freund, Y., Schapire, R. & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780), 1612.

Georges, D. (2019, Janvier, 28). Terminologie médicale [Format]. Repéré à http://www.bio-top.net/Terminologie/V/vertebro.htm.

Ghosh, S. & Chaudhary, V. (2014). Supervised methods for detection and segmentation of tissues in clinical lumbar MRI. *Comp. Med. Imag. and Graph.*, 38(7), 639–649.

Glorot, X. & Bengio, Y. (2010). *Understanding the difficulty of training deep feedforward neural networks*. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. MIT Press.

Guerroumi, N., Playout, C., Laporte, C. & Cheriet, F. (2019). Automatic Segmentation of the Scoliotic Spine from MR Images. *16th IEEE International Symposium on Biomedical Imaging, ISBI , Venice, Italy. accepté pour fins de publication*.

Guo, X., Chau, W.-W., Chan, Y.-L. & Cheng, J.-Y. (2003). Relative anterior spinal overgrowth in adolescent idiopathic scoliosis : results of disproportionate endochondral-membranous bone growth. *The Journal of bone and joint surgery. British volume*, 85(7), 1026–1031.

Hathaway, R. J. & Bezdek, J. C. (1988). Recent convergence results for the fuzzy c-means clustering algorithms. *Journal of Classification*, 5(2), 237–247. doi : 10.1007/bf01897166.

He, K., Zhang, X., Ren, S. & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. CVPR. IEEE Computer Society.

Hille, G., Saalfeld, S., Serowy, S. & Tönnies, K. (2018). Vertebral body segmentation in wide range clinical routine spine MRI data. *Computer Methods and Programs in Biomedicine*, 155, 93 – 99.

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580.

Hinton, G. E., Osindero, S. & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527–1554.

Hu, J., Shen, L. & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141.

Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 2261–2269.

Huang, S., Chu, Y., Lai, S. & Novak, C. L. (2009). Learning-Based Vertebra Detection and Iterative Normalized-Cut Segmentation for Spinal MRI. *IEEE Trans. Med. Imaging*, 28(10), 1595–1605.

Jen-Tang, L., Pedemonte, S., Bizzo, B., Doyle, S., Andriole, K. P., Michalski, M. H., Gonzalez, R. G. & Pomerantz, S. R. (2018). DeepSPINE : Automated Lumbar Vertebral Segmentation, Disc-level Designation and Spinal Stenosis Grading Using Deep Learning. *CoRR*, abs/1807.10215.

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

Keenan, B. E. (2015). *Medical imaging and Biomechanical analysis of scoliosis progression in the growing adolescent spine*. (Thèse de doctorat, Queensland University of Technology).

Kelm, B. M., Wels, M., Zhou, S. K., Seifert, S., Sühling, M., Zheng, Y. & Comaniciu, D. (2013). Spine detection in CT and MR using iterated marginal space learning. *Medical Image Analysis*, 17(8), 1283–1292.

Kim, S., Bae, W. C., Masuda, K., Chung, C. B. & Hwang, D. (2018). Fine-Grain Segmentation of the Intervertebral Discs from MR Spine Images Using Deep Convolutional Neural Networks : BSU-Net. *Applied Sciences*, 8(9).

Kingma, D. P. & Ba, J. (2015). Adam : A Method for Stochastic Optimization. *CoRR*, abs/1412.6980.

Korez, R., Likar, B., Pernuš, F. & Vrtovec, T. (2016). Model-based segmentation of vertebral bodies from MR images with 3D CNNs. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 433–441.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6), 84–90.

Kullback, S. & Leibler, R. A. (1951). On Information and Sufficiency. *Ann. Math. Statist.*, 22, 79-86.

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

Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE.

LeCun, Y. L., Kanter, I. & Solla, S. A. (1991). Eigenvalues of covariance matrices : Application to neural-network learning. *Phys. Rev. Lett.*, 66, 2396–2399. doi : 10.1103/PhysRevLett.66.2396.

Li, X., Dou, Q., Chen, H., Fu, C., Qi, X., Belavy, D., Armbrecht, G., Felsenberg, D., Zheng, G. & Heng, P. (2018). 3D multi-scale FCN with random modality voxel dropout learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR Images. *Medical Image Analysis*, 45, 41–54.

Long, J., Shelhamer, E. & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. CVPR. IEEE Computer Society.

McCulloch, W. S. & Pitts, W. (1988). Neurocomputing : Foundations of Research. 15–27.

Michopoulou, S. K., Costaridou, L., Panagiotopoulos, E. E., Speller, R. D., Panayiotakis, G. & Todd-Pokropek, A. (2009). Atlas-Based Segmentation of Degenerated Lumbar Intervertebral Discs From MR Images of the Spine. *IEEE Trans. Biomed. Engineering*, 56(9), 2225–2231.

Nair, V. & Hinton, G. E. (2010). *Rectified Linear Units Improve Restricted Boltzmann Machines*. ICML. Omnipress.

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.

Oktay, A. B. & Akgul, Y. S. (2013). Simultaneous Localization of Lumbar Vertebrae and Intervertebral Discs With SVM-Based MRF. *IEEE Trans. Biomed. Engineering*, 60(9), 2375–2383.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62–66.

Peng, Z., Zhong, J., Wee, W. & Lee, J.-h. (2006). Automated vertebra detection and segmentation from the whole spine MR images. *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 2527–2530.

Perona, P. & Malik, J. (1990). Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7), 629–639.

Pfandler, M., Lazarovici, M., Stefan, P., Wucherer, P. & Weigl, M. (2017). Virtual reality-based simulators for spine surgery : A systematic review. *The Spine Journal*, 17.

Playout, C. (2018). *Systeme d'apprentissage multitâche dédié à la segmentation des lésions sombres et claires de la rétine dans les images de fond d'oeil.* Thèse de doctorat, Polytechnique Montréal.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1), 145–151.

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-Net : Convolutional Networks for Biomedical Image Segmentation. *MICCAI (3)*, 9351(Lecture Notes in Computer Science), 234–241.

Rosenblatt, F. (1958). The Perceptron : A Probabilistic Model for Information Storage and Organization in The Brain. *Psychological Review*, 65–386.

Roy, A. G., Navab, N. & Wachinger, C. (2018). Concurrent Spatial and Channel 'Squeeze & Excitation' in Fully Convolutional Networks. *MICCAI (1)*, 11070(Lecture Notes in Computer Science), 421–429.

Rubinstein, R. Y. & Kroese, D. P. (2004). *The Cross Entropy Method : A Unified Approach To Combinatorial Optimization, Monte-carlo Simulation (Information Science and Statistics)*. Berlin, Heidelberg : Springer-Verlag.

Rumelhart, D. E., Hinton, G. E., Williams, R. J. et al. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, 5(3), 1.

Shi, J. & Malik, J. (1997). *Normalized Cuts and Image Segmentation*. CVPR. IEEE Computer Society.

Shlens, J. (2014). Notes on Kullback-Leibler Divergence and Likelihood. *CoRR*, abs/1404.2000.

Simonyan, K. & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556.

Sled, J., Zijdenbos, A. & Evans, C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. I. *E.E.E. Transactions on Medical Imaging*, 17, 87-97.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.*, 5, 1–34.

Sutskever, I., Martens, J., Dahl, G. E. & Hinton, G. E. (2013). On the importance of initialization and momentum in deep learning. *ICML (3)*, 28(1139-1147), 5.

Suzani, A., Rasoulian, A., Fels, S., Rohling, R. N. & Abolmaesumi, P. (2014). Semi-automatic segmentation of vertebral bodies in volumetric MR images using a statistical shape+pose model. *Medical Imaging : Image-Guided Procedures*, 9036(SPIE Proceedings), 90360P.

Suzani, A., Rasoulian, A., Seitel, A., Fels, S., Rohling, R. N. & Abolmaesumi, P. (2015). Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images. *Medical Imaging : Image-Guided Procedures, SPIE*, 9415, 941514.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. & Gee, J. C. (2010). N4ITK : Improved N3 Bias Correction. *IEEE Trans. Med. Imaging*, 29(6), 1310-1320.

Van Rijsbergen, C. (1979). Information retrieval : theory and practice. *Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems*, pp. 1–14.

Vrtovec, T., Likar, B. & Pernuš, F. (2005). Automated curved planar reformation of 3D spine images. *Physics in Medicine Biology*, 50(19), 4527.

Zeiler, M. D. (2012). ADADELTA : An Adaptive Learning Rate Method. *CoRR*, abs/1212.5701.

Zeiler, M. D. & Fergus, R. (2013). Visualizing and Understanding Convolutional Networks. *CoRR*, abs/1311.2901.

Zhang, K., Oommen, B. & Lee, W. (1996). Numerical similarity and dissimilarity measures between two trees. *IEEE Transactions on Computers*, 45(12), 1426-1434.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M. & Comaniciu, D. (2008). Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. *IEEE Trans. Med. Imaging*, 27(11), 1668–1681.