



HAL
open science

Apprentissage profond et relations structurelles pour l'analyse d'images : application à l'étude de la lésion cérébrale précoce chez l'enfant après AVC néonatal

Patty Coupeau

► **To cite this version:**

Patty Coupeau. Apprentissage profond et relations structurelles pour l'analyse d'images : application à l'étude de la lésion cérébrale précoce chez l'enfant après AVC néonatal. Médecine humaine et pathologie. Université d'Angers, 2023. Français. NNT : 2023ANGE0008 . tel-04232019

HAL Id: tel-04232019

<https://theses.hal.science/tel-04232019v1>

Submitted on 7 Oct 2023

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

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

THESE DE DOCTORAT

DE
L'UNIVERSITÉ D'ANGERS

SOUS LE SCEAU DE
LA COMUE ANGERS – LE MANS

ECOLE DOCTORALE N° 641

*Mathématiques et Sciences et Technologies du numérique,
de l'Information et de la Communication*

Spécialité : *Signal, Image, Vision*

Par

Patty Coupeau

**Apprentissage profond et relations structurelles pour l'analyse
d'images : application à l'étude de la lésion cérébrale précoce chez
l'enfant après AVC néonatal**

Thèse présentée et soutenue à l'Université d'Angers, le 19/06/2023

Unité de recherche : LARIS

Rapporteurs avant soutenance :

François Rousseau Professeur – IMT Atlantique
Olivier Lézoray Professeur des universités – Université de Caen-Normandie

Composition du Jury :

Présidente : Jessica Dubois Directrice de recherche – Inserm, Unité NeuroDiderot, UMR1141

Examineurs : Harold Mouchère Professeur des universités – Nantes Université
 Jessica Dubois Directrice de recherche – Inserm, Unité NeuroDiderot, UMR1141
 Isabelle Bloch Professeur des universités – Sorbonne Université

Dir. de thèse : Jean-Baptiste Fasquel Professeur des universités – Université d'Angers
Co-dir. de thèse : Mickaël Dinomais Professeur des universités-praticien hospitalier – Université d'Angers

Invité(s)

Sylvain Brochard Professeur des universités-praticien hospitalier – Université Bretagne Occidentale

À mon papa, parti rejoindre les sommets.

REMERCIEMENTS

J'aimerais remercier toutes les personnes qui ont contribué au succès de ma thèse et qui m'ont fait vivre durant ces 3 années une belle aventure scientifique mais aussi humaine.

Naturellement, je pense tout d'abord à mes directeurs de thèse Jean-Baptiste Fasquel et Mickaël Dinomais. Avec Jean-Baptiste l'histoire a commencé il y a 5 ans, durant mes études d'ingénieure, lorsqu'il m'a fait découvrir la recherche dans le domaine médical. Merci de m'avoir ouvert les portes de ce monde scientifique si riche et de m'avoir toujours guidée à travers de précieux conseils. La rencontre avec Mickaël, bien que plus tardive, m'a apporté de nombreuses connaissances sur les neurosciences et m'a permis, à maintes reprises, de prendre du recul sur mon travail de chercheuse. Je vous remercie tous les deux pour votre encadrement, votre bienveillance et votre bonne humeur.

Je tiens à remercier Jessica Dubois et Julien Lefèvre, membres de mon CSI, qui m'ont accompagnée au cours de ces 3 années en apportant un regard extérieur sur mes travaux tout en étant à mon écoute.

Je remercie sincèrement François Rousseau et Olivier Lézoray d'avoir accepté de rapporter cette thèse ainsi que tous les membres de mon jury : Harold Mouchère, Isabelle Bloch, Jessica Dubois à nouveau et Sylvain Brochard.

La thèse serait bien triste si elle n'était pas partagée. Pour cette raison, je souhaite remercier tous les membres du LARIS et tous mes collègues doctorants avec lesquels nous avons passé de bons moments de sciences, de galères parfois mais aussi et surtout de rires.

Je n'oublie pas ma famille et mes amis qui m'ont soutenue durant cette aventure. Vous avez su me faire voir le positif même dans les périodes les plus difficiles. Je pense particulièrement à Louis qui, par sa présence à mes côtés, a rendu tout plus simple. Je ne vous remercierai jamais assez.

TABLE DES MATIÈRES

Introduction	11
I Apprentissage profond et relations structurelles pour l'analyse d'images	15
1 Apprentissage profond et connaissances structurelles	16
1.1 Introduction	16
1.2 Apprentissage profond	16
1.2.1 Introduction	16
1.2.2 Réseaux de neurones convolutifs	17
1.2.3 Exemples d'architectures pour la segmentation sémantique d'images	19
1.2.4 Le défi du nombre de données	21
1.3 Connaissances structurelles	24
1.3.1 Définition	25
1.3.2 Modélisation par un graphe	26
1.4 Exploitation des graphes par réseaux de neurones pour l'analyse d'images .	28
1.4.1 Introduction	28
1.4.2 Les réseaux de neurones sur graphes	30
1.4.3 La convolution sur graphe	32
1.5 Conclusion	37
2 Utilisation des réseaux de neurones sur graphes pour la segmentation sémantique d'images	39
2.1 Introduction	39
2.2 CNNs et GNNs pour la segmentation sémantique	40
2.3 Méthode proposée	41
2.3.1 De l'image au graphe	43
2.3.2 Architecture du GNN	44
2.4 Expériences	45

2.4.1	Protocole expérimental	45
2.4.2	FASSEG	48
2.4.2.1	Jeu de données et architectures CNNs	48
2.4.2.2	Construction des graphes	49
2.4.2.3	Résultats	49
2.4.3	IBSR	52
2.4.3.1	Jeu de données et architectures CNNs	52
2.4.3.2	Construction des graphes	53
2.4.3.3	Résultats	53
2.4.4	Influence de la topologie des graphes	58
2.5	Discussion	62
2.6	Conclusion	64

II Application à l'étude de la lésion cérébrale précoce chez l'enfant après AVC néonatal 67

Introduction à l'étude de la lésion cérébrale précoce 68

3 Apprentissage profond pour l'étude du porcelet : modèle animal de développement neurologique précoce 75

3.1	Introduction	75
3.2	Méthode proposée	78
3.2.1	Architecture du réseau U-Net	80
3.2.2	Stratégie d'extraction de patches 3D	80
3.2.2.1	Entraînement	81
3.2.2.2	Inférence	83
3.2.3	Apprentissage par transfert et analyse longitudinale	83
3.3	Expériences	85
3.3.1	Base de données	85
3.3.2	Protocole expérimental	86
3.3.3	Résultats	89
3.3.3.1	Taille et distribution des patches d'entraînement	89
3.3.3.2	Post-traitement : régularisation spatiale et dilatation	91
3.3.3.3	Comparaison avec l'état de l'art	92

3.3.3.4	Apprentissage par transfert pour l'étude longitudinale . . .	95
3.4	Discussion	99
3.5	Conclusion	101
4	Segmentation des noyaux gris centraux de l'enfant après AVC néonatal : exploitation des connaissances structurelles	103
4.1	Introduction	103
4.2	Adaptation de la méthode de segmentation sémantique	105
4.3	Expériences	108
4.3.1	Base de données	108
4.3.2	Protocole expérimental	109
4.3.3	CNNs et construction des graphes	110
4.3.4	Résultats	111
4.4	Discussion	118
4.5	Conclusion	120
5	Etude de la motricité de l'enfant après AVC néonatal : GNNs et organisation macrostructurale des noyaux gris centraux	123
5.1	Introduction	123
5.2	Méthode	126
5.3	Base de données	128
5.4	Détection de la paralysie cérébrale chez les enfants après AVC néonatal . . .	130
5.4.1	Architecture du GNN considéré	130
5.4.2	Expériences	131
5.4.2.1	Protocole expérimental	132
5.4.2.2	Résultats	133
5.5	Estimation de la motricité manuelle des enfants après AVC néonatal	135
5.5.1	Architecture du GNN considéré	135
5.5.2	Expériences	136
5.5.2.1	Protocole expérimental	136
5.5.2.2	Résultats	138
5.6	Discussion	143
5.7	Conclusion	147

TABLE DES MATIÈRES

Conclusion, limites et perspectives	149
Conclusion	149
Limites et perspectives	152
A Annexes	157
A.1 Annexes du chapitre 2	157
A.2 Annexes du chapitre 3	158
A.3 Annexes du chapitre 5	160
Bibliographie	161

INTRODUCTION

Dans son livre paru en 2023, le médecin oncologue Jean-Emmanuel Bibault déclare "Aujourd'hui, je suis convaincu que l'intelligence artificielle bouleversera profondément la façon d'exercer notre métier et améliorera le devenir des patients [...] Encore faut-il développer nos outils de la bonne façon" [1]. Il faut dire que ce domaine d'étude a connu un essor considérable ces dernières décennies, notamment en médecine. Marvin Lee Minsky, cofondateur du groupe d'intelligence artificielle du Massachusetts Institute of Technology (MIT) définit l'intelligence artificielle (IA) comme "la construction de programmes informatiques qui s'adonnent à des tâches [...] demandant des processus mentaux de haut niveau tels que l'apprentissage conceptuel, l'organisation de la mémoire et le raisonnement critique". Par abus de langage, on regroupe sous le terme d'IA toutes les techniques liées à l'apprentissage machine dont fait partie l'apprentissage profond qui a permis les progrès les plus récents [2]. En effet, la mise à disposition de volumes de données toujours plus importants ainsi que l'augmentation des puissances de calcul ont favorisé le développement de l'apprentissage profond. Dans le cadre de cette thèse, nous nous intéressons à l'usage de ces techniques basées sur des réseaux de neurones profonds pour l'analyse d'images. Les images sont des données prisées du fait de la richesse des informations qu'elles concentrent et de la grande diversité de leurs domaines d'application parmi lesquels on peut citer la télédétection [3], l'imagerie satellite [4], la robotique [5] ou encore évidemment l'imagerie médicale [6]. Apparues, pour les plus anciennes, au tournant du XX^e., les techniques d'imagerie médicale ont révolutionné la médecine en permettant de visualiser les organismes (humains comme animaux) et d'obtenir des informations sur l'anatomie des organes (imagerie dite structurelle : ex. radiologie, scanner, imagerie par résonance magnétique (IRM), échographies) ou sur leur fonctionnement (imagerie dite fonctionnelle : ex. IRM fonctionnelle, électroencéphalographie, tomographie par émission de positons). Parmi cette grande diversité de techniques d'imagerie, nous nous focaliserons dans nos travaux sur des données IRM structurelles.

Dans le domaine de l'imagerie, les algorithmes d'apprentissage profond se sont montrés performants pour diverses tâches d'analyse telles que la classification, la détection d'objet ou la segmentation sémantique d'images [7]. Les réseaux profonds actuels utilisent,

pour faire leur prédiction, les informations dites de "bas-niveau" de l'image (ex.intensités) portées par les pixels et leur voisinage proche. Pourtant, les images fournissent une information bien plus vaste, notamment à travers les relations observables entre les objets (ou régions) présents sur celles-ci (ex. relations spatiales, relations photométriques [8]). Ces informations observables de "plus haut-niveau", que l'on appellera connaissances structurelles tout au long du manuscrit, sont omniprésentes, compréhensibles et semblent, la plupart du temps, plus facile à fournir qu'une annotation à l'échelle du pixel. Ainsi, on peut se demander quel pourrait être l'apport des connaissances structurelles de haut-niveau, en plus des techniques actuelles d'apprentissage profond, pour l'analyse d'images ? Plus précisément et de manière complémentaire, nous nous intéressons à la question de l'apport de ces connaissances structurelles et de leur intégration en s'appuyant sur des réseaux de neurones sur graphes (GNNs) [9]. En effet, comme nous le verrons dans ce manuscrit, les graphes sont largement utilisés pour manipuler ce type d'informations et le récent essor des GNNs offre un cadre prometteur à l'intégration de ces connaissances structurelles.

Afin d'étudier la motricité de l'enfant et plus précisément la réorganisation cérébrale de cette motricité suite à une lésion cérébrale précoce, l'imagerie médicale est essentielle car elle permet, entre autres, de réaliser des analyses morphostructurelles du cerveau. En effet, nous émettons l'hypothèse que la lésion cérébrale précoce peut altérer l'organisation structurelle spatiale de certaines structures anatomiques du cerveau de l'enfant. Ainsi, compte tenu des remarques précédentes relatives aux informations structurelles en traitement d'images et de l'usage de l'intelligence artificielle en médecine souligné par Jean-Emmanuel Bibault [1], on peut se poser la question suivante : dans quelle mesure la combinaison de l'apprentissage profond et des informations structurelles de haut-niveau peut-elle aider dans l'étude de la lésion cérébrale précoce ?

Les travaux présentés dans ce manuscrit, organisés en deux parties, tentent d'apporter des éléments de réponse à ces questions.

La partie I se focalisera sur notre première interrogation en présentant quelques possibilités combinant apprentissage profond et relations structurelles pour l'analyse d'images de manière générale. Cette partie présentera une première contribution à la segmentation sémantique d'images. Elle se divisera en deux chapitres.

Dans le chapitre 1 sera introduit le contexte scientifique de cette thèse relatif à l'apprentissage profond et aux connaissances structurelles. Après des rappels sur l'utilisation de l'apprentissage profond en analyse d'images, nous introduirons plus précisément la notion de connaissances structurelles et leur modélisation sous forme de graphes. Nous

présenterons par la suite diverses stratégies permettant d'exploiter ces informations sous forme de graphes avant de décrire plus en détails la technique choisie consistant à utiliser des réseaux de neurones sur graphes ou GNNs.

Dans le chapitre 2, nous proposerons une méthode de post-traitement, postérieure à l'utilisation d'un réseau de neurones profond classique pour l'analyse d'images, destinée à améliorer les performances de segmentation sémantique en exploitant les connaissances structurelles par un GNN. Nous évaluerons l'approche sur deux jeux de données publiques et avec différents réseaux de neurones profonds afin d'estimer la capacité de la méthode à corriger des erreurs de segmentation (incohérences structurelles) commises par un réseau profond, notamment lorsque peu de données d'entraînement sont disponibles. Nous étudierons dans ce chapitre certains paramètres influençant le comportement du modèle GNN tels que la nature de l'opérateur de convolution sur graphe ou la topologie des graphes.

La partie II du manuscrit aura pour objectif d'apporter quelques réponses à notre deuxième interrogation relative à l'apport de l'apprentissage profond et des connaissances structurelles pour l'étude de la lésion cérébrale précoce chez l'enfant après AVC néonatal. Nous entamerons cette partie par une introduction sur le contexte médical de l'étude, visant à mieux comprendre les notions de lésion cérébrale précoce, d'AVC néonatal et de paralysie cérébrale (troubles moteurs permanents) pouvant en découler. Nous verrons notamment que différentes études (ex. sur l'enfant ayant eu un AVC néonatal, sur des modèles animaux) sont nécessaires pour comprendre cette lésion et aider les cliniciens à proposer des interventions thérapeutiques adaptées et plus ciblées. Cette seconde partie se divisera en trois chapitres. Le premier sera relatif au modèle du porcelet pour l'étude du développement neurologique précoce tandis que les deux autres s'intéresseront à des structures cérébrales liées aux fonctions sensori-motrices des enfants après AVC néonatal : les noyaux gris centraux.

Ainsi, dans le chapitre 3, nous proposerons une méthode d'extraction automatique du cerveau du porcelet basée sur une architecture U-Net 3D manipulant des patches extraits des IRMs cérébrales selon une stratégie spécifique. Nous présenterons également une approche d'apprentissage par transfert afin de permettre à notre méthode d'assurer un suivi longitudinal du cerveau du porcelet. Nous comparerons notre solution à d'autres outils et méthodes d'extraction de cerveau adaptés à l'animal ou à l'humain. Nous espérons que cette contribution pourra faciliter l'analyse du modèle animal porcin, adapté à l'étude de la lésion cérébrale précoce.

Dans le chapitre 4, nous nous attaquerons au problème de la segmentation des noyaux

gris centraux sur IRM structurelle d'enfants avec lésion cérébrale précoce. Nous verrons que les réseaux de neurones profonds dédiés à cette tâche ont tendance à commettre des erreurs dans ces conditions. Nous proposerons d'appliquer la méthode de post-traitement du chapitre 2, basée sur un GNN pour exploiter les relations spatiales entre les structures, afin d'améliorer la segmentation sémantique de ces structures cérébrales profondes. Nous évaluerons la méthode par hémisphère pour distinguer les apports du côté lésé et du côté contra-lésé (la lésion étudiée étant unilatérale) et nous comparerons nos résultats à ceux obtenus avec d'autres méthodes.

Enfin, dans le chapitre 5, nous chercherons à savoir si l'organisation macrostructurelle des noyaux gris centraux des enfants après AVC néonatal permet de mieux comprendre leur motricité, notamment manuelle, et de prévoir ou de détecter le développement d'une paralysie cérébrale unilatérale se caractérisant par des troubles moteurs permanents. Nous proposerons une modélisation sous forme de graphes de l'organisation structurelle des noyaux gris centraux afin d'utiliser des GNNs pour deux applications médicales : la détection de la paralysie cérébrale et l'estimation d'un score de motricité manuelle chez des enfants après AVC néonatal. Nous évaluerons la précision de la prédiction de notre modèle et montrerons l'existence d'une corrélation entre cette prédiction et la performance motrice uni-manuelle réelle des deux mains des enfants.

En fin de manuscrit, nous concluons sur les travaux présentés en mettant en évidence les limites et perspectives associées. Nous verrons que cette thèse n'est que le prémisses d'un long travail de recherche, à la fois en traitement d'images et en neurosciences, autour de l'étude de la lésion cérébrale précoce chez l'enfant après AVC néonatal.

PREMIÈRE PARTIE

Apprentissage profond et relations structurelles pour l'analyse d'images

APPRENTISSAGE PROFOND ET CONNAISSANCES STRUCTURELLES

1.1 Introduction

Ces travaux de thèse s'intéressent à l'apport des relations structurelles, combinées aux méthodes d'apprentissage profond existantes, pour l'analyse d'images. Les informations structurelles considérées incluent les relations observables, à « haut-niveau », entre les régions d'une image.

Dans ce chapitre, nous débutons par un rappel de notions relatives à l'apprentissage profond. Nous poursuivons par une présentation du concept de connaissances structurelles et de leur représentation sous forme de graphes. Nous achevons par une description des réseaux de neurones sur graphe, exploitant les connaissances structurelles, pour l'analyse d'images.

1.2 Apprentissage profond

1.2.1 Introduction

Cette partie reprend les travaux de thèse de Jérémy Chopin [10] et se base sur le livre « Deep learning » de Goodfellow et al. [2].

L'apprentissage machine (*machine learning*) est un domaine d'étude de l'intelligence artificielle qui vise à donner aux machines la capacité d'apprendre à réaliser une tâche, à partir de données, et de s'améliorer au fil du temps. De nombreux modèles d'apprentissage machine ont vu le jour ces dernières années dans le but d'apprendre des tâches variées (régression, classification, clustering). La mise à disposition de volumes de données toujours plus importants ainsi que l'augmentation des puissances de calcul ont contribué, en partie, au développement de l'apprentissage profond [2]. L'apprentissage profond est

un sous-domaine de l'apprentissage machine basé sur des réseaux de neurones de grande taille avec de nombreuses couches. Les modèles d'apprentissage profond sont capables d'apprendre automatiquement des caractéristiques directement à partir des données, s'affranchissant ainsi de la nécessité d'identifier de manière ad-hoc des caractéristiques pertinentes. Ils sont ainsi bien adaptés à des tâches relatives au traitement d'images [7] telles que la classification ou la segmentation sémantique d'images qui nous intéresseront dans cette thèse.

Parmi les architectures les plus utilisées en apprentissage profond, les réseaux de neurones convolutifs (CNNs) s'avèrent très performants pour l'analyse d'images et notamment pour la segmentation d'images [11]. Dans les sections suivantes, nous faisons un rapide rappel sur les réseaux de neurones convolutifs avant de présenter certaines architectures performantes pour la segmentation sémantique d'images médicales et utilisées dans nos travaux. Nous évoquons enfin des solutions existantes pour faire face à un défi de l'apprentissage profond, à savoir le manque de données d'entraînement, particulièrement présent dans le domaine médical que nous aborderons par la suite.

1.2.2 Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs (CNNs) sont spécialisés dans le traitement de données avec une topologie en grille telles que les images (grille 2D de pixels, distribution régulière de voxels en 3D). Ces réseaux se distinguent par l'utilisation d'opérations de convolution dans au moins une de leurs couches. L'opérateur de convolution, décrit par l'eq 1.1 dans le cas d'une image 2D [2], s'applique en tout point (i, j) de l'image et fait intervenir deux arguments : une entrée (ex. une image 2D notée I) et un noyau (ex. un filtre 2D noté K). Sa sortie est appelée carte de caractéristiques (Figure 1.1).

$$(I * K)(i, j) = \sum_{m=1}^M \sum_{n=1}^N I(m, n)K(i - m, j - n) \quad (1.1)$$

Grâce à l'utilisation du noyau ou filtre K glissant sur l'image, les CNNs sont capables de capturer les dépendances spatiales entre pixels ou voxels (dans la limite du champ réceptif formé par le filtre), ce qui n'est pas le cas avec de simples perceptrons multicouches (MLP). Ce même filtre permet d'extraire des caractéristiques de l'image tout en réduisant le nombre de paramètres entraînaables (le filtre ne couvre qu'une sous-partie de l'image), rendant l'opérateur capable de traiter des données de tailles variables. Les premières couches de convolution permettent en général de capturer des caractéristiques

de bas niveau (couleurs, orientation de gradient, etc.). La mise en cascade de couches de convolution enrichit l'information extraite de l'image (caractéristiques de plus haut niveau, dans la limite du champ réceptif de K) et donne une représentation plus complète de celle-ci.

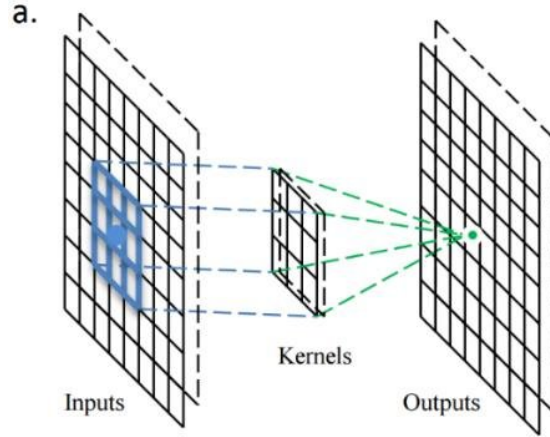


FIGURE 1.1 – Principe de l'opérateur de convolution (extrait de [12]).

Les opérations de convolution des CNNs, combinées à des opérateurs de sous échantillonnage (*pooling*) et à des couches entièrement connectées, comme présenté dans [10], constituent des réseaux profonds efficaces pour la classification d'images (prédiction d'une classe), la régression (prédiction d'une valeur réelle à partir d'une image) mais aussi pour la segmentation sémantique d'images que nous abordons dans les chapitres 2, 3 et 4 de cette thèse. Pour cette tâche, le réseau ne doit plus seulement prédire une valeur pour l'image mais attribuer une classe d'appartenance à chaque pixel (voxel) de l'image (ou à un sous groupe de pixels). Pour se faire, le CNN fournit en sortie un tenseur $S \in \mathbb{R}^{P \times C}$ avec P la dimension de l'image d'entrée ($P = I \times J$ est l'ensemble des pixels pour une image 2D) et C le nombre de classes considérées pour la segmentation. On associe ainsi à chaque pixel (voxel) $p \in P$ de l'image, le vecteur $S(p)$ de dimension C , où $S(p, c) \in [0, 1]$ est la probabilité que le pixel (voxel) p appartienne à la classe $c \in \{1, \dots, C\}$. La carte de segmentation \mathcal{L}^* affecte alors à chaque pixel (voxel) la classe la plus probable :

$$\forall p \in \{1, \dots, P\}, \mathcal{L}^*(p) = \arg \max_{c \in \{1, \dots, C\}} S(p, c). \quad (1.2)$$

De nombreuses architectures de réseaux de neurones pour la segmentation sémantique d'images existent. Nous présentons dans la section suivante des exemples d'architectures, adaptées à l'imagerie médicale, que nous considérerons dans la suite des travaux.

1.2.3 Exemples d'architectures pour la segmentation sémantique d'images

Parmi les modèles de réseaux de neurones pour la segmentation sémantique d'images, nous choisissons de détailler ici deux types d'architectures communément utilisées en imagerie médicale et que nous avons considérées durant nos expériences (chapitres 2, 3 et 4) : l'architecture U-Net [13] et les architectures d'attention telles que les *Vision Transformers*[14].

U-Net

L'architecture U-Net [13] est une extension des réseaux entièrement convolutifs (FCN) [15].

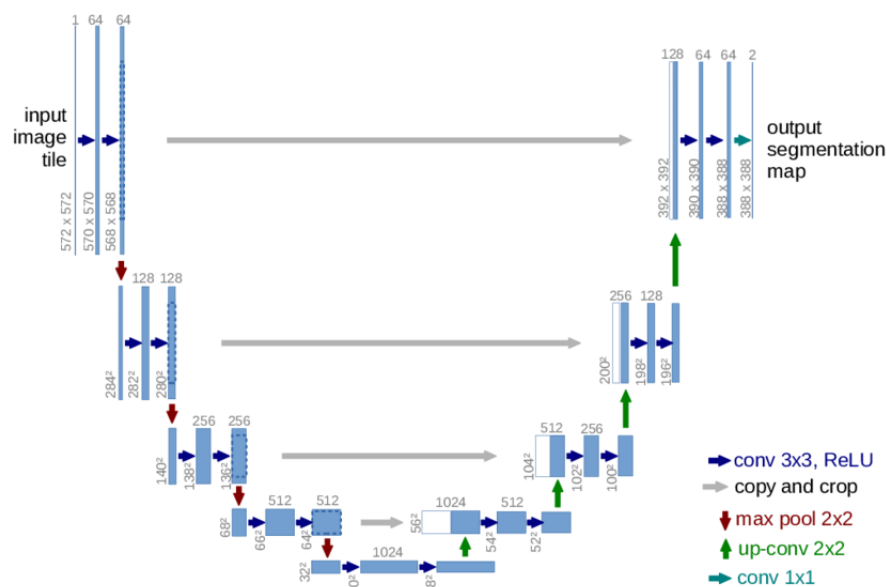


FIGURE 1.2 – Architecture U-Net (extrait de [13]).

Les architectures FCN sont constituées d'un encodeur au sein duquel l'image d'entrée est sous-échantillonnée à travers des couches successives de convolution et de pooling afin de construire une carte de caractéristiques à différentes échelles de l'image. Le réseau U-Net (Figure 1.2) possède également cet encodeur (partie gauche), mais il est suivi par un autre chemin similaire dénommé décodeur (partie droite), lui donnant sa forme en « U ». Dans le décodeur, les opérations de sous-échantillonnage sont remplacées par des opérations de sur-échantillonnage dans le but de revenir à l'échelle haute-résolution

de l'image. Des connexions sont faites à chaque échelle (flèches grises - Figure 1.2) de manière à combiner les informations de l'échelle courante et de l'échelle supérieure.

Les architectures d'attention

Les architectures d'attention ont suscité un intérêt croissant en apprentissage profond avec la proposition d'une architecture majoritairement composée de couches d'attention, nommée *Transformer* [16]. Ce modèle a montré son intérêt en traitement du langage naturel [16] avant de s'étendre plus récemment à la vision par ordinateur avec les *Vision Transformers (ViT)* [14, 17].

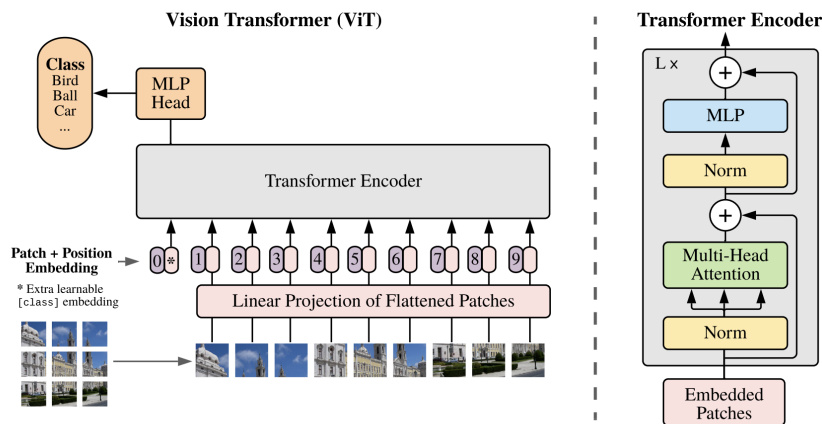


FIGURE 1.3 – Exemple d'architecture de *Vision Transformer* (extrait de [14]).

Les mécanismes d'attention sont apparus avec une volonté de focaliser le modèle sur des parties spécifiques de l'image, ayant un impact plus important dans la tâche à effectuer. Ainsi, dans le cas d'un problème de classification d'images par exemple, les pixels de l'arrière-plan contiennent des informations moins pertinentes pour la classification. Les modules d'attention produisent des cartes d'attention qui donnent un poids différent aux différentes zones de l'image. Dans le cas des *Vision Transformers*, plusieurs dizaines (voire centaines) de cartes d'attention par couche et par inférence sont générées, rendant l'interprétation des décisions prises par ces réseaux difficile [18]. Les ViT capturent des relations spatiales entre différentes parties de l'image. Pour se faire, ils calculent des relations entre pixels dans plusieurs patches de petite taille (ex. 16x16 pixels) découpés dans l'image (Figure 1.3). Les patches sont intégrés linéairement et combinés à des informations relatives

à la position (position embedding). La séquence de vecteurs résultante est envoyée à un *Transformer* (Figure 1.3 - droite).

Les *Transformers* permettent de capturer des relations globales et à plus grande échelle que les CNNs (limités par la taille du champ réceptif lors de la convolution), au prix d'un nombre de données d'entraînement nécessaire plus onéreux. Mais, le champ réceptif plus réduit des CNNs fournit une meilleure stabilité et précision que l'utilisation des ViT seuls. C'est pourquoi, ViT et couches convolutives sont souvent combinés pour bénéficier des avantages de chacun [19]. L'idée est d'intégrer de multiples couches d'attention entre les opérations de convolution afin de générer de nombreuses cartes d'attention. Les architectures hybrides de cette sorte obtiennent de bonnes performances en analyse d'images [20]. Dans nos expériences, décrites dans les chapitres 2 et 4, nous considérerons deux architectures combinant mécanismes d'attention et couches convolutives : un modèle hybride de ViT (UNETr) [21] et un modèle d'attention multi-échelle (MSGSE-Net) [22].

Pour rendre ces modèles d'apprentissage profond performants et généralisables, il est nécessaire de les entraîner sur un grand nombre de données. Malheureusement, en pratique, la quantité de données disponibles est limitée. Dans le domaine médical, que nous aborderons dans la Partie II de cette thèse, ceci est particulièrement vrai. Dans la prochaine section, nous présenterons des solutions existantes pour compenser le manque de données d'entraînement.

1.2.4 Le défi du nombre de données

De nombreuses techniques ont vu le jour afin de compenser le besoin d'une base de données large et représentative pour l'entraînement des réseaux de neurones profonds et pour éviter le sur-apprentissage des modèles. Parmi ces techniques, on peut citer l'augmentation de données [23], l'apprentissage par transfert [24] ou encore l'apprentissage à partir de peu d'exemples [25] utilisés en analyse d'images.

L'augmentation de données

L'augmentation des données d'image regroupe les techniques utilisées pour augmenter artificiellement la taille d'un groupe de données d'apprentissage en créant des versions modifiées d'images à partir des images d'apprentissage disponibles [23]. La technique utilisée doit avoir un sens par rapport au type d'analyse souhaitée et par conséquent influencer positivement les performances du modèle : l'application d'un grand nombre

d'augmentation n'est pas forcément la meilleure solution. Il existe différentes sortes de transformations (quelques exemples sont illustrés dans la Figure 1.4), notamment utilisées en imagerie médicale, et pouvant être classées en différentes catégories :

- **transformations géométriques** : rotation, réflexion, translation, cisaillement.
- **transformations diverses** : ajout de bruit, application d'un filtre, modification de contraste.

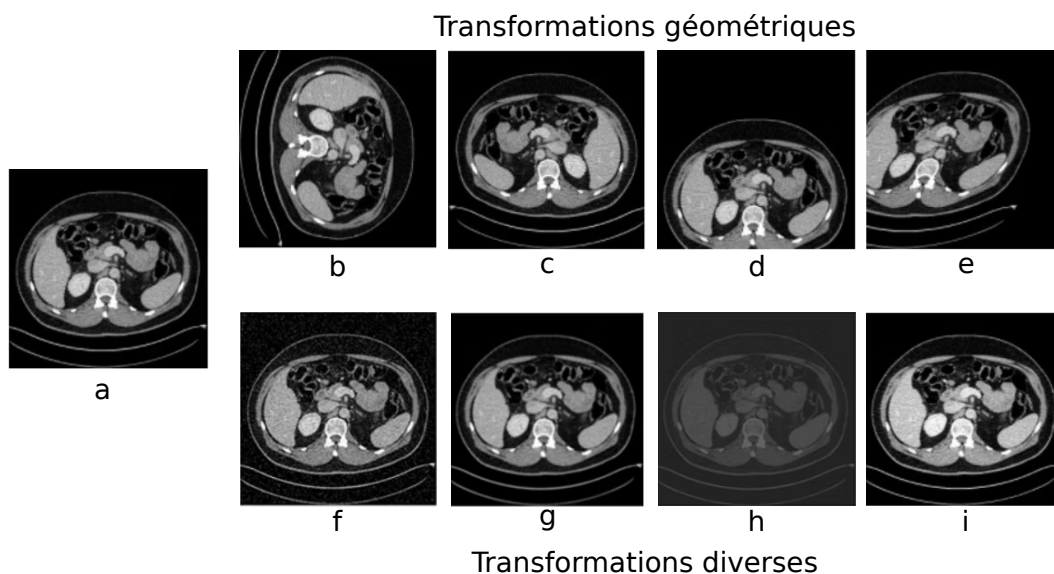


FIGURE 1.4 – Exemples de transformations utilisées pour l'augmentation de données¹. a : image originale, b : rotation de 90° , c : réflexion horizontale, d : translation selon y , e : cisaillement, f : image bruitée avec un bruit impulsionnel, g : image filtrée par filtre Gaussien, h : faible contraste, i : variation de luminosité.

Depuis quelques années, une nouvelle méthode d'augmentation de données se généralise à travers l'utilisation de réseaux antagonistes génératifs (GANs) [26].

L'apprentissage par transfert

L'apprentissage par transfert [24] désigne l'ensemble des méthodes qui permettent de transférer les connaissances acquises à partir de la résolution d'un problème donné (source) pour traiter un autre problème (cible) (Figure 1.5). En utilisant des modèles pré-entraînés comme point de départ, l'apprentissage par transfert permet de développer rapidement des modèles performants et de résoudre des problèmes complexes avec moins de données.

Différents types d'apprentissage par transfert peuvent être distingués :

1. www.imaios.com/fr/ressources/blog/ia-pour-l-imagerie-medecale-augmentation-des-donnees

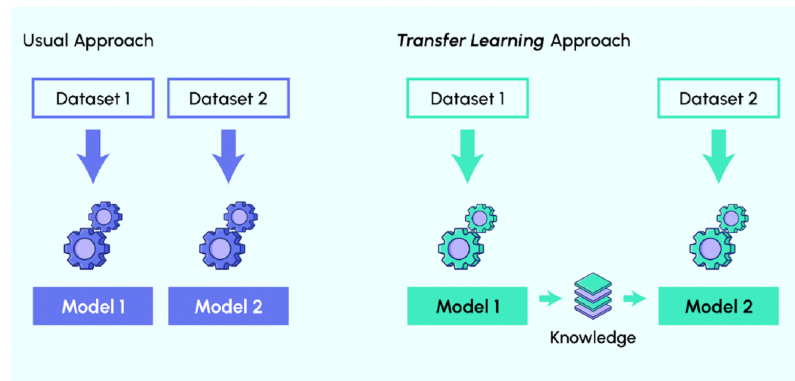


FIGURE 1.5 – Apprentissage traditionnel en apprentissage profond versus apprentissage par transfert (extrait de [27]).

- **L'apprentissage par transfert inductif** : Les domaines source et cible sont les mêmes (même données), mais les tâches source et cible sont différentes (bien que relativement proches). Le principe consiste à utiliser les modèles existants pour réduire le champ d'application des modèles possibles (biais de modèle). Par exemple, dans un contexte médical, il est possible d'utiliser un modèle entraîné pour la segmentation des tissus cérébraux sur IRM pour construire un modèle capable de segmenter des tumeurs cérébrales sur IRM [28].
- **L'apprentissage par transfert transductif** : Les tâches source et cible sont similaires, mais les domaines correspondants sont différents soit en termes de données ou de distributions de probabilités marginales. Ce type d'apprentissage est exploité dans nos expériences présentées dans le chapitre 3.

Par ailleurs, il existe différentes stratégies de transfert. Si l'on souhaite utiliser le modèle pré-entraîné comme extracteur de caractéristiques, il suffit de ré-entraîner uniquement la couche finale du modèle destinée à la classification finale. Dans d'autres situations, on souhaite également ré-entraîner d'autres couches de manière sélective. L'idée est de fixer les poids de certaines couches pendant l'entraînement (pour profiter des connaissances de l'architecture globale du réseau) et d'affiner le reste du modèle.

L'apprentissage à partir de peu d'exemples

Ce type d'apprentissage (Few-Shot Learning : FSL) [25, 29], plus récent et pouvant être vu comme une sorte d'apprentissage par transfert [30], permet à un modèle pré-entraîné sur un jeu de données étendu (ex. ImageNet [31]) de généraliser sur de nouvelles catégories de données (que le modèle pré-entraîné n'a pas vu) en utilisant seulement quelques échan-

tillons étiquetés par classe nouvelle. Ainsi, le modèle ne nécessite pas un grand volume de données annotées et permet d'économiser de la puissance de calcul (pas d'entraînement depuis "zéro"). Les échantillons de données intervenant dans le FSL se décomposent en :

- **données de support** : quelques échantillons étiquetés par nouvelle catégorie de données qui seront utilisés par le modèle pré-entraîné pour généraliser sur ces nouvelles classes.
- **données de requête** : échantillons des nouvelles et anciennes catégories de données sur lesquelles le modèle doit généraliser en utilisant les connaissances antérieures et les informations obtenues à partir des données de support.

L'apprentissage FSL est considéré pour la segmentation sémantique d'images [32] afin de mieux modéliser la variation intra-classe des caractéristiques des objets.

Au cours de nos expériences décrites par la suite, nous utiliserons parfois des techniques d'augmentation de données ou d'apprentissage par transfert pour compenser les petites bases de données à disposition. En combinant les méthodes traditionnelles d'apprentissage profond à l'exploitation des relations structurelles, nous espérons améliorer les performances des CNNs en bénéficiant de relations spatiales globales (à l'image des *Transformers* décrits précédemment) tout en réduisant le nombre de données d'entraînement nécessaires. La notion d'information structurelle est présentée dans la section suivante.

1.3 Connaissances structurelles

Cette partie reprend les travaux de thèse de Jérémy Chopin [10]. Les connaissances structurelles sont des informations (relations spatiales, description qualitative, etc.) permettant de modéliser et d'exploiter la structure globale d'une scène dans divers contextes (analyse de documents [33], robotique [34]). Dans le cadre de l'analyse d'images par apprentissage profond, leur utilisation par les réseaux de neurones reste superficielle [35] et limitée à des informations de bas niveau (intensité des pixels, texture) ou à des informations de plus haut niveau mais à basse échelle (petites structures globales contenues dans le champ réceptif des réseaux de neurones, comme présenté dans la section 1.2.3).

Dans nos travaux, nous souhaitons réintroduire ces connaissances structurelles de haut niveau pour différentes tâches d'analyse d'images : la segmentation sémantique (chapitres 2 et 4), la classification d'images (chapitre 5 - section 5.4) et la régression à partir d'images (chapitre 5 - section 5.5). Dans cette partie, nous définissons plus précisément ce que sont

les relations structurelles avant de présenter comment celles-ci peuvent être modélisées par un graphe.

1.3.1 Définition

Une image est constituée de différentes régions (objets) dotées de propriétés spécifiques et entre lesquelles il est possible d'identifier des relations. Ces relations peuvent être de natures diverses, notamment spatiales (ex. distance, position directionnelle relative [36]), topologiques [37] ou photométriques (ex. luminosité relative, différence de colorimétrie [8]) comme illustré dans la Figure 1.6.

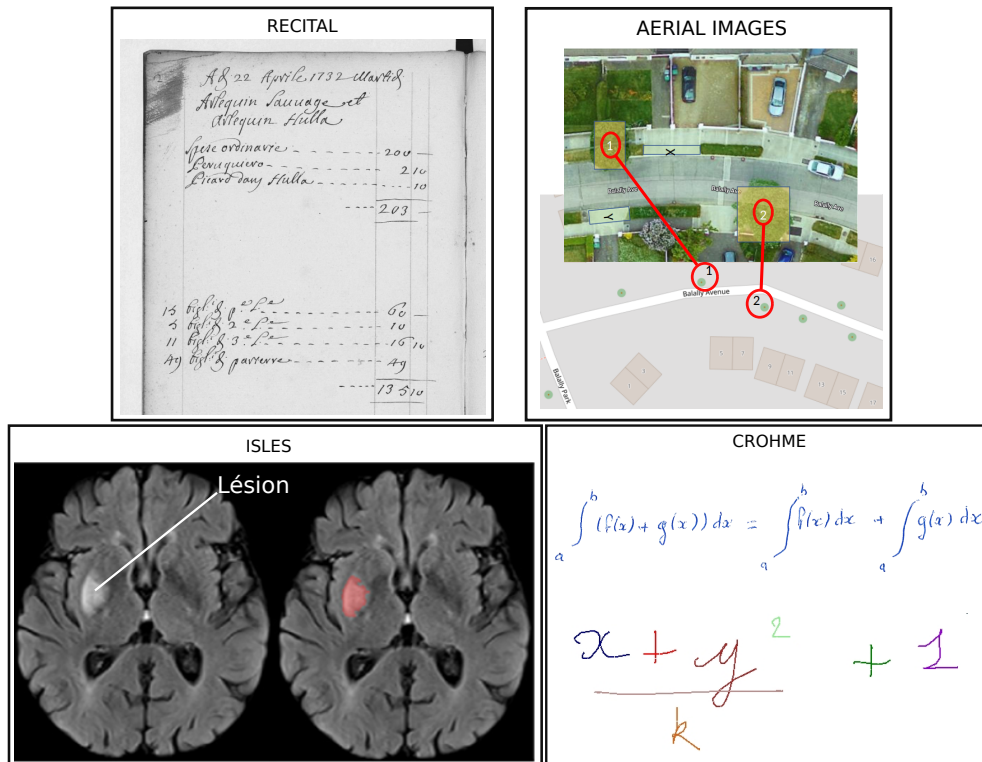


FIGURE 1.6 – Exemples de relations structurelles (images extraites des jeux de données RECITAL [38], ISLES [39], CROHME [40] et des travaux de Liu et al. [41]). On y observe des relations spatiales (les bâtiments sont à une certaine distance, l'exposant 2 est en haut à droite du y), topologiques (la lésion cérébrale est dans le cerveau) et photométriques (la lésion ischémique est plus claire que la substance grise, l'écriture sur le document est plus sombre que le papier).

Les information structurelles se sont montrées pertinentes dans différents domaines dont l'imagerie médicale [42, 43] qui nous intéressera particulièrement dans cette thèse.

Ces relations présentent l'avantage de pouvoir être formulées intuitivement par des personnes non spécialistes du traitement d'images [44] et de regrouper des connaissances *a priori* qui aideront à l'interprétation des modèles les incluant. Parmi l'ensemble de ces relations, on peut distinguer deux familles :

- Les relations **qualitatives** : elles représentent des concepts non numériques (adjacence, intersection, inclusion, etc.). Dans la Figure 1.6, l'inclusion de la lésion cérébrale ischémique dans le cerveau est une relation qualitative.
- Les relations **quantitatives** : elles représentent des relations représentables numériquement par un scalaire, un vecteur numérique ou un intervalle (distances, angles, différences d'intensité, etc.). Dans la Figure 1.6, la distance entre les bâtiments sur l'image aérienne est une relation quantitative.

Dans le cas de la segmentation sémantique d'images, une difficulté repose dans la combinaison de ces connaissances structurelles de haut niveau avec les informations de bas niveau à l'échelle du pixel. Une manière habituelle de modéliser les informations structurelles de haut niveau consiste à utiliser la notion de graphes. Chaque noeud du graphe représente une région de l'image et porte des informations propres à celle-ci (ex. intensité moyenne, volume) tandis que les arêtes entre noeuds informent sur les relations entre régions connectées (ex. distances, angles). Dans la section suivante, nous détaillons la notion de graphe et sa capacité à modéliser les informations structurelles.

1.3.2 Modélisation par un graphe

Soit une image I composée d'un ensemble de régions V connectées entre elles par des relations notées E . On définit un graphe $G = (V, E)$, tel que chaque noeud $v \in V$ correspond à une région (objet) de l'image et E indique l'ensemble des arêtes reliant les noeuds du graphe [45]. On dénote une arête reliant les noeuds $u \in V$ et $v \in V$ comme $(u, v) \in E$. Lorsque les arêtes indiquent une relation structurelle dans laquelle la source et la destination ont une importance (ex. la région source est au-dessus de la région destinataire : Figure 1.7 - a), elles peuvent être orientées ($(u, v) \neq (v, u)$). Dans nos travaux, nous considérerons des « graphes simples » [45] où une arête, au plus, relie deux noeuds, aucune arête ne relie un noeud à lui-même et les arêtes sont non orientées ($(u, v) = (v, u)$) comme dans la Figure 1.7 - b. Le voisinage d'un noeud v est l'ensemble des noeuds connectés à v par une arête [9] : $\mathcal{N}(v) = \{u \in V | (u, v) \in E\}$.

Un graphe peut modéliser diverses propriétés des régions sous forme d'attributs de noeuds et des relations structurelles entre régions sous forme d'attributs d'arêtes. On dé-

fini alors un graphe valué $G = (V, E, X, L)$ où $X : V \rightarrow \mathbb{R}^{d_v}$ est une fonction d'affectation d'attributs de noeuds (vecteurs numériques $A_v \in \mathbb{R}^{d_v}$) et $L : E \rightarrow \mathbb{R}^{d_e}$ est une fonction d'affectation d'attributs d'arêtes (vecteurs numériques $A_e \in \mathbb{R}^{d_e}$).

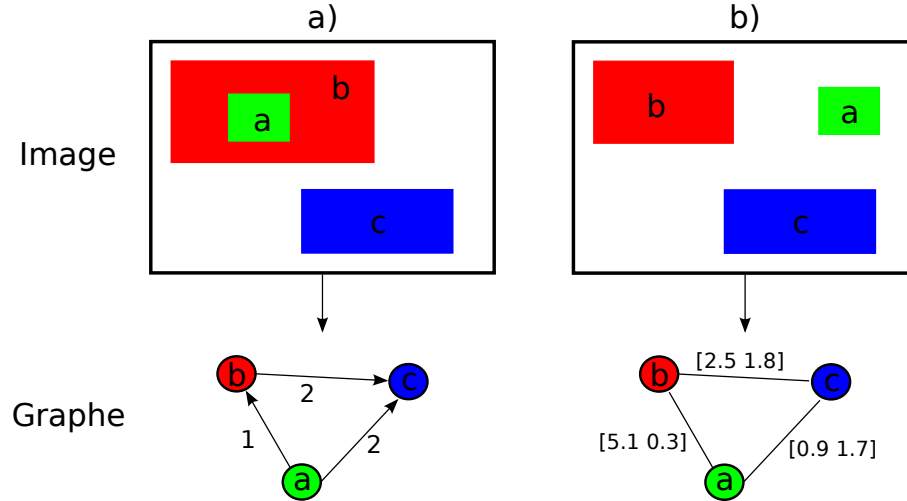


FIGURE 1.7 – Exemples de représentations d'une image sous forme de graphe. a) Représentation sous forme d'un graphe orienté où les attributs d'arêtes indiquent un type de relation structurelle (1 : « inclus dans », 2 : « au-dessus de »). b) Représentation sous forme d'un graphe non orienté où les arêtes portent des attributs multi-dimensionnels (distances au barycentre selon les axes x et y).

Les arêtes peuvent porter une simple information scalaire ($d_e = 1$) indiquant le poids de la relation entre les deux régions ou encore le type de connexion [46] mise en jeu (ex. Figure 1.7 - a : un type d'arête pour la relation « inclus dans » et un autre pour la relation « au-dessus de »). Dans d'autres cas, les attributs d'arêtes peuvent être des vecteurs multi-dimensionnels décrivant de multiples relations structurelles entre régions ($d_e > 1$). Ainsi, dans la Figure 1.7 - b, les attributs d'arêtes indiquent la distance au barycentre selon les 2 dimensions (x et y) entre les régions connectées.

A partir de la représentation sous forme de graphe d'une image, il est possible de réaliser un grand nombre de tâches parmi lesquelles la reconnaissance d'objets, la prédiction de relations entre régions, le regroupement de régions ou encore la segmentation sémantique de l'image considérée. Dans la section suivante, nous mettons en avant différentes utilisations des graphes pour l'analyse d'images et diverses approches employées pour y parvenir avant de décrire plus précisément la méthode retenue dans nos travaux de thèse : les réseaux de neurones sur graphes (GNNs).

1.4 Exploitation des graphes par réseaux de neurones pour l'analyse d'images

Nous débuterons cette partie par une introduction aux différentes applications pouvant être résolues grâce à l'exploitation de graphes et sur quelques méthodes de résolution indépendantes des réseaux de neurones. Nous poursuivrons par une description plus détaillée des techniques d'analyse d'images basées sur l'apprentissage profond sur graphes et considérées dans nos travaux de thèse. Nous terminerons cette section en nous focalisant sur une opération clé de ces réseaux de neurones sur graphes : la convolution.

1.4.1 Introduction

L'exploitation des informations structurelles sous forme de graphes a montré ses preuves pour aider à l'analyse d'images en proposant des solutions à de nombreuses applications. On peut, entre autres, mentionner l'identification d'objets [47], la segmentation sémantique d'images [48], la reconnaissance d'émotions [49] ou encore l'interprétation de documents manuscrits [50]. Ces applications peuvent être résolues grâce à diverses méthodes et approches. On peut notamment citer les méthodes d'analyse syntaxique de graphe (*graph parsing*) qui consistent à analyser une image en s'appuyant sur un graphe d'hypothèses décrivant le contenu de l'image et sur une grammaire propre au domaine d'application (connaissances *a priori*). Prenons l'exemple de l'interprétation d'équations mathématiques manuscrites [50], comme sur la Figure 1.6 - CROHME. La région associée au caractère "2" est en haut à droite de la région associée au caractère "y". À partir de l'analyse du graphe associé et de la grammaire associée au domaine mathématique, l'analyse syntaxique va permettre de vérifier que "2" est l'exposant de "y". À l'inverse, ces techniques montreront que "y", bien que situé en haut à droite du caractère "k", n'est pas l'exposant de "k". On réalise alors une prédiction de relation entre régions.

Certaines approches vont chercher à mesurer des similarités entre graphes [51] (c.a.d. entre deux objets ou images) pour différencier des graphes (ex. des images) ou encore pour identifier des noeuds du graphe (ex. des éléments de l'image). Ainsi, si l'on reprend les exemples de la Figure 1.6, on peut imaginer que la mesure de similarité d'un graphe décrivant un cerveau sain avec un graphe décrivant un cerveau pathologique fournira une différence permettant de distinguer les deux populations. On réalise alors de la classification de graphes, problème qui sera abordé dans la section 5.4 du chapitre 5 de cette thèse.

Cette mesure de similarité entre graphes peut être effectuée par diverses méthodes [51]. On peut citer, entre autres, la distance d'édition sur graphe (*Graph edit distance* (GED) [52]) qui mesure une distance entre deux graphes (i.e. similarité forte si distance faible). Si l'on considère deux graphes G_1 et G_2 , GED mesure la distance entre ces deux graphes comme une série de transformations (ou opérations d'édition) à appliquer au graphe G_1 pour obtenir le graphe G_2 (ajout, suppression ou substitution de noeuds ou d'arêtes). Elle est définie par :

$$\text{GED}(G_1, G_2) = \min_{(e_1, \dots, e_k) \in \mathcal{P}(G_1, G_2)} \sum_{i=1}^k c(e_i) \quad (1.3)$$

où $\mathcal{P}(G_1, G_2)$ dénote l'ensemble des chemins d'édition permettant de transformer G_1 en un graphe isomorphe à G_2 et $c(e) \geq 0$ représente le coup de chaque opération d'édition.

La similarité entre des graphes peut également être mesurée en appariant les graphes entre eux [53]. Ainsi, dans le cas de l'image aérienne (Figure 1.6), on peut imaginer un graphe où chaque noeud correspond à un objet (bâtiment, arbre, etc.). Grâce à un appariement de ce graphe avec un graphe dit modèle (construit à partir de connaissances *a priori* de paysages urbains), on peut imaginer classer chaque noeud du graphe et ainsi identifier chaque objet de la scène. On fait alors de la classification de noeuds, problème abordé dans les chapitres 2 et 4 de cette thèse. Différentes méthodes existent pour appairer des graphes entre eux. Sans chercher à être exhaustif, on peut citer le problème d'affectation quadratique (*Quadratic assignment problem* (QAP) [54, 55]). Considérons encore les deux graphes G_1 et G_2 . Si ces deux graphes ont la même structure et les mêmes attributs, on définit l'appariement comme un appariement exact de graphes basé sur la notion d'isomorphisme de graphes. En analyse d'images, on rencontre souvent des variations dans les images à analyser qui conduisent à des graphes de structures (nombre de noeuds et d'arêtes) différentes et avec des attributs différents. On définit alors l'appariement comme inexact. Le QAP définit l'appariement de graphes comme une matrice d'affectation $Z \in \{0, 1\}^{|V_1| \times |V_2|}$, où $Z_{ij} = 1$ indique que le noeud i de G_1 est associé au noeud j de G_2 . L'objectif du QAP consiste à trouver l'appariement optimal Z^* minimisant les dissimilarités (représentées par une matrice K) entre noeuds et arêtes des graphes G_1 et G_2 . Il peut être formulé, comme mentionné dans [10, 54], par :

$$Z^* = \arg \min_Z \text{vec}(Z)^T K \text{vec}(Z) \quad (1.4)$$

Les méthodes sur graphes introduites dans cette partie sont limitées par leur nature intrinsèque hautement combinatoire, bien que des travaux récents tentent de réduire leur

complexité par des algorithmes cherchant des solutions approchées (GED [56, 57], QAP [58]). Pour pallier cette limitation, d'autres approches se basent sur des réseaux de neurones pour apprendre les relations entre structures et donc apprendre à réaliser ces tâches (classification de graphes, de noeuds, etc.). Ces réseaux de neurones sur graphes, au coeur de nos travaux de thèse, sont présentés dans la section suivante.

1.4.2 Les réseaux de neurones sur graphes

Les réseaux de neurones sur graphes (GNNs) constituent un domaine émergent et actif dans le contexte de l'apprentissage profond [9, 59]. Cet engouement est notamment lié à la grande habilité des graphes pour modéliser diverses données structurées, comme présenté dans la section 1.3.2. Les GNNs permettent de réaliser différentes tâches sur graphes, parmi lesquelles on trouve la classification [60] et la régression [61] de graphes, la classification de noeuds [62, 63], la prédiction de liens [64], le regroupement de noeuds [65] ou encore la génération de graphes [66] (illustrées sur la Figure 1.8).

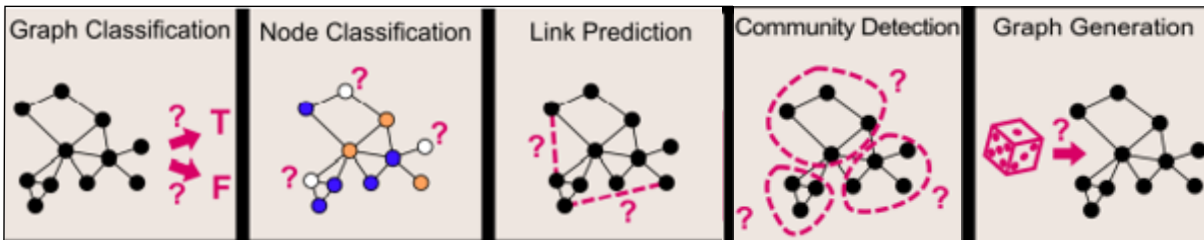


FIGURE 1.8 – Exemples de tâches pouvant être réalisées avec les réseaux de neurones sur graphes (images issues de [67]).

Dans la suite de la thèse, nous exploiterons la classification de noeuds (chapitres 2 et 4), la classification de graphes (chapitre 5 - section 5.4) et la régression de graphes (chapitre 5 - section 5.5). Le but de la classification de noeuds est de prédire la classe y_v des noeuds $v \in V$ d'un graphe à partir de données d'entraînement avec des noeuds étiquetés [45]. Les noeuds étiquetés peuvent faire partie du même graphe que les noeuds à classer (comme dans la Figure 1.8 : il s'agit alors d'apprentissage semi-supervisé) ou appartenir à d'autres graphes totalement étiquetés utilisés pour l'entraînement (on parle alors d'apprentissage supervisé). Nous avons considéré cette deuxième approche dans nos travaux. Dans le cas de la classification ou de la régression de graphes, l'objectif est de faire une prédiction sur le graphe entier (classe, valeur numérique) et non sur les composants individuels de celui-ci (noeuds et arêtes).

Pour réaliser ces tâches, les GNNs se basent sur une architecture profonde partageant beaucoup de similarités avec les réseaux de neurones convolutifs présentés dans la section 1.2.2. La Figure 1.9 donne un exemple d'architecture pour la classification de graphes [68].

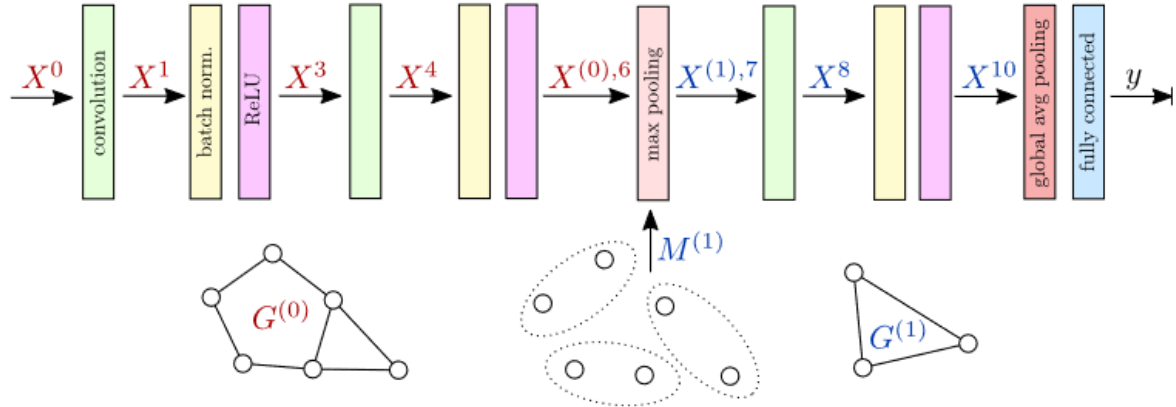


FIGURE 1.9 – Exemple d'architecture de GNN pour la classification de graphes (image extraite de [67]).

On retrouve une architecture générale proche de celle des CNNs avec des opérations de convolution suivies de normalisation et de fonctions d'activation (ex. ReLU), des opérations de sous-échantillonnage (*pooling*) et une couche entièrement connectée pour la classification finale. Ces opérateurs sont évidemment adaptés aux graphes. Ainsi, l'opération de sous-échantillonnage consiste à regrouper des noeuds du graphe entre eux en combinant leurs attributs afin de conserver l'information générale du graphe. Le regroupement peut se faire en plusieurs sous-groupes (ex. *max pooling* dans la Figure 1.9) ou à l'échelle du graphe (*global pooling*). Dans le cas du regroupement en sous-ensembles, une stratégie habituelle consiste à utiliser des techniques de *clustering* sur graphe [65] pour définir les noeuds à fusionner [45]. L'attribut du nouveau noeud de regroupement peut se calculer comme la somme, la moyenne ou encore la valeur maximum des attributs $X(v)$ des noeuds v inclus dans le regroupement, à l'instar des opérations de sous-échantillonnage des CNNs sur images (regroupement de pixels voisins).

A chaque couche l du réseau, les GNNs produisent des représentations de noeuds notées $X^l(v)$ ($X^0(v)$ étant les attributs du noeud v du graphe $G(V, E, X, L)$ initial). Tout comme pour les CNNs, l'un des opérateurs clés des réseaux de neurones sur graphes est la convolution. Nous présentons, dans la section suivante, l'opération de convolution sur graphe et différentes variations ayant été proposées pour cet opérateur.

1.4.3 La convolution sur graphe

Dans le domaine des graphes, deux approches existent pour l'opération de convolution : l'approche spectrale et l'approche spatiale [9]. L'approche spectrale est basée sur la théorie spectrale des graphes et suppose une décomposition en valeurs propres de la matrice Laplacienne du graphe afin d'identifier des sous groupes dans celui-ci. Nous faisons le choix de ne pas présenter cette approche plus en détail car nous ne nous y sommes pas intéressés dans nos travaux. En effet, nous avons choisi d'utiliser une approche spatiale qui formule la convolution sur graphe comme une agrégation d'informations de voisinage. Les approches spatiales, en opérant directement dans l'espace des graphes et non dans l'espace de Fourier, ne requièrent pas le traitement simultané de l'ensemble du graphe pouvant entraîner des temps de calcul relativement conséquents [69]. Par ailleurs, comme les modèles spatiaux opèrent localement sur chaque noeud du graphe, ils peuvent facilement s'appliquer à des graphes de tailles arbitraires. Ce n'est pas le cas des approches spectrales, qui en opérant sur le graphe entier, suppose des graphes de taille fixe [69]. Pour ces raisons, nous avons considéré une convolution dans le domaine spatial.

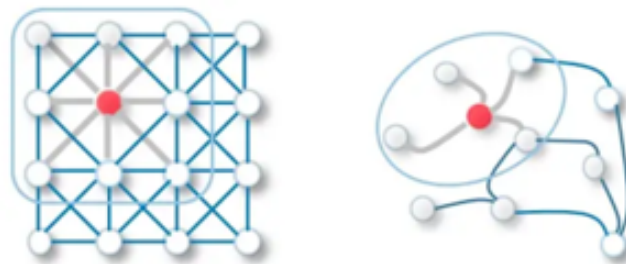


FIGURE 1.10 – Analogie entre convolution sur image 2D par CNN et convolution sur graphe (image extraite de [69]).

La convolution est une étape importante des réseaux de neurones sur graphes au cours de laquelle la représentation de chaque noeud $v \in V$ est actualisée par un mécanisme de diffusion de l'information à travers le graphe (*message passing*) [9] pouvant être divisé en deux étapes :

1. Chaque noeud $v \in V$ du graphe génère un "message" basé sur sa représentation $X^l(v)$ actuelle qu'il envoie à ses voisins $\mathcal{N}(v) = \{u \in V | (u, v) \in E\}$. Le message peut être actualisé au cours du transfert par les attributs des arêtes empruntées pour rejoindre chaque voisin. Ainsi, le noeud indiqué en rouge sur la Figure 1.10,

envoie son "message" uniquement aux 4 noeuds auxquels il est connecté par une arête.

2. Chaque noeud $v \in V$ actualise sa représentation (conduisant à $X^{l+1}(v)$) à partir des messages entrants reçus par ses voisins et de sa représentation actuelle $X^l(v)$. Le noeud représenté en rouge sur le graphe de la Figure 1.10 actualise son état à partir des messages parvenus par ses 4 arêtes incidentes (c.a.d. par ses 4 voisins).

Comme illustré dans la Figure 1.10, la convolution sur graphe est similaire à la convolution des réseaux de neurones convolutifs sur image. Tout comme pour les CNNs, où l'information de chaque pixel de l'image est mise à jour à partir de l'information des pixels voisins (ensemble des voisins défini par la taille du filtre de convolution), l'information de chaque noeud du graphe est actualisée à partir de l'information des noeuds voisins. La différence repose dans le fait que, dans le cas des graphes, les voisins de chaque noeud ne sont pas ordonnés et leur nombre varie (deux noeuds différents n'ont pas toujours le même nombre de voisins). La convolution, modifiant la représentation $X^{l+1}(v)$ d'un noeud v à la couche $l + 1$, peut être formulée comme (notation inspirée de [9]) :

$$X^{l+1}(v) = \phi^{l+1}(X^l(v), \Psi(\{e^{l+1}(L((u, v)))\psi^{l+1}(X^l(u)) | u \in \mathcal{N}(v)\})) \quad (1.5)$$

Le nouvel état $X^{l+1}(v)$ du noeud v dépend donc de sa propre représentation $X^l(v)$, d'une fonction quelconque $\psi^{l+1}(\cdot)$ (ex. perceptron multi-couches) sur les noeuds voisins $X^l(u)$ et parfois d'une fonction $e^{l+1}(\cdot)$ sur les attributs d'arêtes $L((u, v))$ incidentes au noeud v . La fonction $\Psi(\cdot)$ agrégeant les informations de voisinage est une fonction invariante par permutation (ex. moyenne, maximum, produit, somme). Après agrégation de l'ensemble des paramètres, une fonction non linéaire $\phi^{l+1}(\cdot)$ (ex. sigmoïde, ReLU, Tanh) est appliquée à la sortie résultante.

Ces dernières années, une grande variété d'opérateurs de convolution a été proposée [59]. Parmi les différences entre ces opérateurs, nous nous sommes intéressés à la fonction $e^{l+1}(\cdot)$ tenant compte des attributs d'arêtes dans le *message passing*. Ainsi, certains opérateurs considèrent uniquement les attributs de noeuds voisins (ex. GCNConv [63], GraphConv [70]) tandis que d'autres combinent à la fois les attributs (multidimensionnels) sur noeuds et sur arêtes lors de l'agrégation de voisinage (ex. ECCConv [68], GATConv [71]). Pour mieux illustrer cette différence, nous prenons l'exemple de GraphConv (eq 1.6) et de ECCConv (eq 1.7) appliqués au graphe de la Figure 1.11.

$$X^{l+1}(v) = W_1^{l+1}X^l(v) + W_2^{l+1} \sum_{u \in \mathcal{N}(v)} w_{u,v}X^l(u) \quad (1.6)$$

$$X^{l+1}(v) = W_1^{l+1}X^l(v) + \sum_{u \in \mathcal{N}(v)} F^{l+1}(L((u, v), W_e^{l+1}))X^l(u) \quad (1.7)$$

où W_1 , W_2 et W_e sont des matrices de poids entraînaables et $w_{u,v}$ est un poids d'arête (scalaire) fixé à 1 le plus souvent.

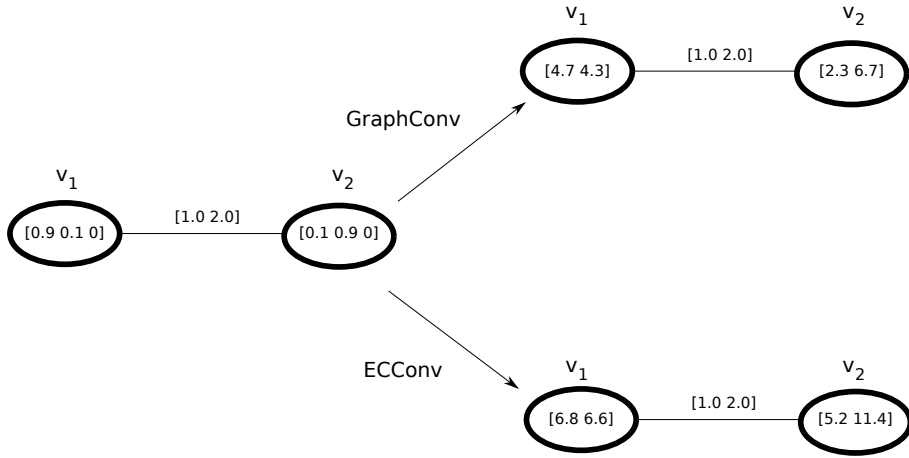


FIGURE 1.11 – Exemple de résultats des opérateurs de convolution GraphConv et ECConv sur un graphe simple constitué de 2 noeuds (v_1 et v_2) reliés par une arête. La dimension des attributs de noeuds en sortie de convolution est fixée à 2.

Dans les deux cas, nous fixons arbitrairement la dimension de sortie des attributs de noeuds $d^{l+1} = 2$ et $W_1^{l+1} \in \mathfrak{R}^{d^{l+1} \times d^l} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 5 & 3 \end{pmatrix}$. Considérons, par exemple, le noeud v_2 et son voisin v_1 tels que :

$$X^l(v_2) = \begin{pmatrix} 0.1 \\ 0.9 \\ 0 \end{pmatrix} \text{ et } X^l(v_1) = \begin{pmatrix} 0.9 \\ 0.1 \\ 0 \end{pmatrix}$$

Pour GraphConv, nous fixons la matrice appliquée aux noeuds voisins $W_2^{l+1} \in \mathfrak{R}^{d^{l+1} \times d^l}$:

$$W_2^{l+1} = \begin{pmatrix} 2 & 3 & 2 \\ 2 & 3 & 2 \end{pmatrix} \text{ On a alors : } X^{l+1}(v_2) = W_1^{l+1} \times X^l(v_2) + W_2^{l+1} \times X^l(v_1) = \begin{pmatrix} 2.3 \\ 6.7 \end{pmatrix}$$

Dans le cas de ECCConv, la fonction $F^{l+1} : \mathfrak{R}^{d_e} \rightarrow \mathfrak{R}^{d^{l+1}}$ est une fonction de *mapping*, mettant en correspondance les poids spécifiques aux attributs d'arêtes avec les dimensions des noeuds [68]. Elle peut être modélisée par un réseau de neurones caractérisé par les poids W_e^{l+1} . Dans cet exemple, on se limite à un simple produit par la matrice de poids $W_e^{l+1} \in \mathfrak{R}^{d_e \times d^{l+1}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 2 & 2 & 3 & 2 \end{pmatrix}$. F se calcule alors comme :

$$F^{l+1}(L((v_1, v_2), W_e^{l+1})) = (1 \ 2) \times W_e = (5 \ 7 \ 5 \ 5 \ 7 \ 5) \rightarrow \theta_{v_1 v_2}^{l+1} = \begin{pmatrix} 5 & 5 & 7 \\ 7 & 5 & 5 \end{pmatrix}$$

On a alors :

$$X^{l+1}(v_2) = W_1^{l+1} \times X^l(v_2) + \theta_{v_1 v_2}^{l+1} \times X^l(v_1) = \begin{pmatrix} 5.2 \\ 11.4 \end{pmatrix}$$

On observe bien que GraphConv ne considère pas les attributs d'arêtes $(1 \ 2)$ dans son calcul contrairement à ECCConv. On peut alors s'interroger sur l'impact de l'opérateur de convolution dans les performances du GNN. Les résultats de ces différents opérateurs de convolution (GCNConv, GraphConv, ECCConv et GATConv) pour la segmentation sémantique d'images seront étudiés dans le chapitre 2. L'opérateur ECCConv sera utilisé dans les travaux des chapitres 4 et 5.

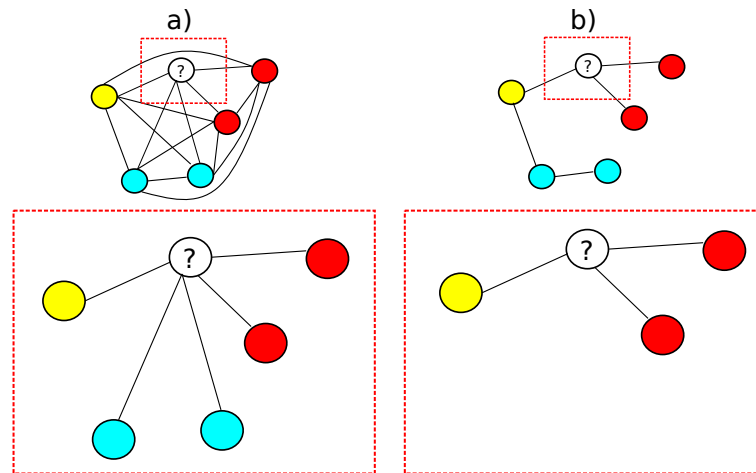


FIGURE 1.12 – Quel est l'impact de la topologie des graphes lors de l'opération de convolution dans une tâche comme la classification de noeuds ? a) Graphe complet. b) Graphe non complet. La classe du noeud non étiqueté est "rouge".

Une autre question à laquelle nous nous sommes intéressés repose sur la considération de graphes complets (c.a.d où tous les noeuds sont connectés entre eux), notamment pour l’application de la convolution. En effet, de par la nature invariante par permutation de la fonction $\Psi(\cdot)$ (eq 1.5) intervenant lors de la convolution, on peut craindre un manque de discrimination entre les noeuds voisins intervenant dans l’actualisation de la représentation d’état de chaque noeud du graphe. Par exemple, pour la tâche de classification de noeuds, on souhaiterait que chaque noeud se différencie, entre autre par son voisinage, pour distinguer les différentes classes. Ainsi, dans l’exemple de la Figure 1.12, considérons que le noeud non étiqueté appartient en fait à la classe rouge. On peut imaginer qu’en considérant un graphe complet (Figure 1.12 - a), l’information générale provenant des noeuds voisins (de 3 classes différentes) sera confuse (autant de voisins rouges que bleus). En modifiant de façon appropriée la topologie du graphe (Figure 1.12 - b), il n’y a plus cette ambiguïté pour la classification du noeud.

La confrontation à des graphes complets est un cas de figure présent dans l’analyse d’images car il est possible de donner ou de calculer des relations entre toutes les régions d’une image. Dans d’autres contextes, les graphes peuvent ne pas être complets (l’absence d’arête n’est pas un choix). Si l’on prend les exemples connus de Citeseer [72] ou Cora [73], un noeud (représentant un article) est connecté à un autre seulement s’il y fait référence. De la même manière, dans un graphe associé à un réseau social [74], il ne peut pas y avoir d’arête entre deux personnes non connectées. En analyse d’images, la considération des graphes d’adjacence de régions (RAG) [75, 76] permet d’arriver à la même situation : seuls les noeuds correspondant à des régions adjacentes seront connectés par une arête. Mais, cette technique ne permet alors pas d’exploiter toutes les informations structurelles d’intérêt dans l’image, comme présentées dans la section 1.3.

La question de la topologie des graphes dans le contexte des GNNs a fait l’objet de différents travaux dans la communauté. Ainsi, *Yang et al.* [77] ont développé un algorithme de mise à jour de la topologie de graphe, supprimant les arêtes inter-classes et ajoutant des arêtes intra-classes selon les classes de noeuds prédites par un modèle GNN. *Li et al.* [78] propose le GNN *Sparsified Graph Convolutional Network* formulant le problème d’élagage d’arêtes comme un problème d’optimisation résolu par la méthode des multiplicateurs à direction alternée (ADMM). D’autres proposent un mécanisme de *dropout* sur arêtes [79], similaire à celui des CNNs. On trouve également des mécanismes de suppression d’arêtes (*edge-cutting*) développés plutôt pour le regroupement de noeuds [80].

Dans le chapitre 2, nous présenterons une méthode de segmentation sémantique d’images

combinant CNNs, connaissances structurelles et GNNs. Nous verrons que, dans ce contexte, la question relative à la topologie des graphes a une influence.

1.5 Conclusion

Les réseaux de neurones, notamment convolutifs (CNNs), ont montré leur capacité pour diverses applications liées à l'analyse d'images. Les nombreuses architectures existantes témoignent de l'engouement de la communauté pour ces méthodes d'apprentissage profond. Les CNNs exploitent les informations de l'image à l'échelle du pixel (ex. intensité) avec les relations spatiales de voisinage, dans la limite du champ réceptif de l'opérateur de convolution. Si certaines architectures comme les *Vision Transformers* tentent de capturer des relations spatiales à plus grande échelle via des mécanismes d'attention, cela reste limité. Dans ses travaux [10], J. Chopin a montré qu'un post-traitement intégrant des relations spatiales de haut-niveau, basé sur de l'appariement de graphes formulé sous forme de QAP, permet d'améliorer la segmentation sémantique de ViT. Par ailleurs, un défaut des architectures de type *Vision Transformer* concerne le nombre de paramètres entraînaibles démultiplié. Or, cette complexification induit un plus grand nombre de données d'entraînement nécessaire. Les techniques existantes pour faire face aux petite bases de données (ex. augmentation de données, apprentissage par transfert) ne sont alors pas toujours suffisantes.

Pour intégrer un contexte structurel global sans exiger trop de données d'entraînement, nous proposons d'exploiter les connaissances structurelles disponibles à haut niveau (ex. relations spatiales, description qualitative). Ces informations spatiales, ayant déjà montré leur pertinence dans le domaine médical notamment, peuvent être modélisées sous forme de graphes où les noeuds représentent des régions de l'image et les arêtes portent des informations sur les relations entre celles-ci. Les graphes construits peuvent alors être utilisés pour guider l'interprétation d'une image à l'aide de nombreuses approches basées sur graphes. Parmi ces méthodes, nous nous intéressons aux réseaux de neurones sur graphes (GNNs) dont l'architecture, similaire à celle des CNNs, permet de réaliser différentes tâches abordées dans cette thèse : la classification de noeuds (chapitres 2 et 4), la classification de graphes et la régression de graphes (chapitre 5).

Dans ce contexte, nous nous posons certaines questions relatives à l'exploitation des connaissances structurelles par les GNNs. Dans le chapitre 2, nous nous demandons si les connaissances structurelles peuvent permettre d'améliorer les performances des CNNs

pour la segmentation sémantique d’images, en s’appuyant sur des GNNs et non sur des techniques combinatoires comme le QAP (cela ayant déjà été démontré par les travaux de J. Chopin [10]). L’influence de l’opérateur de convolution utilisé dans l’architecture du GNN et de la topologie des graphes considérés en entrée de GNN est étudiée. Dans la partie II de cette thèse, nous tentons d’exploiter CNNs, informations structurelles et GNNs pour aider dans l’étude de la lésion cérébrale précoce de l’enfant après AVC néonatal. Plus précisément, nous cherchons à savoir si ces outils peuvent aider à comprendre la paralysie cérébrale de l’enfant, déficience motrice liée à la lésion cérébrale précoce et développée par seulement une partie des enfants avec lésion.

À retenir

- Les réseaux de neurones profonds, notamment convolutifs (CNNs), se sont montrés efficaces pour diverses tâches d’analyse d’images (classification, segmentation sémantique) et peuvent faire face au possible manque de données d’entraînement, particulièrement présent dans le domaine médical.
- Les connaissances structurelles, représentant les régions d’une image et les relations entre elles, restent partiellement utilisées par les CNNs pour l’analyse d’images.
- Les connaissances structurelles peuvent être modélisées sous forme de graphes pouvant être exploités par diverses approches dont les réseaux de neurones sur graphes (GNNs).
- La performance des GNNs dans diverses tâches (classification de graphes, classification de noeuds), dont l’architecture se rapproche de celle des CNNs, semble être influencée par divers facteurs tels que le choix de l’opérateur de convolution (fixant les attributs du graphe pris en compte) ou la topologie des graphes.

UTILISATION DES RÉSEAUX DE NEURONES SUR GRAPHES POUR LA SEGMENTATION SÉMANTIQUE D'IMAGES

2.1 Introduction

Dans ce chapitre, nous nous intéressons à l'impact des connaissances structurelles (section 1.3) pour améliorer les performances des réseaux d'apprentissage profonds (ex. CNNs, *Vision Transformers* présentés section 1.2.3) pour une tâche spécifique et fondamentale de l'analyse d'images : la segmentation sémantique (présentée dans la section 1.2.2). Nous proposons d'utiliser des réseaux de neurones sur graphes (GNNs introduits en section 1.4.2) pour exploiter les informations structurelles d'images segmentées par un réseau de neurones profond. Les motivations à l'origine de l'approche que nous allons présenter ont été évoquées dans le chapitre précédent et peuvent se résumer ainsi :

- Nous espérons améliorer les performances de segmentation des réseaux d'apprentissage profonds basés sur des couches convolutives et des mécanismes d'attention en exploitant des relations structurelles à plus grande échelle entre les régions de l'image (limitées à la taille du champ réceptif pour les CNNs) et en limitant la complexité des architectures, ceci étant un défi de l'apprentissage profond [81].
- Nous souhaitons représenter les informations structurelles de l'image sous forme de graphes qui seront exploités par des GNNs afin d'éviter la nature hautement combinatoire de solutions alternatives telles que les QAP [54] ou les GED [52] évoquées section 1.4.1.

Le principe de notre méthode consiste à considérer un GNN en post-traitement d'un réseau de neurones profond (CNN) pour effectuer une classification de noeuds (section 1.4.2) sur des graphes encodant les probabilités d'appartenance des régions aux différentes classes et les informations structurelles liées aux régions segmentées par le CNN.

Nous avons évoqué dans le chapitre précédent (section 1.4.3), l’influence de certains facteurs dans la performance des GNNs (choix de l’opérateur de convolution, topologie des graphes manipulés). Ces deux aspects seront considérés au travers de comparaisons entre 4 opérateurs de convolution (considérant ou non les attributs d’arêtes) et entre différents élagages de graphes (basés sur l’information spatiale des arêtes). A noter que nous étudions également la robustesse de notre méthode vis à vis de la taille de la base de données d’entraînement. Notre proposition vise à être moins sensible au manque de diversité et de représentativité dans la base d’apprentissage, en faisant l’hypothèse que les informations structurelles de « haut-niveau » utilisées, extraites des images d’entraînement segmentées, sont suffisamment constantes d’une image à l’autre. Nous espérons ainsi limiter le besoin de techniques alternatives telles que l’augmentation de données ou l’apprentissage par transfert (section 1.2.4).

Nous commencerons ce chapitre en présentant le positionnement de notre méthode par rapport aux travaux existants combinant CNNs et GNNs pour la segmentation sémantique d’images. Nous détaillerons ensuite la méthode proposée, c’est-à-dire la manière de modéliser les informations structurelles et la carte probabiliste découlant de l’image segmentée par un CNN ainsi que l’architecture du GNN utilisé pour la classification de noeuds. Nous poursuivrons par une description des expériences menées pour évaluer la pertinence de la méthode sur deux bases de données publiques, avant d’analyser les résultats obtenus. Nous achèverons le chapitre par une discussion des avantages et limites de notre approche.

2.2 CNNs et GNNs pour la segmentation sémantique

Les GNNs ont été considérés dans le contexte de la vision par ordinateur pour différentes tâches (certaines sont présentées dans la section 1.4.2) dont la segmentation sémantique d’images, au centre de ce chapitre. Ces dernières années, de nombreuses méthodes ont émergé combinant GNNs et CNNs pour la segmentation d’images.

Certaines approches proposent d’exploiter les informations relatives aux superpixels (petits groupes de pixels voisins partageant des caractéristiques communes comme l’intensité) pour la construction des graphes. Chaque noeud du graphe est ainsi associé à un superpixel. Néanmoins, ces méthodes supposent l’utilisation d’algorithmes supplémentaires de sur-segmentation, tels que SLIC [82] pour générer les superpixels. Ainsi, *Ouyang et al* [83] considèrent un réseau résiduel d’attention en "U" pour extraire des caractéris-

tiques de « haut-niveau » avant d’appliquer SLIC pour extraire les superpixels de l’image segmentée. Un GNN est ensuite utilisé pour effectuer une classification de noeuds sur les graphes construits à partir des superpixels, et corriger ainsi les erreurs de segmentation commises par le réseau d’apprentissage profond. De la même manière, *Diao et al* [84] proposent un GNN avec mécanismes d’attention basé sur des superpixels (SAGNN) pour la segmentation sémantique d’images aériennes. Chaque noeud des graphes, correspondant à un superpixel obtenu avec SLIC, est associé à un vecteur de représentation extrait par un CNN. Dans le domaine médical, on observe la même tendance à combiner GNNs et CNNs à travers la considération de superpixels [85].

D’autres propositions récentes visent à intégrer les relations spatiales des images à l’aide des GNNs dans l’espace des caractéristiques des réseaux de neurones profonds afin de capturer les dépendances spatiales à plus grande échelle [86, 87, 88, 89]. Néanmoins, ce type d’approche requiert le développement d’architectures spécifiques, démultipliant potentiellement le nombre de paramètres entraînaibles.

Enfin, certains travaux combinant GNNs et CNNs ont été conduits pour des applications spécifiques comme la segmentation binaire de vaisseaux rétiniens [90] ou encore pour la segmentation d’images biomédicales basée sur une approche par contours [91].

Comparée à toutes ces méthodes, notre approche propose d’utiliser les GNNs comme une couche de post-traitement pouvant être appliquée à la sortie de n’importe quel CNN produisant une carte de segmentation et pour n’importe quelle application, sans nécessiter le développement d’une architecture spécifique basée sur le CNN (comme dans [86, 87, 88, 89]) et sans intégrer d’algorithmes additionnels de sur-segmentation produisant des superpixels (comme dans [83, 84, 85]).

Notre approche ayant été positionnée dans l’état de l’art relatif à la combinaison de CNNs et GNNs pour la segmentation sémantique d’images, nous la présentons plus en détail dans la section suivante.

2.3 Méthode proposée

La Figure 2.1 fournit une représentation d’ensemble de la méthode proposée. Dans un premier temps, un réseau de neurones profond (ex. CNN, *Vision Transformer*) est entraîné à réaliser de la segmentation sémantique d’images à partir d’une base de données d’apprentissage annotée (Figure 2.1-A). Un réseau de neurones sur graphes (GNN) est également entraîné, à partir de la base de données annotée, à classer correctement les

noeuds des graphes générés à partir des images segmentées par le réseau de neurones profond. Dans le même temps, un rayon d'élagage des arêtes du graphe (paramètre ρ - Figure 2.1), détaillé en section 2.3.1 est optimisé à partir des données d'entraînement.

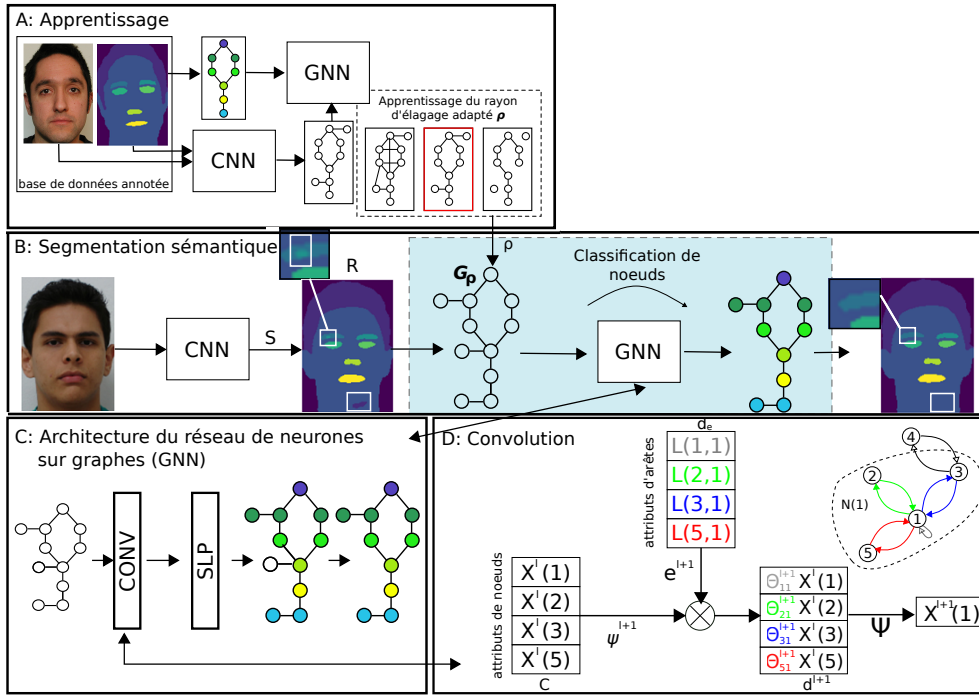


FIGURE 2.1 – Vue générale de la méthode proposée.

Ainsi, à la rencontre d'une nouvelle image (Figure 2.1-B), le réseau de neurones profond entraîné fournit une première segmentation pouvant contenir des erreurs. Dans l'exemple, on observe notamment des cheveux détectés au niveau du menton ou encore des yeux dans les sourcils. A partir de cette proposition de segmentation, un graphe, élagué en considérant le rayon ρ optimal, est construit. Le GNN précédemment entraîné est ensuite considéré pour effectuer une opération de classification de noeuds sur le graphe en question. A partir de cette classification des noeuds du graphe, les régions de l'image sont finalement identifiées (impliquant la correction des erreurs structurelles mentionnées précédemment dans cet exemple).

Dans les sections qui suivent, nous verrons comment les graphes sont construits à partir des images segmentées par les CNNs (section 2.3.1) et nous décrirons plus en détail l'architecture du GNN considéré pour la classification de noeuds (section 2.3.2).

2.3.1 De l'image au graphe

Dans un contexte de segmentation sémantique d'images, nous avons vu dans la section 1.2.2 que les réseaux de neurones profonds (ex. CNNs, *Vision Transformers*) entraînés sur une base de données annotées, fournissent une carte de probabilités à partir d'une image d'entrée. Nous avons vu que cette carte de probabilité est un tenseur $S \in \mathfrak{R}^{P \times C}$, où P est la dimension de l'image et C est le nombre total de classes considérées lors de la segmentation. Pour chaque pixel (voxel) de position p , $S(p, c) \in [0, 1]$ est la probabilité que ce pixel (voxel) appartienne à la classe c . La carte de probabilité S est ensuite utilisée pour construire une carte de segmentation \mathcal{L}^* (eq 1.2) en attribuant à chaque pixel (voxel) la classe la plus probable. Comme précédemment mentionné, cette étape peut être réalisée par n'importe quel réseau de neurones profond dédié à la segmentation sémantique tels que ceux présentés dans la section 1.2.3. A partir de la carte de segmentation \mathcal{L}^* , on construit un ensemble R des composantes connexes résultantes (c.a.d. des ensembles de pixels connectés a priori appartenant à la même classe selon \mathcal{L}^*) comme illustré sur la Figure 2.1. A partir de cet ensemble R , nous construisons un graphe $G = (V, E, X, L)$ où chaque noeud $v \in V$ correspond à une région $R_v \in R$. Pour chaque région R_v , nous définissons un vecteur de probabilité d'appartenance moyen S_v , basé sur la carte de probabilité S définie précédemment, limitée à la zone R_v considérée :

$$\forall v \in V, \forall c \in C, S_v[c] = \frac{1}{|R_v|} \sum_{p \in R_v} S(p, c) \quad (2.1)$$

Le vecteur S_v peut être assimilé à un vecteur de probabilité car $\sum_{i \in \{1, \dots, C\}} S_v[i] = 1$ et $\forall i \in \{1, \dots, C\}, 0 \leq S_v[i] \leq 1$. La fonction $X : V \rightarrow \mathfrak{R}^C$ de G , attribue à chaque noeud $v \in V$ le vecteur S_v de probabilité d'appartenance moyen à chaque classe $c \in C$. $L : E \rightarrow \mathfrak{R}^{d_e}$, associe à chaque arête un attribut de dimension d_e dépendant de la relation spatiale considérée, celle-ci étant un hyperparamètre de notre méthode. Nous verrons dans nos expériences (section 2.4), des exemples de relations pouvant être envisagées.

A partir de la base d'apprentissage (entraînement et validation), nous cherchons un niveau d'élagage ρ optimal (donnant les meilleures performances de classification de noeuds sur les données de validation). Dans notre méthode actuelle, nous considérons une série de rayons d'élagage définis de façon arbitraire à partir des distances moyennes entre les régions du graphe (voir section 2.4) afin d'étudier l'impact de la topologie du graphe sur les résultats. Notons que d'autres techniques d'élagage auraient pu être employées comme la technique des k plus proches voisins [92] ou celles présentées dans la section 1.4.3. A partir

du rayon d’élagage ρ optimisé, on élague le graphe G en un graphe $G_\rho = (V, E_\rho, X, L)$ tel que $E_\rho = \{e \in E \mid \|L(e)\| < \rho\}$, $\|\cdot\|$ étant la norme euclidienne.

2.3.2 Architecture du GNN

Comme illustré dans la Figure 2.1-C, le GNN considéré n’est composé que de deux couches. La première est une couche de convolution et la deuxième est un perceptron à couche unique. Le nombre de noeuds à classifier est *a priori* inconnu et variable (le nombre de régions candidates détectées par le CNN varie d’une image à l’autre et d’un CNN à un autre). On fait l’hypothèse que ce nombre de noeuds est supérieur au nombre de classes C (hypothèse réaliste de sur-segmentation [8]). Les graphes traités sont donc de taille arbitraire. Par conséquent, pour les raisons expliquées dans la section 1.4.3, nous considérons une approche spatiale et non spectrale pour l’opérateur de convolution.

La première couche du réseau (couche de convolution : $(\mathbb{R}^C \rightarrow \mathbb{R}^{d^{+1}})$) a pour objectif d’agréger les informations de voisinage de chaque noeud (notion de *message passing* introduite dans la section 1.4.3 et pouvant être définie par l’eq 1.5). La dimension d^{+1} des attributs de noeuds à la sortie de la convolution (Figure 2.1-D) est un autre hyperparamètre de notre méthode défini empiriquement pendant nos expériences. Nous avons vu dans la section 1.4.3 que de nombreux opérateurs de convolution sur graphe existaient et que la nature de l’opérateur choisi pouvait modifier les résultats du GNN, notamment s’il considère ou non les attributs d’arêtes multi-dimensionnels. Afin d’étudier l’influence de cet opérateur sur les performances de notre méthode, nous considérons également l’opérateur de convolution comme un hyperparamètre (plusieurs opérateurs seront comparés). Enfin, notons que notre architecture ne présente qu’une seule couche de convolution bien que plusieurs pourraient être mises en cascade (Figure 1.9). Cependant, les GNNs obtiennent de meilleures performances quand ils sont peu profonds (moins de 4 couches de convolution selon [93]).

La deuxième couche du réseau est un perceptron à couche unique ($(\mathbb{R}^{d^{+1}} \rightarrow \mathbb{R}^C)$) fournissant à chaque noeud $v \in V$ un vecteur de probabilité d’appartenance à chaque classe $c \in C$.

Les paramètres du réseau sont optimisés, à partir des données d’entraînement, pour maximiser le taux de bonne classification de noeuds. Cette optimisation correspond à une minimisation de la fonction de coût du logarithme négatif de la vraisemblance :

$$\text{Loss}(Y, \hat{y}) = - \sum_{v=1}^{|V|} \sum_{c=1}^C Y_{v,c} \times \log(\hat{y}_{v,c}) \quad (2.2)$$

où $Y_{v,c}$ indique si c est la classe réelle du noeud v (1 si v appartient à la classe c , 0 sinon) et $\hat{y}_{v,c}$ est la probabilité en sortie du GNN que le noeud v appartienne à la classe c . Pour chaque noeud v , sa classe finale y_v attribuée est définie comme : $y_v = \text{argmax}_c \hat{y}_{v,c}$. A partir de la classification de noeuds proposée par le GNN, les pixels (voxels) de chaque région R_v sont fixés à la valeur de la classe assignée au noeud v correspondant. La carte de segmentation \mathcal{L}^* est alors mise à jour.

2.4 Expériences

Pour nos expériences, nous nous sommes basés sur deux applications de segmentation sémantique. La première, relativement simple pour tester la méthode, consiste à segmenter des images 2D en couleur de visages humains sur la base de données publique FASSEG¹. La deuxième application repose sur la segmentation de structures cérébrales, dans une anticipation des travaux menés dans la partie II de cette thèse. Pour cela, nous avons utilisé le jeu de données public IBSR² constitué d'IRMs en 3D de cerveaux adultes humains. Dans la section 2.4.1, nous décrivons le protocole expérimental considéré. Nous détaillerons ensuite, dans les sections 2.4.2 et 2.4.3, les expériences menées sur les jeux de données FASSEG et IBSR respectivement. Dans chaque cas, nous présenterons la base de données et les réseaux de neurones considérés pour la segmentation initiale, puis la construction des graphes à partir des images avant de détailler les résultats de segmentation obtenus. Enfin, dans la section 2.4.4, nous discuterons de l'influence de la topologie des graphes sur les résultats, et plus précisément du rayon d'élagage considéré.

2.4.1 Protocole expérimental

Nos expériences ont été menées dans un environnement Python en utilisant la librairie *PyTorch Geometric* [94] sur un GPU *Nvidia Quadro RTX 3000*. Ce sera le cas pour toutes les expériences présentées dans cette thèse.

Pour les deux applications, l'architecture GNN est entraînée avec l'algorithme d'optimisation Adam, sur 600 époques, avec un *dropout* de 0.5 afin de minimiser la fonction de

1. <https://github.com/massimomauro/FASSEG-repository>

2. <https://www.nitrc.org/projects/ibsr>

coût de l’eq 2.2. Une stratégie de réduction du taux d’apprentissage sur plateau est utilisée avec un taux d’apprentissage initial $\Xi_0 = 0.01$ et un facteur de réduction $\sigma = 5e^{-4}$. La dimension d^{l+1} des attributs de noeuds en sortie de l’opérateur de convolution est fixée à 7 pour FASSEG et à 22 pour IBSR.

Dans le but d’étudier l’impact de l’opérateur de convolution employé, 4 opérateurs sont comparés. Nous considérons deux opérateurs maniant uniquement un poids scalaire sur les arêtes : GCNConv [63] et GraphConv [70] (présenté dans la section 1.4.3, eq 1.6). L’opérateur GCNConv, reconnu pour de la classification de noeud semi-supervisée (et non, supervisée comme dans notre cas) actualise la représentation du noeud $v \in V$ de la façon suivante [63] :

$$X^{l+1}(v) = W^{l+1} \sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{w_{u,v}}{\sqrt{\hat{d}_u \hat{d}_v}} X^l(u) \quad (2.3)$$

où W^{l+1} est une matrice de poids entraînable et $\hat{d}_v = 1 + \sum_{u \in \mathcal{N}(v)} w_{u,v}$ avec $w_{u,v}$ étant le poids scalaire sur l’arête reliant les noeuds u et v . Afin d’observer si l’inconsidération des attributs d’arêtes influe sur les performances, nous fixons le poids scalaire ($w_{u,v}$ dans les eq 1.6 et 2.3) à 1. Les deux autres opérateurs de convolution considérés dans nos expériences manient les informations structurelles multi-dimensionnelles portées par les arêtes : ECConv [68] (présenté dans la section 1.4.3, eq 1.7) et GATConv [71]. Avec ce dernier, nous souhaitons voir si l’ajout de mécanismes d’attention dans l’opérateur est bénéfique pour ce type d’application. En effet, le *message passing* avec l’opérateur GATConv est défini comme dans [71] par :

$$X^{l+1}(v) = \sum_{u \in \mathcal{N}(v) \cup \{v\}} \alpha_{v,u}^{l+1} W^{l+1} X^l(u) \quad (2.4)$$

$\alpha_{v,u}^{l+1}$ étant les coefficients d’attention calculés comme :

$$\alpha_{v,u}^{l+1} = \frac{e^{a(W^{l+1}X^l(v), W^{l+1}X^l(u), W_e^{l+1}L((u,v)))}}{\sum_{k \in \mathcal{N}(v) \cup \{v\}} e^{a(W^{l+1}X^l(v), W^{l+1}X^l(k), W_e^{l+1}L((k,v)))}} \quad (2.5)$$

a désignant le mécanisme d’attention et W_e^{l+1} étant une matrice entraînable de poids spécifique aux attributs d’arêtes.

Pour évaluer la qualité des segmentations obtenues, nous calculons pour chaque classe présente deux métriques : le coefficient de Dice (DSC) [95] et la distance d’Hausdorff (HD) [96]. DSC $\in [0, 1]$ mesure la similarité entre l’image segmentée (I') et l’image annotée (I),

appelée aussi vérité terrain (1 indique une équivalence parfaite). HD mesure la distance maximale entre deux ensembles I et I' , mettant donc plutôt en évidence les composantes éloignées de la région cible (artefacts). Ces métriques, qui seront considérées à plusieurs reprises dans cette thèse, sont détaillées dans les eq 2.6 et 2.7.

$$\text{DSC}(I, I') = \frac{2|I \cap I'|}{|I| + |I'|} \quad (2.6)$$

$$\text{HD}(I, I') = \max(\max_{a \in I} \min_{b \in I'} |b - a|, \max_{b \in I'} \min_{a \in I} |a - b|) \quad (2.7)$$

Pour analyser l'apport de notre méthode, nous comparons les résultats de segmentation obtenus en sortie de CNN seul et ceux obtenus avec notre post-traitement basé sur GNN.

Par ailleurs, pour évaluer la robustesse de notre méthode à des jeux de données plus petits (cela étant fréquent dans le domaine médical qui nous intéressera par la suite), nous considérons des bases de données d'entraînement (c.a.d. images utilisées pour entraîner le CNN et le GNN) de différentes tailles correspondant aux configurations A, B et C présentées dans la Table 2.1. L'ensemble d'images utilisé pour tester la méthode est toujours le même (50 images pour FASSEG, 6 pour IBSR).

TABLE 2.1 – Taille et distribution des jeux de données utilisés dans chacune des applications.

Configuration	Nombre d'images	entraînement (validation)			test
		A	B	C	
FASSEG	70	15 (5)	10 (3)	5 (2)	50
IBSR	18	8 (4)	6 (3)	4 (2)	6

Pour chaque configuration, les résultats sont moyennés sur 3 tirages aléatoires des données d'entraînement/validation, donnant lieu à 3 réseaux de neurones entraînés.

Enfin, pour montrer que notre approche est généralisable à tout réseau de neurones fournissant une carte de segmentation, nous considérons dans nos expériences plusieurs CNNs (détaillés par la suite). Ces derniers ont été sélectionnés parmi des travaux récents, avec une implémentation disponible en ligne, spécialisés pour des images en couleur 2D ou des images médicales 3D.

2.4.2 FASSEG

2.4.2.1 Jeu de données et architectures CNNs

Le jeu de données public FASSEG fournit des images 2D en couleur de visages humains dans différentes poses pour la segmentation sémantique multi-classes [97]. Dans nos expériences, afin d’évaluer la méthode dans un premier temps sur un problème simple, nous considérons un sous-ensemble de cette base de données correspondant à une pose de face. Ce sous-ensemble contient 70 images segmentées manuellement. Nous avons utilisé une variante de cette base avec des annotations affinées, afin d’avoir une région unique par classe, conduisant au jeu de données *FASSEG-Instances*³. Par simplicité, ce nouveau jeu de données sera nommé FASSEG par la suite. Les nouvelles annotations considèrent neuf classes : le fond, les cheveux, le visage, les yeux gauche et droit, les sourcils gauche et droit, le nez et la bouche.

Pour ce jeu de données, nous avons considéré 4 réseaux de neurones profonds : un U-Net (présenté section 1.2.3), un U-Net combiné à des champs aléatoires conditionnels (CRF) [98], un EfficientNet [99] et un PSPNet [100]. Concernant l’architecture U-Net, nous avons utilisé une implémentation PyTorch⁴, entraînée sur 200 époques avec une stratégie d’arrêt prématuré pour éviter le sur-entraînement. La régularisation spatiale basée sur CRF a été placée en sortie du U-Net pré-entraîné. Nous avons appliqué un filtre Gaussien de taille 11 et avons entraîné le modèle CRF sur 50 époques pour affiner le modèle d’ensemble. Le modèle EfficientNet s’appuie sur une méthode de mise à l’échelle composée (longueur, largeur et résolution) pour améliorer la précision des CNNs. Nous avons choisi EfficientNet-B3 (pre-entraîné sur ImageNet), avec les hyperparamètres de [99] car cette implémentation utilise un U-Net pour le décodage et sera donc comparable avec les réseaux précédents. L’architecture PSP-Net a été considérée car, grâce à son module pyramidal, elle est capable de capturer des informations, notamment spatiales, à différentes échelles. Nous souhaitons voir l’apport de notre méthode sur un réseau capturant déjà, en partie, de l’information spatiale globale. Pour ce réseau, nous avons utilisé les hyperparamètres de [100], c’est-à-dire, un ResNet pré-entraîné pour le calcul initial de la carte de caractéristiques et une mise en commun des informations avec un module pyramidal à 4 niveaux. Quel que soit le CNN, un filtre médian est appliqué en sortie pour éliminer les petits artefacts de la carte de segmentation fournie.

3. <https://github.com/Jeremy-Chopin/FASSEG-instances>

4. <https://doi.org/10.5281/zenodo.3522306>

2.4.2.2 Construction des graphes

Un graphe G_ρ est construit comme présenté dans la section 2.3.1. Pour extraire les composantes connexes de la carte de segmentation du CNN, une connectivité de 8 est considérée. Face à la concavité de certaines régions, notamment les cheveux dont le barycentre peut se trouver au milieu du visage (c.a.d. en-dehors des cheveux), nous avons considéré comme relations spatiales (attributs d'arêtes) les distances minimales ($d_{min}^{R_v, R_u} = \min_{a \in R_v, b \in R_u} |a - b|$) et maximales ($d_{max}^{R_v, R_u} = \max_{a \in R_v, b \in R_u} |a - b|$) entre régions R_v et R_u , normalisées par la plus grande distance D observée dans l'image (la diagonale de l'image) :

$$L((u, v)) = \left[\frac{d_{min}^{R_v, R_u}}{D}, \frac{d_{max}^{R_v, R_u}}{D} \right] \quad (2.8)$$

Ainsi, chaque noeud a un attribut de dimension 9 puisque 9 classes sont présentes ($X : V \rightarrow \mathbb{R}^9$) et les arêtes portent des informations de dimension 2 ($L : E \rightarrow \mathbb{R}^2$).

2.4.2.3 Résultats

TABLE 2.2 – FASSEG - nombre de régions et donc de noeuds fournis par les différentes architectures de CNNs (moyenne sur les 3 tirages aléatoires de la base d'apprentissage) selon la configuration d'entraînement considérée.

	A	B	C
U-Net	12.0	13.2	13.9
U-Net + CRF	12.1	12.7	14.7
EfficientNet	11.3	11.7	14.6
PSPNet	11.9	12.2	17.5

La Table 2.2 rapporte les tailles de graphes impliquées pour les différentes architectures de CNNs et configurations d'entraînement. Nous sommes confrontés à des graphes de taille arbitraire. Par ailleurs, le nombre de noeuds augmente avec la réduction de la taille de la base de données d'entraînement. En effet, le réseau, moins efficace, fournit plus de régions candidates par classe.

La Table 2.3 compare les performances de segmentation avec et sans la méthode proposée sur FASSEG. Les résultats obtenus avec les 4 différents opérateurs de convolution sont rapportés. On observe que les performances du CNN se dégradent avec la réduction du nombre de données d'entraînement (troisième ligne de la figure 2.2 avec de nombreuses erreurs de segmentation). Néanmoins, quels que soient le CNN et la configuration

TABLE 2.3 – Comparaison des performances de segmentation sur FASSEG avec CNN seul et avec la combinaison proposée de CNN et GNN (les résultats avec les 4 opérateurs de convolution sur graphe sont indiqués : EC pour ECConv, GCN pour GCNConv, Graph pour GraphConv et GAT pour GATConv). DSC : Dice, HD : distance d’Hausdorff.

CNN	Configuration A									
	DSC					HD				
	CNN	EC	GCN	Graph	GAT	CNN	EC	GCN	Graph	GAT
U-Net	0.70	0.70	0.70	0.70	0.70	37.40	30.79	28.80	29.88	28.78
U-Net+CRF	0.71	0.71	0.71	0.71	0.71	35.39	27.29	26.24	27.96	26.24
EfficientNet	0.84	0.84	0.82	0.84	0.84	22.10	19.01	20.05	18.29	16.75
PSPNet	0.83	0.83	0.83	0.83	0.83	24.18	20.16	19.62	19.49	18.30
Moyenne	0.77	0.77	0.76	0.77	0.77	29.76	24.31	23.68	23.90	22.52
CNN	Configuration B									
	DSC					HD				
	CNN	EC	GCN	Graph	GAT	CNN	EC	GCN	Graph	GAT
U-Net	0.72	0.73	0.71	0.72	0.72	43.02	29.23	32.87	30.71	30.72
U-Net+CRF	0.70	0.70	0.68	0.70	0.70	39.88	30.92	36.53	31.00	30.60
EfficientNet	0.83	0.83	0.82	0.83	0.83	26.39	19.93	22.19	21.11	18.77
PSPNet	0.82	0.82	0.81	0.83	0.82	28.05	20.86	21.17	19.94	20.16
Moyenne	0.77	0.77	0.75	0.77	0.77	34.33	25.23	28.19	25.69	25.06
CNN	Configuration C									
	DSC					HD				
	CNN	EC	GCN	Graph	GAT	CNN	EC	GCN	Graph	GAT
U-Net	0.71	0.71	0.69	0.71	0.71	54.25	41.12	42.98	41.57	36.61
U-Net+CRF	0.69	0.68	0.64	0.69	0.67	60.00	45.26	55.44	42.71	41.88
EfficientNet	0.81	0.81	0.74	0.81	0.81	49.80	28.54	43.30	27.90	26.39
PSPNet	0.75	0.78	0.72	0.76	0.76	81.51	40.66	55.47	48.52	45.77
Moyenne	0.74	0.75	0.70	0.74	0.74	61.39	38.90	49.30	40.17	37.66

d’entraînement, notre proposition améliore les résultats (pour la plupart des opérateurs de convolution sur graphe) tout en ajoutant un nombre négligeable de paramètres par rapport à ceux du CNN (Tableau 2.4). Notons que l’architecture EfficientNet est plus complexe que celle du traditionnel U-Net, ce qui conduit à un meilleur Dice au prix d’un plus grand nombre de paramètres entraînaibles. PSPNet semble plus efficace que U-Net et similaire à EfficientNet, mais avec moins de paramètres, ce qui souligne l’importance de capturer différents éléments d’information à différentes échelles.

L’amélioration avec le post-traitement basé sur GNN est particulièrement significative en termes de distance d’Hausdorff (jusqu’à 47% selon la Table 2.3 - EfficientNet(C)) mais négligeable en terme de Dice. Cela signifie que les erreurs de segmentation consistent en de petites régions relativement éloignées de la région cible (modifiant considérablement le

TABLE 2.4 – Nombre de paramètres entraîables pour chaque architecture (CNNs et GNNs) considérée pour FASSEG. Pour les architectures GNN, les paramètres incluent également le perceptron à couche unique.

Architecture CNN	# paramètres	Architecture GNN	# paramètres
U-Net	7 865 529	ECCConv	387
U-Net + CRF	7 865 532	GCNConv	142
EfficientNet	17 779 713	GraphConv	241
PSPNet	3 849 420	GATConv	184

HD). Le bénéfice de la méthode est illustré sur la Figure 2.2 où l’on observe que le GNN élimine la majorité des artefacts et corrige les erreurs de classification du CNN. Les bonnes performances sur les configurations B et C démontrent la robustesse de la proposition face au manque de représentativité. Notre méthode, combinée à PSPNet, s’avère même plus performante avec un petit jeu de données (configuration B) qu’un réseau de neurones plus complexe comme EfficientNet avec plus de données (configuration A) (HD de 19.94 vs 22.10 selon le Tableau 2.3).

Concernant l’opérateur de convolution sur graphe utilisé, il semble que les 4 opérateurs considérés améliore la segmentation en terme de distance d’Hausdorff. Néanmoins, le GNN basé sur GCNConv est moins efficace et a tendance à dégrader le Dice. La première ligne de la Figure 2.2 illustre certaines erreurs pouvant être commises par cet opérateur. Cela peut s’expliquer par l’inconsidération des attributs d’arêtes et par un nombre trop faible de paramètres entraîables (Table 2.4) pour apprendre toutes les variations entre les visages.

Les performances des trois autres opérateurs sont équivalentes, bien que GATConv fournisse, en moyenne, le plus petit HD (deux premières lignes de la Figure 2.2) malgré son petit nombre de paramètres entraîables. Cela peut s’expliquer par les mécanismes d’attention mis en oeuvre. La plus grande complexité d’ECCConv (Table 2.4) lui permet de mieux corriger les incohérences spatiales lorsque le nombre de régions est plus élevé (Configuration C - PSPNet selon Table 2.2) comme l’illustre la Figure 2.2-C. De manière générale, lorsque le nombre de noeuds augmente (Configuration C), les opérateurs de convolution considérant à la fois les attributs sur noeuds et sur arêtes (c.a.d. ECCConv et GATConv) semblent plus adaptés pour corriger les erreurs de segmentation (Table 2.3).

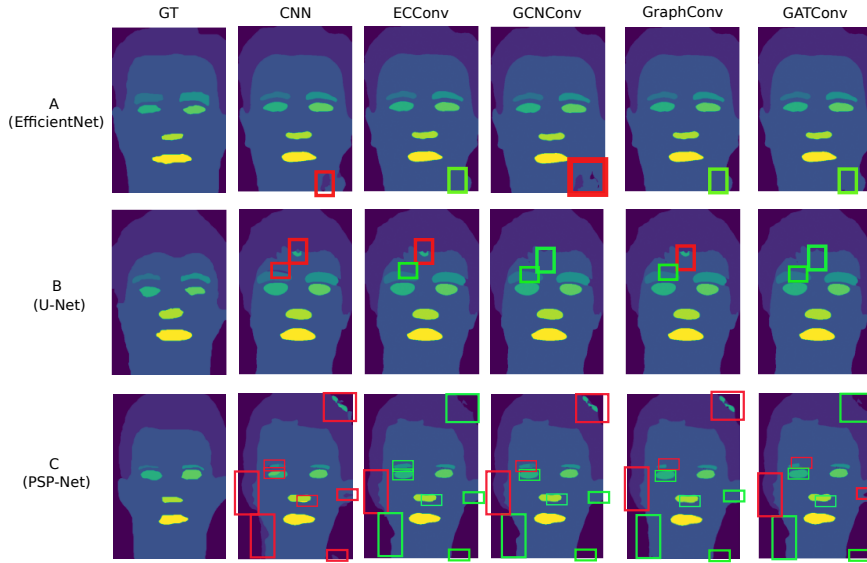


FIGURE 2.2 – Exemples de segmentation sur FASSEG avec et sans GNN. Une comparaison des 4 opérateurs de convolution est fournie. Les boîtes englobantes montrent les régions avec des améliorations significatives. A, B et C correspondent aux configurations d'entraînement considérées. GT : vérité terrain.

2.4.3 IBSR

2.4.3.1 Jeu de données et architectures CNNs

Le jeu de données public IBSR est composé de 18 IRMs (volumes 3D) de cerveaux adultes humains associées à la segmentation de 32 régions cérébrales. Dans nos expériences, comme dans [101], seulement 15 classes du jeu de données annoté sont considérées : le fond, le thalamus, le noyau caudé, le putamen, le pallidum, l'hippocampe, l'amygdale et le noyau accumbens (gauche et droit pour toutes les structures).

Pour cette application, nous avons considéré trois réseaux de neurones profonds : un U-Net 3D [102] semblable à l'architecture présentée dans la section 1.2.3, un U-Net 3D combiné à des CRF et le récent UNETr [21], *Transformer* s'appuyant également sur une architecture en "U" avec des couches convolutives. La spécificité de ce jeu de données repose sur la taille des images médicales et sur le déséquilibre de représentativité entre classes (certaines classes cibles sont de très petites tailles par rapport à d'autres). Cela nécessite des stratégies spécifiques telles que l'utilisation de patches, qui sera davantage détaillée dans le chapitre 3 de cette thèse.

Pour le réseau U-Net 3D, nous avons considéré l'implémentation générique 2D/3D utilisée pour FASSEG, entraînée sur 60 époques avec une stratégie d'arrêt prématuré pour

éviter le sur-entraînement. Nous avons, pour les raisons mentionnées précédemment, employé une approche par patches 3D [103]. Des patches de dimension 32^3 voxels ont été extraits autour du barycentre de chaque structure d'intérêt (tirage aléatoire) avec la librairie *Torchio* [104]. Pour chaque IRM, 64 patches ont été sélectionnés avec une fréquence proportionnelle à la probabilité inverse de la classe correspondante (c.a.d. davantage de patches pour les classes les plus petites). Concernant le post-traitement par CRF, nous avons considéré les mêmes paramètres que pour FASSEG (section 2.4.2). Le modèle CRF a, cette fois, été entraîné sur 40 époques à partir du U-Net préentraîné.

Le réseau UNETr est un *Visual Transformer*, comme présenté dans la section 1.2.3, appliqué à la segmentation d'images médicales 3D. Nous avons vu que de tels modèles capturent le contexte global grâce à leurs mécanismes d'attention et les dépendances spatiales à grande échelle mises en oeuvre. Là encore, nous avons choisi un tel réseau pour analyser la capacité de notre méthode à améliorer les performances de modèles englobant déjà, en partie, des informations structurelles globales. Les hyperparamètres du UNETr ont été fixés comme dans [21] avec des patches de dimension 16^3 pour la décomposition de l'image avant projection linéaire.

2.4.3.2 Construction des graphes

Un graphe G_ρ est construit comme détaillé dans la section 2.3.1. Pour extraire les composantes connexes de la carte de segmentation du CNN, une connectivité de 26 est considérée. Comme dans les travaux de [76], les arêtes portent les distances absolues entre les barycentres des régions connectées R_v et R_u . Ces distances sont découpées selon les 3 dimensions de l'image (x, y, z) :

$$L((u, v)) = [d_x^{R_v, R_u}, d_y^{R_v, R_u}, d_z^{R_v, R_u}] \quad (2.9)$$

Ainsi, chaque noeud a un attribut de dimension 15 correspondant aux 15 classes considérées ($X : V \rightarrow \mathfrak{R}^{15}$) et les arêtes portent des informations de dimension 3 ($L : E \rightarrow \mathfrak{R}^3$).

2.4.3.3 Résultats

La Table 2.5 indique la taille des graphes impliqués pour les différentes architectures de réseaux de neurones et les différentes configurations des données d'entraînement. Comme pour FASSEG, nous sommes confrontés à des graphes de taille arbitraire et le nombre de noeuds augmente lorsque la taille de la base d'apprentissage diminue.

TABLE 2.5 – IBSR - Nombre de régions (c.a.d. de noeuds) fournies par les différentes architectures de CNNs (moyenne sur les 3 tirages aléatoires de la base d’entraînement) et selon la configuration d’entraînement considérée.

	A	B	C
U-Net	20.6	23.9	23.6
U-Net + CRF	22.4	22.8	28.3
UNETr	15.4	15.6	21.3

TABLE 2.6 – Comparaison des performances de segmentation sur IBSR avec CNN seul et avec la combinaison proposée de CNN et GNN (les résultats avec les 4 opérateurs de convolution sur graphe sont reportés : EC pour ECCConv, GCN pour GCNConv, Graph pour GraphConv et GAT pour GATConv). DSC : Dice, HD : distance d’Hausdorff.

CNN	Configuration A									
	DSC					HD				
	CNN	EC	GCN	Graph	GAT	CNN	EC	GCN	Graph	GAT
U-Net	0.80	0.81	0.64	0.81	0.79	22.51	5.73	57.36	5.74	8.83
U-Net+CRF	0.81	0.82	0.74	0.82	0.80	25.77	6.60	29.06	4.40	11.74
UNETr	0.75	0.76	0.57	0.76	0.73	11.01	5.27	62.95	5.27	13.55
Moyenne	0.79	0.80	0.65	0.80	0.77	19.76	5.87	49.79	5.14	11.37
CNN	Configuration B									
	DSC					HD				
	CNN	EC	GCN	Graph	GAT	CNN	EC	GCN	Graph	GAT
U-Net	0.79	0.81	0.70	0.81	0.76	22.99	4.72	41.08	4.72	17.08
U-Net+CRF	0.79	0.81	0.74	0.81	0.77	23.49	5.69	22.84	4.53	18.88
UNETr	0.73	0.73	0.58	0.73	0.73	13.77	5.82	59.75	6.03	7.19
Moyenne	0.77	0.78	0.67	0.78	0.75	20.08	5.41	41.22	5.09	14.38
CNN	Configuration C									
	DSC					HD				
	CNN	EC	GCN	Graph	GAT	CNN	EC	GCN	Graph	GAT
U-Net	0.76	0.78	0.70	0.78	0.71	26.94	6.76	26.88	5.04	28.78
U-Net+CRF	0.77	0.79	0.72	0.80	0.69	28.50	8.47	24.59	5.15	33.33
UNETr	0.67	0.67	0.47	0.67	0.62	25.75	9.69	76.89	12.26	20.65
Moyenne	0.73	0.75	0.63	0.75	0.67	27.06	8.31	42.79	7.49	27.59

La Table 2.6 compare les performances de segmentation avec et sans la méthode proposée sur IBSR. Une fois de plus, les performances de segmentation par le CNN se dégradent avec la diminution du nombre de données d’entraînement. On peut noter que pour la configuration C utilisant UNETr, les résultats sont beaucoup plus faibles que ceux obtenus en considérant un jeu de données plus grand ou un U-Net classique (Dice de 0.67 contre 0.77 avec U-Net+CRF). Ce mauvais résultat illustre le fait qu’un tel réseau, basé sur un *Trans-*

former, nécessite davantage de données d’entraînement que les autres CNNs traditionnels en raison de son très grand nombre de paramètres entraînaibles (Table 2.7). Cependant, pour les configurations A et B, le réseau UNETr, avec son mécanisme d’attention considérant l’information spatiale à plus grande échelle pour guider la segmentation, donne de meilleurs résultats en termes de distance d’Hausdorff que le U-Net classique (HD divisé par 2). Ceci est également illustré dans la Figure 2.3 (détaillée ci-après) où nous observons que U-Net produit plus d’artefacts (noeuds environnants - 1e ligne) que UNETr (2e ligne).

TABLE 2.7 – Nombre de paramètres entraînaibles pour chaque architecture CNN utilisée sur IBSR.

Arch. CNN	# paramètres
U-Net	15 711 887
U-Net + CRF	15 711 891
UNETr	94 197 951

TABLE 2.8 – Nombre de paramètres entraînaibles pour chaque architecture GNN utilisée sur IBSR (les paramètres incluent le perceptron à couche unique).

Arch. GNN	# paramètres
ECCConv	1995
GCNConv	697
GraphConv	1027
GATConv	829

Quelle que soit la configuration d’entraînement et le CNN, il existe au moins un opérateur de convolution sur graphe améliorant la segmentation du CNN. Comme pour FASSEG, l’amélioration apparaît significative en termes de distance d’Hausdorff avec une réduction moyenne d’environ 75% avec toutes les configurations, selon la Table 2.6 (ex. de 20.08 à 5.09 avec la configuration B). Cela souligne la pertinence de notre approche pour corriger les erreurs structurelles de segmentation. En effet, même avec un *Visual Transformer* comme UNETr, fournissant moins d’artefacts grâce à son mécanisme d’attention (Figure 2.3), notre proposition réduit significativement le HD de 11.01 à 5.27 (réduction de 52%) pour la configuration d’entraînement A, de 13.77 à 5.82 (réduction de 58%) pour la configuration d’entraînement B et de 25.75 à 9.69 (réduction de 62%) pour la configuration d’entraînement C. Cela illustre également la capacité de notre méthode à compenser de manière significative le manque de représentativité lorsque la base de données d’entraînement est petite : les relations apprises par le GNN pendant l’entraînement permettent de corriger les erreurs structurelles réalisées par un CNN. De cette manière, notre approche conduit à de meilleurs résultats avec peu de données qu’un réseau de neurones profond complexe comme UNETr entraîné avec plus de données, ceci constituant un défi [81]. Ainsi, notre proposition combinée à un U-Net, ajoutant au maximum 1995 paramètres (ECCConv), atteint de meilleures performances avec un petit jeu de données (configuration C) que UNETr seul entraîné avec plus de données (configuration A) (HD de

5.04 contre 11.01, Dice de 0.78 contre 0.75 selon la Table 2.6). Ceci illustre la pertinence de considérer des relations de haut niveau en post-traitement d’une carte de segmentation fournie par un CNN, non seulement en qualité de segmentation mais aussi en termes de nombre de paramètres entraînaables.

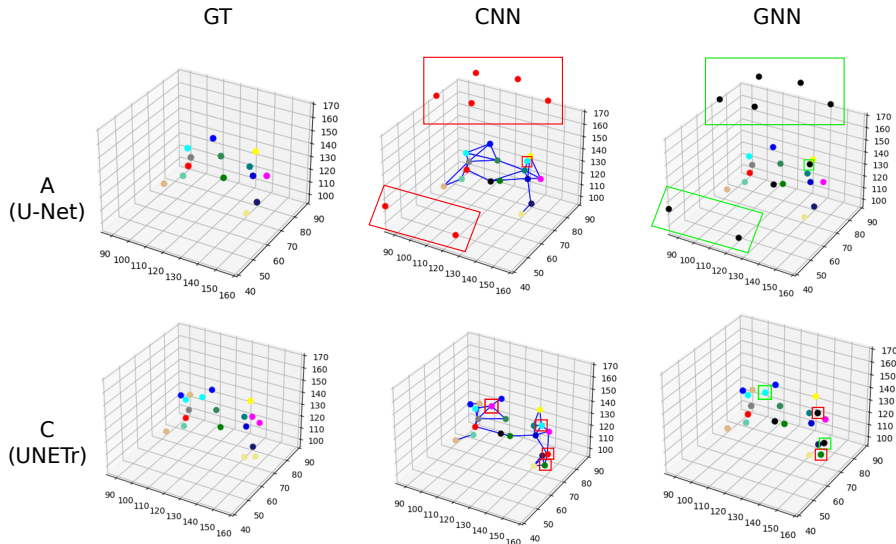


FIGURE 2.3 – Noeuds des graphes construits à partir des segmentations sur IBSR (position 3D des barycentres des régions segmentées). De gauche à droite : classe réelle des noeuds (vérité terrain - GT), classes assignées à la sortie du CNN et arêtes prises en compte par le GNN (élagage du graphe), classes assignées à la sortie du GNN. Deux exemples de deux configurations sont donnés. Chaque couleur correspond à une structure cérébrale (noir étant le fond). A et C correspondent aux configurations d’entraînement.

Pour cette application, ECCConv et GraphConv semblent être les meilleurs opérateurs, capable d’améliorer significativement les segmentations des CNNs dans toutes les configurations (en termes de Dice et de distance d’Hausdorff). Cela peut s’expliquer par leur plus grand nombre de paramètres entraînaables (Table 2.8), leur donnant davantage de degrés de liberté. L’efficacité de GraphConv montre qu’il n’est pas indispensable, dans ce cas (ce qui n’était pas vrai pour FASSEG), de considérer des attributs d’arêtes multidimensionnels. On peut penser que, pour cette application, la discrimination entre noeuds apportée par l’élagage d’arêtes (voir section 2.4.4) est suffisante. Notons que, comparé à un travail récent sur le jeu de données IBSR [101] mais impliquant un atlas, notre approche conduit à des résultats légèrement moins bons en Dice (0.82 contre 0.84 [101]) mais légèrement meilleurs en HD (4.40 contre 4.49 in [101]).

Sur la Figure 2.3, on visualise mieux les corrections fournies par le GNN. Ainsi, dans

le premier exemple (1e ligne), les artefacts produits par U-Net (entourés en rouge) sont correctement classifiés par le GNN comme appartenant au fond. On peut imaginer que c'est l'isolement de ces noeuds (reliés à aucun autre noeud du fait de l'élagage) qui conduit à les considérer comme du fond. Dans le deuxième exemple, le UNETr a fait quelques erreurs de classification (encadrées en rouge). Le GNN élimine (c.a.d. classe comme fond) la mauvaise structure rouge à droite et corrige le noeud à gauche classifié comme rose en cyan. Le GNN détecte également l'erreur liée au noeud cyan détecté à droite (incohérence spatiale car la classe cyan est de l'autre côté) mais il se trompe en le classifiant comme fond (noir) au lieu de rose. Cela illustre la capacité de notre méthode à détecter les anomalies structurelles et à les corriger la plupart du temps, même si des erreurs subsistent dans certains cas (ex. le noeud vert en bas à droite).

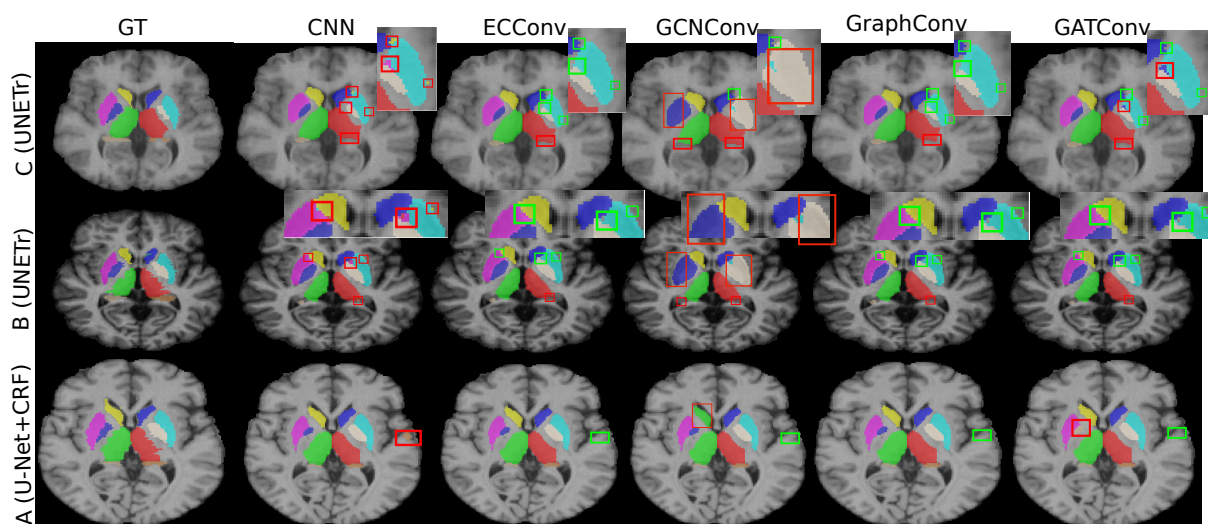


FIGURE 2.4 – Exemples de segmentation sur IBSR avec et sans GNN. Une comparaison des 4 opérateurs de convolution est fournie. Les boîtes englobantes montrent les régions avec des améliorations significatives. A, B et C correspondent aux configurations d'entraînement. GT : vérité terrain.

Sur la Figure 2.4-ligne 3, nous voyons que tous les opérateurs parviennent à corriger les artefacts fournis par U-Net avec CRF en configuration A. ECCConv et GraphConv parviennent à corriger les mauvaises classifications faites par le CNN dans quasiment tous les cas. Néanmoins, sur les deux premières lignes, nous pouvons voir que lorsqu'une classe n'est pas détectée par le CNN (par exemple l'hippocampe gauche encadré en rouge), le GNN, logiquement, ne parvient pas à récupérer cette structure manquante (boîte rouge restante avec ECCConv et GraphConv). Contrairement à l'application sur FASSEG, GCN-

Conv et GATConv semblent avoir plus de difficultés à classer correctement les noeuds. Cela peut peut-être s’expliquer par la plus grande taille des graphes concernés (plus de 20 noeuds en moyenne selon la Table 2.5 contre moins de 15 pour FASSEG selon la Table 2.2), par le plus grand nombre de classes étudiées (dimension plus grande sur les attributs de noeuds) et par leur plus petit nombre de paramètres entraînaables (Table 2.8). En effet, ils ont tendance à perdre des informations, ce qui conduit à des incohérences spatiales voire à la disparition totale de classes. Par exemple, sur la Figure 2.4, GCNConv remplace le putamen par le pallidum (lignes 1 et 2) et GATConv remplace le pallidum par l’hippocampe (ligne 3), ce qui dégrade fortement les résultats (Table 2.6).

2.4.4 Influence de la topologie des graphes

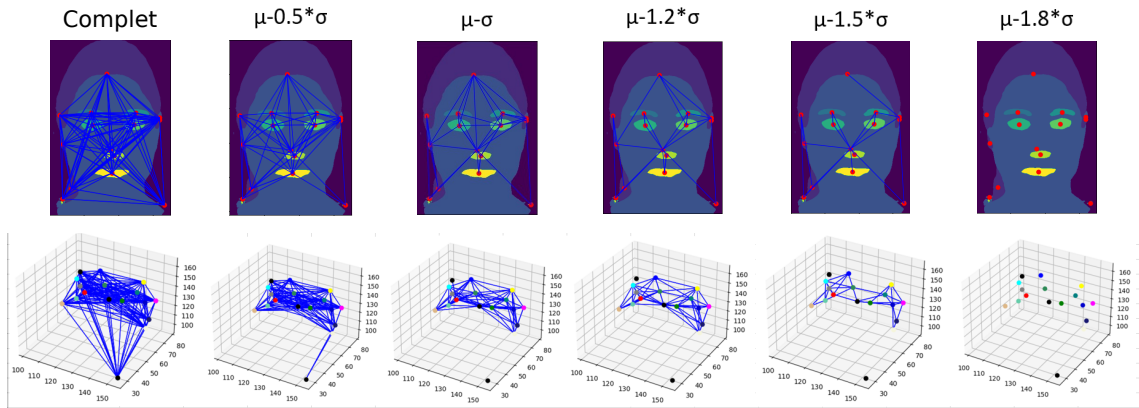


FIGURE 2.5 – Évolution du nombre d’arêtes en fonction du rayon d’élague. La première ligne montre un exemple sur FASSEG (EfficientNet-C) et la deuxième ligne sur IBSR (U-Net-C).

Comme nous l’avons vu dans la section 1.4.3, la topologie des graphes mis en jeu en entrée de GNN semble jouer un rôle important à en croire les divers travaux développés ces dernières années pour élaguer les graphes. L’objectif de cette section est d’étudier l’impact du nombre d’arêtes dans les graphes (ou plutôt du niveau d’élague) sur les performances de segmentation. Pour cela, nous proposons une méthode d’élague très simple basée sur un rayon d’élague définissant le voisinage de chaque noeud comme expliqué dans la section 2.3.1. Pour chaque configuration étudiée, six rayons d’élague ρ ont été comparés durant la phase d’optimisation sur les données d’entraînement : graphes complets (rayon infini), $\mu - 0.5\sigma$, $\mu - \sigma$, $\mu - 1.2\sigma$, $\mu - 1.5\sigma$ et $\mu - 1.8\sigma$. Dans chaque cas, μ est la valeur moyenne de la relation $L((u, v))$ entre toutes les régions d’une image, moyen-

née sur l'ensemble des images d'entraînement et σ est l'écart-type. Le plus grand rayon d'élagage ($\mu - 1.8\sigma$) conduit la plupart du temps à des graphes sans arêtes. Notons que l'implémentation *torch-geometric* considérée pour la convolution ne gère pas les situations de graphes sans arêtes. Par conséquent, dans ce cas, nous considérons que le GNN est réduit au perceptron à couche unique.

Ainsi, comme illustré sur la Figure 2.5, avec l'instauration d'un rayon d'élagage basé sur les relations spatiales entre régions, les noeuds correspondant aux régions les plus éloignées entre elles finissent par ne plus être connectés. En fonction du rayon d'élagage, le nombre d'arêtes dans les graphes varie drastiquement (Figures 2.5 et 2.6).

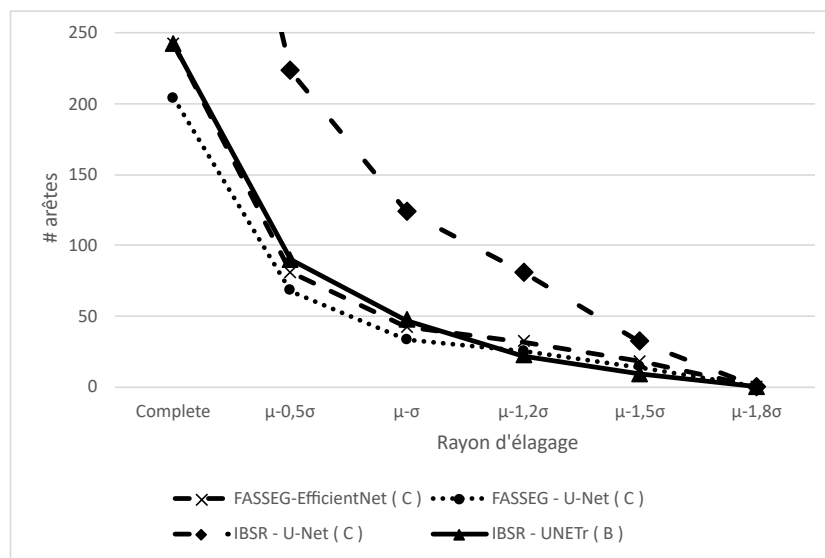


FIGURE 2.6 – Nombre moyen d'arêtes sur les données de test en fonction du rayon d'élagage considéré. Deux configurations sur FASSEG (EfficientNet-C, U-Net-C) et deux sur IBSR (U-Net-C, UNETr-B) sont illustrées.

La Figure 2.6 représente le nombre moyen d'arêtes sur les données de test en fonction du rayon d'élagage considéré. Pour plus de clarté, seuls les résultats sur 4 configurations sont reportés : deux avec FASSEG (EfficientNet et U-Net avec configuration d'entraînement C) et deux sur IBSR (U-Net (C) et UNETr (B)). On voit que le nombre d'arêtes évolue de manière similaire pour toutes les configurations. Typiquement, un rayon de $\mu - 1.5\sigma$ divise par environ 20 le nombre d'arêtes par rapport à l'utilisation de graphes complets. Le plus grand nombre d'arêtes avec U-Net (C) sur IBSR peut s'expliquer par le nombre plus élevé de noeuds présents (Table 2.5).

La Figure 2.7 représente l'évolution de la distance d'Hausdorff des segmentations sur

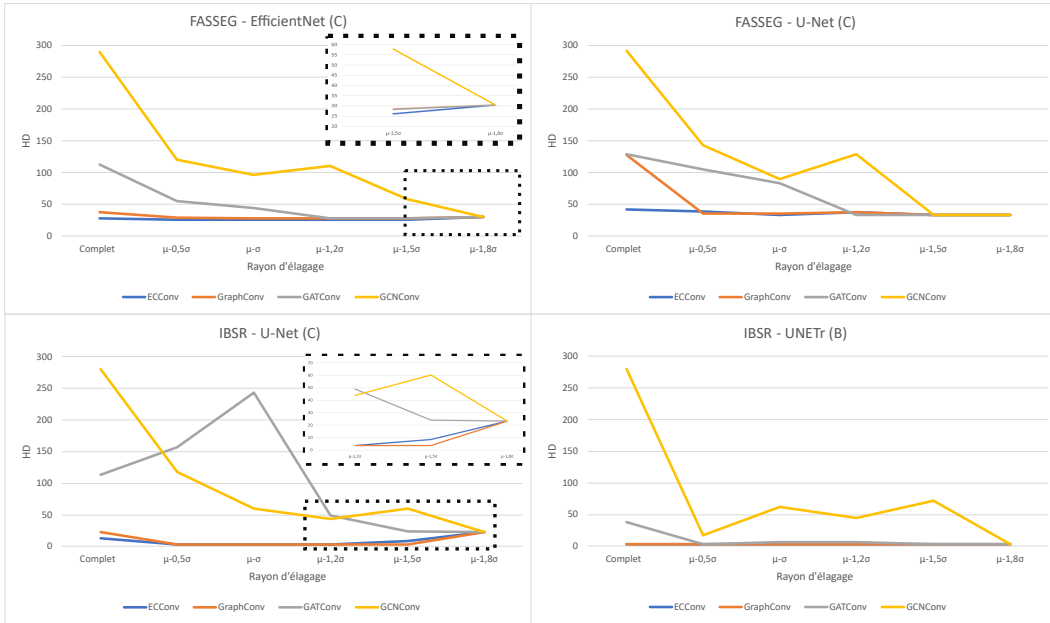


FIGURE 2.7 – Évolution de la distance d’Hausdorff des segmentations en fonction du rayon d’élargage considéré (étude sur les données d’entraînement). Deux exemples sur FASSEG (EfficientNet-C et U-Net-C) et deux sur IBSR (U-Net-C et UNETR-B) sont reportés.

les données d’entraînement (validation) en fonction du rayon d’élargage dans les 4 configurations considérées. Selon les sorties de CNN, le nombre de noeuds (et donc d’arêtes) est différent. Il en résulte un rayon d’élargage optimal différent dans chaque cas, justifiant la nécessité de définir le rayon optimal, pour chaque configuration, en fonction des données d’entraînement. Sur les deux graphes de la ligne supérieure, correspondant à l’application FASSEG, on observe des comportements similaires avec les deux réseaux EfficientNet et U-Net. Dans les deux cas, GCNConv et GATConv sont moins efficaces (parfois même moins que le CNN seul) lorsque le rayon devient trop grand ($\rho \geq \mu - 1.2\sigma$), c’est-à-dire, lorsque le nombre d’arêtes augmente (supérieur à environ 40 selon la Figure 2.6). Cela peut s’expliquer par leur trop petit nombre de paramètres entraînaibles (Table 2.4) pour gérer un tel nombre d’informations structurelles connexes. GCNConv atteint même ses meilleures performances pour $\rho = \mu - 1.8\sigma$, c’est-à-dire lorsque le GNN est réduit à un SLP, ce qui remet en question son efficacité. GraphConv et ECCConv sont plus flexibles aux différents rayons d’élargage. Néanmoins, ECCConv semble être le seul opérateur capable de travailler correctement avec des graphes complets dans tous les cas (FASSEG - U-Net sur la Figure 2.7), peut-être en raison de sa prise en compte des attributs d’arêtes combinée à son nombre élevé de paramètres entraînaibles (Table 2.4). Le troisième graphe

(IBSR - U-Net) montre des informations similaires avec la perte d'efficacité de GCNConv et GATConv lorsque le nombre d'arêtes augmente. Il met également en évidence l'impact de la nature de l'opérateur de convolution avec la ré-augmentation de HD dans le cas $\rho = \mu - 1.8\sigma$, passant d'environ 4 à 28 pour ECConv et GraphConv. Ceci souligne la valeur ajoutée de la convolution sur graphe comparé à la simple utilisation d'un SLP. Dans le cas de UNETr, qui correspond à un faible nombre de noeuds (Table 2.5), tous les rayons d'élagage semblent équivalents. De manière générale, l'élagage considéré (rayon de voisinage) améliore la segmentation par rapport à l'utilisation de graphes complets. A partir des résultats sur données d'entraînement (Figure 2.7), il apparaît judicieux de travailler avec un rayon compris entre $\mu - 1.2\sigma$ et $\mu - 1.5\sigma$ pour maximiser les performances.

Ainsi, nous avons utilisé ces 2 rayons optimaux (définis sur données d'entraînement et validation) pour entraîner le GNN sur les données d'entraînement avant de l'appliquer sur les données test. Pour confirmer l'efficacité de l'opérateur de convolution comparé à un SLP seul, nous avons comparé les performances obtenues à celles avec un GNN réduit à un unique perceptron unicouche (c.a.d. avec $\rho = \mu - 1.8\sigma$). Les résultats obtenus sont illustrés en Annexe A.1.1. Pour FASSEG, nous observons que l'utilisation d'un opérateur de convolution améliore les résultats par rapport à un SLP unique sauf pour GCNConv combiné à EfficientNet. Cette observation confirme l'importance de la convolution pour la classification de noeuds. Pour IBSR, les courbes valident notre intuition selon laquelle GCNConv et GATConv ne sont pas adaptés à ce problème trop complexe (voir Table 2.6). En effet, ils donnent de moins bons résultats qu'un SLP seul. Avec U-Net, ECConv et GraphConv enrichissent la prédiction des noeuds en réduisant considérablement la distance d'Hausdorff. La topologie du graphe influence dans ces cas les performances de segmentation. Dans le cas de UNETr, aucune valeur ajoutée n'est apportée par aucun des opérateurs de convolution. Cela peut s'expliquer par le fait que le nombre moyen de noeuds (15.6 selon la Table 2.5) est approximativement égal au nombre de classes (15), rendant la tâche de classification plus facile à résoudre. Dans ce cas, l'exploitation des attributs des noeuds par un SLP est suffisante. Cependant, cela ne représente pas la majorité des cas.

Les résultats obtenus valident notre intuition selon laquelle la topologie des graphes influe sur les résultats de segmentation et confirment la pertinence de notre approche visant à chercher un rayon d'élagage optimal.

2.5 Discussion

A travers nos expériences, nous avons montré la capacité d’un réseau de neurones sur graphes, exploitant les prédictions d’un CNN et les relations spatiales entre régions, à améliorer la qualité des segmentations sémantiques d’images initialement produites par des CNNs. Les résultats décrits dans la section 2.4.2, pour la segmentation de visages sur images 2D (FASSEG), montrent la pertinence des opérateurs de convolution sur graphe gérant les attributs d’arêtes multi-dimensionnels (fonction L) dans l’agrégation d’informations de voisinage pour améliorer la précision lors de la classification de noeuds. Cependant, lorsque le nombre de noeuds et de classes (c.a.d. la dimension des attributs de noeuds) augmente, la complexité des opérateurs de convolution (c.a.d. leur nombre de paramètres entraînaables) devient un facteur clé dans la bonne classification des noeuds. Ainsi, dans le cas de IBSR (section 2.4.3), les opérateurs plus complexes comme ECConv et GraphConv (Table 2.8) obtiennent de meilleurs résultats que des opérateurs plus simples comme GCNConv ou GATConv. Dans la section 2.4.4, nous avons montré que GCNConv n’était pas adapté au problème de segmentation sémantique. On peut supposer que cet opérateur se prête davantage à des problèmes d’apprentissage semi-supervisés.

La méthode proposée correspond à une architecture générique, applicable à tout réseau de neurones profond fournissant une carte de segmentation. Nous verrons d’ailleurs dans le chapitre 4 une application de cette méthode au problème médical au centre de cette thèse. Notre approche a l’avantage de ne pas nécessiter l’implémentation d’une architecture spécifique comme dans [86, 87, 88, 89] ou des algorithmes additionnels de sur-segmentation fournissant des superpixels comme dans [83, 84, 85]. Elle peut être étendue à d’autres types d’informations structurelles que celles exploitées (distances minimum et maximum pour FASSEG, distances au barycentre pour IBSR). Notons que des stratégies de segmentation ad-hoc pourraient améliorer les résultats dans chacune des applications (ex. méthodes de détection des points de repère faciaux pour FASSEG), au prix d’une moins bonne généralité de la méthode. La simplicité du réseau GNN proposé (seulement 2 couches) le rend relativement rapide. Ainsi, le temps d’inférence sur IBSR par exemple (construction du graphe et classification des noeuds) est de 10 secondes en moyenne (GPU Nvidia Quadro RTX 3000). L’ajout de couches de convolution supplémentaires a été testé et n’a pas fourni de meilleures performances. Enfin, notre méthode est robuste aux petits jeux de données, qui conduisent souvent à des graphes plus grands (le CNN moins bien entraîné fournit davantage de composantes connexes dans sa segmentation). La combinaison de

notre méthode avec une architecture classique comme U-Net égale voire surpasse avec peu de données d’entraînement les réseaux d’apprentissage profonds complexes comme UNETr et EfficientNet entraînés sur davantage de données. Cet aspect de la méthode sera particulièrement exploité dans le chapitre 4 de cette thèse, du fait de la difficulté d’avoir beaucoup de données médicales, notamment pathologiques et pédiatriques.

Nous avons confirmé l’hypothèse selon laquelle la topologie des graphes avait une influence sur les performances des GNNs, notamment pour cette tâche de classification de noeuds pour la segmentation sémantique (section 2.4.4). En effet, l’élagage des arêtes aide à garder l’information spatiale pertinente en réduisant le nombre d’arêtes (Figure 2.6) et donc de voisins de chaque noeud dans l’étape de *message passing* des opérateurs de convolution. Ainsi, dans le cas de IBSR, l’efficacité d’un opérateur comme GraphConv, ignorant pourtant les attributs sur arêtes, peut être liée à la topologie des graphes impliqués. Nous avons mis en évidence qu’un rayon d’élagage ρ entre $\mu - 1.2\sigma$ et $\mu - 1.5\sigma$ fournissait les meilleures performances dans les deux applications, quels que soient la configuration et le CNN considérés. Dans la Figure 2.7, nous avons observé que l’opérateur de convolution ECCConv, gérant les attributs d’arêtes multi-dimensionnels (comme montré dans la section 1.4.3), était le moins sensible à la topologie des graphes, obtenant de bonnes performances dans toutes les configurations. On peut imaginer que son plus grand nombre de paramètres entraînaibles compense mieux l’effet de la fonction invariante par permutation (eq 1.5) lors du *message passing* (moyenne, somme, etc.), pouvant partiellement éliminer le caractère discriminant de chaque arête. Pour ces raisons, nous avons décidé d’utiliser l’opérateur ECCConv dans nos expériences des chapitres 4 et 5.

L’approche présentée, bien que prometteuse, présente des limites et suppose des améliorations dont nous sommes conscients. Tout d’abord, l’efficacité de notre méthode semble particulièrement liée à la topologie des graphes considérés. Notre volonté était seulement de montrer l’impact du nombre d’arêtes dans un problème de segmentation sémantique. Néanmoins, pour l’instant, nous nous contentons de tester et comparer différents niveaux d’élagage basés sur un rayon de voisinage. Les rayons considérés sont définis arbitrairement et ne sont peut-être pas toujours les plus appropriés. Il serait judicieux d’inclure l’apprentissage de ce rayon d’élagage (paramètre ρ) dans le GNN afin de bénéficier d’une procédure d’apprentissage homogène de bout en bout, apprenant simultanément le niveau d’élagage et la tâche de classification de noeuds. Par ailleurs, il est possible que la méthode d’élagage proposée (basée sur une relation spatiale) ne soit pas la plus adaptée. On pourrait alors tenter d’appliquer des techniques d’élagage existantes comme [77, 78, 79]

mentionnées dans la section 1.4.3. Dans le chapitre 4 de ce manuscrit, nous utiliserons une solution alternative à l’élagage des arêtes. Celle-ci consistera à complexifier la fonction de projection F de ECCConv (*mapping* décrit dans la section 1.4.3) pour apprendre plus distinctement les poids de chaque attribut de chaque arête.

Une autre limite de notre méthode repose sur la nécessité que le CNN ait préalablement détecté et sur-segmenté toutes les classes recherchées. On suppose que le CNN produit des sortes de superpixels pertinents (ce qui semble être une hypothèse raisonnable d’après nos expériences) et que le post-traitement avec GNN assure une cohérence spatiale globale. Cependant, si le CNN perd une classe ou segmente comme une seule classe une région contenant plusieurs classes différentes, notre méthode ne peut pas y faire face. Pour surmonter cette limitation, une alternative pourrait être de considérer l’incertitude du CNN [105] pour subdiviser plus finement les régions produites par celui-ci. Nous pensons traiter cette idée dans des travaux futurs et l’évaluer sur d’autres jeux de données comme *Pascal VOC* [106] ou *Cityscapes* [107].

2.6 Conclusion

Dans ce chapitre, nous avons proposé une méthode de post-traitement basée sur un réseau de neurones sur graphes, afin d’améliorer la segmentation sémantique d’images fournie par des réseaux de neurones profonds dédiés (CNNs). Notre approche convertit la carte de segmentation du CNN en graphe, où chaque noeud correspond à une région segmentée. Elle exploite les vecteurs d’appartenance issus de la sortie du CNN comme attributs de noeuds et les relations spatiales entre régions segmentées comme attributs d’arêtes dans l’objectif de réaliser une classification de noeuds. Nous avons étudié des paramètres, introduits dans le chapitre 1, influençant la performance des GNNs : la nature de l’opérateur de convolution considéré et la topologie des graphes en entrée. Les résultats nous ont emmenés à la conclusion qu’un opérateur comme ECCConv offrait de bonnes performances avec une certaine flexibilité quant à la topologie des graphes donnés en entrée grâce à sa prise en compte des attributs d’arêtes multi-dimensionnels et à son nombre plus important de paramètres entraînaibles. Par ailleurs, nous avons observé que l’élagage des arêtes des graphes simplifiait la tâche d’agrégation d’informations de voisinage en ne conservant que les informations structurelles pertinentes, conduisant souvent à de meilleurs résultats. C’est pourquoi, dans les travaux de la partie II, nous emploierons ECCConv et tiendrons une attention particulière à la structure des graphes considérés.

Nous avons validé notre méthode sur deux applications (images 2D et volumes 3D) et avec divers CNNs. Notre proposition semble être robuste aux petits ensembles de données d’entraînement peu représentatifs, surpassant des architectures de réseaux de neurones profonds plus complexes (EfficientNet, UNETr) entraînées avec davantage de données. Le post-traitement proposé est une méthode générique complémentaire pour la segmentation sémantique d’images, pouvant être étendu à tout type d’information structurelle et à toute autre propriété supplémentaire décrivant les régions (section 1.3). Il est à noter que ces travaux ont fait l’objet d’une communication à une conférence internationale [108].

À retenir

- Un GNN est utilisé en post-traitement d’un CNN pour améliorer la segmentation sémantique d’images au travers d’une tâche de classification de noeuds en se basant sur la carte de segmentation fournie par le CNN et sur les relations structurelles entre régions détectées.
- Une stratégie d’élagage des arêtes, définie par un rayon de voisinage, est utilisée pour limiter les informations structurelles prises en compte lors de la convolution sur graphe.
- La méthode développée s’avère générique respectivement à l’application, à la nature de l’information structurelle considérée et au CNN utilisé pour la pré-segmentation.
- La pertinence de l’approche a été évaluée sur deux bases de données publiques, avec plusieurs tailles de jeux de données d’entraînement, en considérant différents CNNs et avec divers opérateurs de convolution sur graphe. Notre approche combinée à un réseau classique comme U-Net et entraînée sur peu de données surpasse des réseaux plus complexes (ex. *Visual Transformers*) entraînés avec davantage de données.

Dans la partie I de cette thèse, nous avons présenté diverses notions techniques (CNNs, informations structurelles, GNNs) et une méthode combinant ces concepts pour la segmentation sémantique d’images qui seront exploitées dans la partie II du manuscrit pour traiter une application médicale : l’étude de la lésion cérébrale précoce chez l’enfant après AVC néonatal.

DEUXIÈME PARTIE

**Application à l'étude de la lésion
cérébrale précoce chez l'enfant après
AVC néonatal**

INTRODUCTION À L'ÉTUDE DE LA LÉSION CÉRÉBRALE PRÉCOCE

Dans cette partie, nous allons aborder une application médicale qui a été au coeur de nos travaux de thèse : l'étude en IRM de la morphométrie cérébrale en présence d'une lésion cérébrale précoce chez l'enfant après AVC néonatal ainsi que la prédiction, à partir d'IRMs, de la présence ou non de troubles moteurs permanents définis comme paralysie cérébrale de l'enfant. Nous nous appuyerons sur certaines notions (CNNs, GNNs, informations structurelles - chapitre 1) et travaux (chapitre 2) présentés dans la partie I.

La lésion cérébrale précoce est une lésion intervenue sur un cerveau en cours de développement et de maturation, durant la vie foetale ou dans les deux premières années de vie qui a des conséquences sur le développement morphostructurel du cerveau de l'enfant. Cette lésion, progressive ou non, peut prendre plusieurs formes (ex. lésion de la substance blanche, lésion corticale ou sous-corticale, malformations), intervenir à des moments de développement cérébral différents et avoir des caractéristiques variées (lésion localisée ou diffuse). Pour étudier l'impact sur le cerveau de la lésion cérébrale précoce en neuroimagerie, nous avons besoin de modèles animaux ou d'études sur le cerveau humain (Figure II.1-gauche).

L'AVC néonatal est considéré comme un bon modèle d'étude de l'impact de la lésion cérébrale précoce sur la plasticité du cerveau en cours de développement car il correspond à une lésion bien caractérisée (lésion corticale ou sous-corticale), non progressive, intervenant toujours à la même période chez des enfants nés à terme [109]. Ainsi, l'utilisation de ce modèle permet de se libérer de facteurs de confusion liés à la temporalité ou aux caractéristiques de la lésion. Avec une prévalence à la naissance de [37-67]/100 000, l'AVC néonatal est un terme général englobant "*un groupe de maladies ayant comme caractéristique commune l'interruption focale du flux sanguin cérébral, secondaire à une occlusion artérielle ou veineuse, survenant entre la 20ème semaine de vie foetale et le 28ème jour post-natal*" [110]. Cette dénomination englobe elle-même plusieurs pathologies (ex. thrombose néonatale des sinus et des veines cérébraux, hémorragie cérébrale) dont l'infarctus cérébral artériel néonatal (anglais - NAIS) est le plus courant avec une prévalence estimée

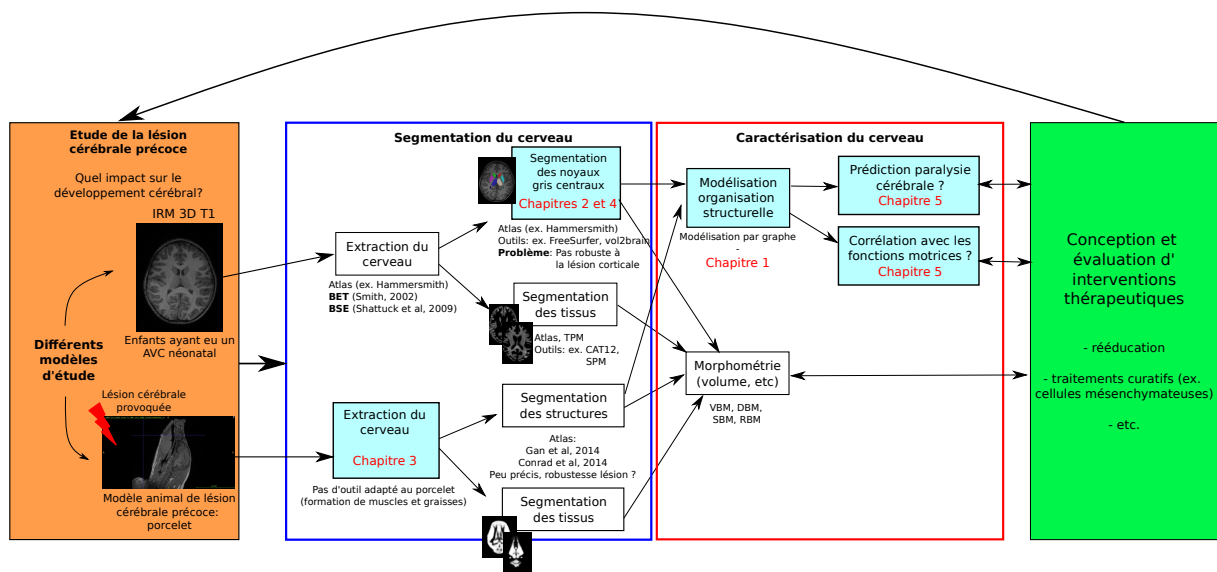


FIGURE II.1 – Vue d’ensemble des travaux relatifs à l’étude de la lésion cérébrale précoce. Les éléments de la chaîne abordés durant la thèse sont indiqués sur fond bleu. Pour chaque autre étape, des outils/techniques existants sont renseignés.

à environ 1/4000 (25/100 000) naissances à terme [111]. Le handicap moteur (paralysie cérébrale) est la principale conséquence de l’AVC néonatal [109]. En effet, on estime que environ 20-30% des enfants après NAIS développent une paralysie cérébrale (PC) unilatérale, première cause de handicap moteur de l’enfant en France [112]. De manière générale, les nouveaux-nés présentant les lésions cérébrales les plus étendues sont les plus exposés aux handicaps moteurs à long terme [113]. Toutefois, certains nouveaux-nés chez lesquels la lésion cérébrale est relativement petite mais spécifiquement localisée le long des voies motrices peuvent également développer une PC [114]. Parmi les structures cérébrales clés dans les fonctions sensori-motrices, on trouve notamment les noyaux gris centraux [115] (thalamus, noyau caudé, putamen et pallidum illustrés sur la Figure II.2) constituant la "plaque" tournante du cerveau, spécialement en terme de contrôle de la motricité. On peut faire l’hypothèse que la présence, directe ou indirecte (par phénomène de dégénérescence Wallerienne, c.a.d. de perte neuronale), d’une lésion corticale au niveau des noyaux gris centraux, subissant normalement une importante croissance après la naissance [116], puisse provoquer l’apparition de troubles moteurs. Sur la Figure II.2, on observe macroscopiquement que la présence d’une lésion dans le cerveau semble entraîner une désorganisation des noyaux gris centraux. En effet, la relative symétrie présente dans les cerveaux sains paraît déformée chez les enfants ayant eu un AVC néonatal. On peut

faire l'hypothèse que plus le "désordre" dans ce réseau profond est important, plus les fonctions motrices de l'enfant seront affectées. Ces caractéristiques relatives aux noyaux gris centraux en font des structures très souvent étudiées dans le contexte de l'AVC néonatal [117, 118, 119].

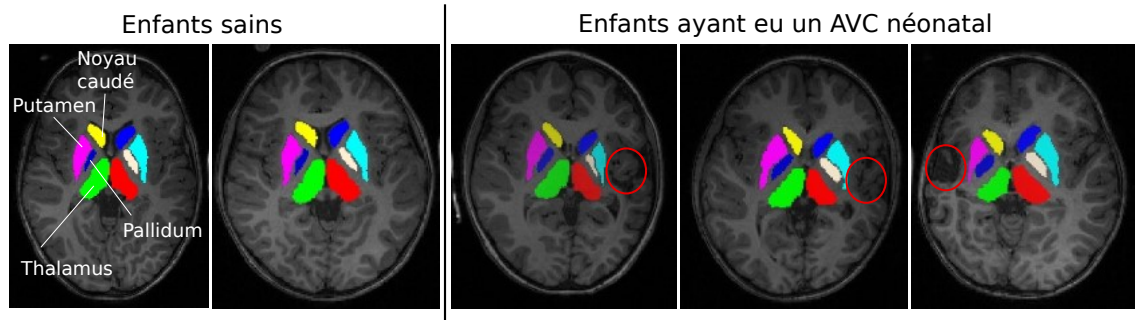


FIGURE II.2 – Noyaux gris centraux sur IRM cérébrale d'enfants sains (à gauche) et d'enfants après AVC néonatal (à droite). La lésion dans les cerveaux affectés est repérée en rouge. Une sorte de désorganisation des noyaux gris centraux semble découler de la présence de la lésion. Les IRMs et segmentations proviennent de nos travaux.

L'imagerie par résonance magnétique (IRM) est un outil adapté pour étudier les conséquences *in vivo* de la lésion cérébrale précoce sur le cerveau, à l'échelle macroscopique, avec des mesures objectives, et pour assurer des comparaisons à la fois longitudinales (intra-sujet) et inter-sujets, ceci étant très compliqué avec des analyses radiologiques classiques. Excepté pour les cas les plus graves, il est difficile de prédire l'apparition d'un handicap moteur à partir d'une simple IRM conventionnelle. La prédiction du développement moteur de l'enfant reste un défi et de nouveaux outils cliniques et/ou radiologiques sont en cours de développement pour mieux prédire la motricité à long terme [120]. Il est important d'établir un diagnostic précoce de la PC et de comprendre l'apparition de celle-ci car cela peut accélérer la mise en place d'interventions thérapeutiques plus adaptées à l'enfant (Figure II.1-droite) et mieux informer les parents sur le devenir de leur enfant.

Une étape préalable aux études relatives à l'influence des noyaux gris centraux sur la motricité des enfants après AVC néonatal consiste à segmenter, le plus fidèlement possible, ces structures cérébrales sur IRM. Nous avons vu dans le chapitre 1 que les techniques d'apprentissage profond, notamment basées sur des réseaux convolutifs (CNNs), obtiennent de très bons résultats pour ce type d'application. Néanmoins, la tâche de segmentation des noyaux gris centraux est complexifiée par divers facteurs dans le contexte de l'enfant avec lésion cérébrale précoce. D'une part, le jeune enfant avec lésion cérébrale est un sujet d'étude difficile. En effet, à ce stade de développement, le cerveau

évolue rapidement et entraîne un contraste tissulaire sur IRM variant en fonction de l'âge de l'enfant (contraste particulièrement peu prononcé avant l'âge de 2 ans). De plus, les enfants ont davantage tendance à bouger durant l'acquisition IRM, rendant certaines images inexploitable. Enfin, pour des raisons éthiques évidentes combinées à la prévalence de la pathologie et à la mise en place de protocoles d'acquisition plus difficile que chez l'adulte, les cohortes d'enfants sont généralement petites, confrontant les CNNs à un manque de données de bonne qualité pour entraîner les modèles. Les techniques alternatives (augmentation de données, apprentissage par transfert), présentées dans la section 1.2.4, peuvent alors ne pas être suffisantes. D'autre part, les noyaux gris centraux sont des structures difficiles à segmenter car elles sont petites et leurs contours sont ambigus sur IRM à cause du faible contraste des tissus cérébraux dans cette région profonde du cerveau [22] (observation accentuée chez l'enfant). Leur segmentation est encore plus difficile après lésion corticale (type AVC néonatal) car celle-ci provoque des déformations et modifications des structures (Figure II.2), même à distance de la lésion, altérant parfois le fonctionnement des algorithmes existants pour la segmentation des structures cérébrales [22, 121, 122, 123, 124].

Par ailleurs, afin de proposer des interventions plus ciblées améliorant la réorganisation des réseaux neuronaux, il est important de comprendre les processus complexes de neuroplasticité post-lésionnelle se mettant en place dans le cerveau immature de l'enfant. Pour étudier les mécanismes *in vivo* de la lésion cérébrale précoce et de la plasticité cérébrale, les modèles animaux sont également nécessaires car ils offrent des conditions expérimentales contrôlées. Il existe des modèles de rongeurs avec PC [125] mais ces derniers, comparés à l'humain, peuvent présenter de potentiels biais et différences pouvant entraver l'obtention de résultats translationnels fiables. D'autres modèles animaux sont donc nécessaires [126]. Parmi les modèles animaux de plus en plus étudiés, on retrouve le macaque (étroites similitudes dans la structure cérébrale, la connectivité et la fonction avec l'humain) [127], le mouton (études hormonales de l'adulte et du fœtus ainsi que des mécanismes endocriniens) [128], le furet (études sur le processus de myélinisation) [129] ou encore le chien (études relatives aux troubles d'anxiété, épilepsie, déclin cognitif, vieillissement) [130]. Concernant l'étude du développement neurologique précoce, un modèle d'animal a suscité l'intérêt ces dernières années : le porcelet [131]. En effet, les caractéristiques anatomiques du cerveau du porcelet (circonvolutions, rapport substance blanche/substance grise, développement du système nerveux, croissance et maturation de la myéline) sont très similaires à celles de l'être humain [132]. Par ailleurs, ce type de

gros animal permet d'utiliser les scanners IRM adaptés à l'humain. Ainsi, le modèle du porcelet permet de mieux comprendre le processus de maturation du cerveau et d'analyser les futurs troubles du développement neurologique [133]. A cette fin, un modèle de porcelet reproduisant la lésion cérébrale corticale a été développé dans le cadre du projet REPAR [134]. L'objectif de ce projet était de caractériser l'impact de la lésion cérébrale précoce sur le développement du cerveau (volumes tissulaires, organisation structurelle, etc.) ainsi que d'évaluer la capacité de cellules mésenchymateuses à réparer la lésion en testant l'injection de cellules souches à certains porcelets lésés de l'étude. En effet, une grande avancée serait de proposer des thérapies curatives pour "guérir" le cerveau suite à la lésion cérébrale précoce puisqu'il n'en existe pas à ce jour. Le modèle du porcelet est encore peu étudié et présente un nombre très réduit de données du fait de restrictions éthiques relatives à l'étude sur l'animal. Cela s'associe à un petit nombre d'atlas existants [131, 135], encore peu précis car réalisés avec peu de données (15 porcelets pour l'atlas de [135]) et spécifiques à un certain stade de développement (1 mois pour l'atlas de [135]). Par ailleurs, même si le modèle du porcelet présente des similarités avec le cerveau humain, ses différences s'accroissent rapidement avec l'âge (formation de nombreux sinus (air), de muscles et de graisses péri-cérébraux), complexifiant les études longitudinales sur cet animal avec les outils existants sur l'humain [136, 137, 138, 139]. Pour ces raisons, il n'existe pas d'outil de traitement d'images automatisé pour l'ensemble de la chaîne de segmentation, présentée Figure II.1 (extraction du cerveau, segmentation des structures et tissus cérébraux). Pourtant, la segmentation des différentes parties du cerveau est nécessaire pour assurer un suivi longitudinal du développement cérébral avec des mesures quantitatives objectives. Ces segmentations se font donc entièrement à la main, ceci étant long, opérateur-dépendant et non propice à la construction de grands jeux de données.

Ainsi, l'étude en neuroimagerie de la lésion cérébrale précoce est difficile chez l'enfant comme chez l'animal mais nécessaire à la conception et à l'évaluation d'interventions thérapeutiques. Dans ce contexte, nous proposons des algorithmes contribuant à faciliter l'étude de la lésion cérébrale précoce, à partir d'IRMs, sur l'humain (enfants ayant eu un AVC néonatal) ou sur des modèles animaux (porcelets), à différents niveaux de la chaîne de traitement (extraction du cerveau pour le porcelet, segmentation des structures cérébrales pour l'enfant) comme illustré par la Figure II.1. Les algorithmes développés ont pour objectif d'automatiser la segmentation du cerveau et de proposer des caractéristiques anatomiques du cerveau après AVC néonatal permettant de mettre en place et d'évaluer des interventions thérapeutiques (ex. thérapies de rééducation, traitements curatifs).

Ainsi, dans le chapitre 3 de cette thèse, nous proposons une méthode d'extraction automatique et longitudinale du cerveau du porcelet basée sur un réseau U-Net 3D (section 1.2.3) par patches et sur de l'apprentissage par transfert (section 1.2.4). L'objectif est de proposer une méthode efficace et robuste à la formation de sinus (air), de muscles et de graisses péri-cérébraux rendant peu efficaces les outils existants sur l'humain (ex. BET [136], BSE [137], LABEL [138], MONSTR [139]). Nous n'avons pas abordé le problème de la segmentation des structures et tissus cérébraux sur le modèle porcin. Si les atlas existants [131, 135] permettent de pré-segmenter le cerveau, des corrections manuelles restent indispensables du fait de l'imprécision de ces atlas, notamment en présence d'une lésion cérébrale. Ces étapes de traitement d'images devront faire l'objet de futurs travaux pour être entièrement automatisées. Nous pourrions ensuite tenter de transposer la chaîne de segmentation du cerveau automatisée à d'autres modèles animaux (chiens, moutons, furets, etc.). Enfin, les analyses morphométriques (ex. VBM, DBM, SBM, RBM) liées à l'impact de la lésion cérébrale et à l'injection de cellules souches chez le modèle porcin n'ont pas été menées à bout pour des motifs externes à notre responsabilité.

Dans le chapitre 4, nous appliquons une version légèrement modifiée de la méthode de segmentation sémantique présentée dans le chapitre 2 pour segmenter les noyaux gris centraux d'enfants (âgés de 7 ans) ayant eu un AVC néonatal. Nous montrons que notre post-traitement, basé sur l'exploitation d'informations structurelles de haut-niveau par GNN, permet de corriger les incohérences structurelles de segmentation, liées à la lésion cérébrale, commises par les outils automatiques existants (FreeSurfer [121], vol2Brain [122]) ou des CNNs dédiés à cette tâche [22, 123, 124].

Dans le chapitre 5, nous proposons de caractériser le cerveau, et plus précisément, l'organisation structurelle des noyaux gris centraux par une modélisation sous forme de graphes comme présenté dans le chapitre 1. Nous cherchons à voir si l'information structurelle des noyaux gris centraux, sous forme de graphes exploités par GNNs, permet d'aider d'une part à déterminer la présence d'une paralysie cérébrale chez des enfants après NAIS à partir d'IRMs acquises à 7 ans (section 5.4) et d'autre part à estimer la motricité (manuelle) de ces enfants (section 5.5).

Dans les expériences décrites dans les chapitres suivants, nous nous baserons sur deux bases de données :

1. Pour l'étude relative à l'extraction du cerveau du porcelet : cohorte issue du projet REPAR [134] de 27 porcelets (26 mâles et 1 femelle) suivis durant les 4 premiers mois de vie. Ce jeu de données et les IRM 3D T1 acquises sont décrites plus

précisément dans le chapitre 3.

2. Pour l'étude des noyaux gris centraux : cohorte *AVCnn* de 69 enfants de 7 ans incluant 31 enfants dits contrôles et 38 enfants avec NAIS (lésion dans l'hémisphère droit ou gauche) [120, 140]. Ce jeu de données et son utilisation dans les expériences sont détaillés dans les chapitres 4 et 5.

À retenir

- L'AVC néonatal, bon modèle d'étude de la lésion cérébrale précoce, se traduit par l'apparition, aux alentours de la naissance, d'une lésion cérébrale non progressive et peut, dans environ 30% des cas, entraîner le développement d'une paralysie cérébrale (PC) unilatérale provoquant des troubles moteurs permanents.
- Il est important de comprendre l'impact de la lésion sur le développement cérébral de l'enfant, notamment sur l'apparition de la PC, pour mettre en place très tôt des interventions thérapeutiques adaptées (rééducation, traitements curatifs).
- Les noyaux gris centraux jouent un rôle clé dans les fonctions sensori-motrices. Des anomalies, suite à un AVC néonatal, dans cette région du cerveau semble être liées à l'apparition de troubles moteurs. Mais, l'étude de ces structures sur IRM est difficile du fait du faible contraste des tissus cérébraux. Cela est particulièrement vrai chez le jeune enfant et est d'autant plus complexe avec la présence d'une lésion dans le cerveau et le peu de données à disposition.
- Les modèles animaux sont aussi nécessaires pour mieux comprendre les mécanismes *in vivo* de la lésion cérébrale. Les similarités du cerveau du porcelet avec celui du jeune enfant en font un bon modèle pour l'étude du développement neurologique précoce mais difficile à étudier par l'absence d'outils automatisant l'entièreté de la chaîne de segmentation du cerveau de cet animal, à commencer par l'extraction du cerveau.

APPRENTISSAGE PROFOND POUR L'ÉTUDE DU PORCELET : MODÈLE ANIMAL DE DÉVELOPPEMENT NEUROLOGIQUE PRÉCOCE

3.1 Introduction

Dans ce chapitre, nous proposons une méthode automatique d'extraction de cerveau, basée sur l'apprentissage profond et appliquée au cerveau du porcelet. Cette méthode est destinée à faciliter l'analyse du cerveau d'un tel modèle de gros animal adapté à l'étude du développement cérébral précoce. Nous reprenons brièvement ici des éléments de contexte, que nous avons commencé à présenter dans l'introduction de la Partie II.

Tout d'abord, nous avons vu que les modèles animaux sont nécessaires, en plus des études sur l'humain, pour comprendre les mécanismes *in vivo* de la lésion cérébrale précoce car ils fournissent des conditions expérimentales contrôlées. En effet, « *les remarquables similitudes anatomiques et physiologiques entre les hommes et les animaux, [. . .] ont incité les chercheurs à étudier de nombreux mécanismes et à évaluer les nouveaux traitements sur des modèles animaux avant d'appliquer leurs découvertes à l'homme* » [141]. Nous avons vu que plusieurs modèles animaux sont utilisés pour comprendre le fonctionnement cérébral humain (rongeurs [142], macaques [127], moutons [128], furets [129], chiens [130], etc.) mais que, selon le type d'étude, ils ne sont pas tous adaptés de par leurs différences anatomiques, morphologiques et développementales. A titre d'exemple, si le rongeur est le modèle le plus exploité du fait de sa petite taille, de sa résistance et de la facilité à le nourrir et à le soigner, les différences de son cerveau comparé au cerveau humain (ex. absence de sulci et gyri, surface corticale lisse) sont autant de biais pouvant entraver l'obtention de résultats translationnels fiables. Ainsi, nous avons constaté que le modèle

du porcelet a suscité l'intérêt ces dernières années pour l'étude du développement cérébral précoce du fait des similarités de son cerveau avec celui du jeune enfant (morphologie et croissance périnatale, nombre de neurones néo-corticaux, présence de plis, distribution tissulaire) [132] et de sa taille suffisamment grande pour utiliser les scanners IRM adaptés à l'humain.

Par ailleurs, nous avons expliqué que l'imagerie par résonance magnétique (IRM) est un outil adapté pour étudier les conséquences de la lésion cérébrale précoce car on peut en extraire des mesures objectives (ex. analyses morphométriques) tout en assurant une comparaison à la fois longitudinale (intra-sujet) et inter-sujets, ceci étant difficile avec des analyses radiologiques classiques. Néanmoins, avant de pouvoir appliquer des analyses morphométriques caractérisant le cerveau, il est nécessaire de segmenter ce dernier (segmentation des structures et des tissus cérébraux). Or, nous avons souligné la difficulté de telles opérations de traitement d'images dans le cas du porcelet. En effet, la rapide croissance du cerveau du porcelet durant les premières semaines de vie, s'accompagnant de la formation de nombreux sinus (air), muscles et graisses péri-cérébraux, complexifie les études longitudinales sur cet animal avec les outils existants pour l'humain (ex. BET [136], BSE [137], LABEL [138], MONSTR [139]). L'absence d'outils adaptés au porcelet s'explique par le peu d'études existantes sur ce modèle. Certains travaux ont été menés sur le cochon nain (Göttingen [143], Yucatan [144]) et sur le porc domestique [145]. Mais, dans le cas du porcelet, on trouve uniquement quelques travaux s'attelant à la création d'un atlas du porcelet [131, 135] afin d'automatiser la segmentation des structures cérébrales avec une approche basée sur atlas. Cependant, une étape préliminaire de la chaîne de traitement reste un défi : l'extraction du cerveau du porcelet. Actuellement, le masque du cerveau est fait manuellement avec des logiciels adaptés (ex. ITKSnap [146], MITK [147]). C'est une opération longue (environ 1h30 par IRM), dépendante d'une expertise humaine et inappropriée à la construction de grands jeux de données annotés. Il est donc nécessaire de trouver une alternative pour automatiser l'extraction du cerveau du porcelet sur IRM. Notre objectif est de proposer une méthode, efficace sur le porcelet, et suffisamment générique pour être transposée à d'autres modèles animaux de grande taille. De plus, on souhaite que cette méthode reste efficace même avec un petit jeu de données d'entraînement, sans nécessiter d'atlas contrairement à d'autres travaux analogues [148].

L'approche proposée pour l'extraction du cerveau se base sur une segmentation par apprentissage profond (section 1.2.2) et plus particulièrement sur une architecture U-Net [13] (section 1.2.3). L'utilisation des réseaux de neurones profonds pour l'extraction du

cerveau sur IRM a déjà montré ses performances sur le cerveau de l'enfant, en particulier celui du fœtus. Ainsi, *Ebner et al.* [149] proposent une chaîne de traitement pour la reconstruction du cerveau du fœtus débutant par une extraction du cerveau basée sur un CNN 2D. De la même manière, *Salehi et al.* [150] considèrent un U-Net 2D pour segmenter le cerveau du fœtus sur IRM. L'utilisation de réseaux convolutifs manipulant directement des volumes 3D est encore peu exploitée pour la segmentation de cerveaux animaux malgré de bons résultats sur les rongeurs (PCNN3D [151]) ou les macaques (DIKA-Nets [152]). Ils n'ont encore jamais été considérés pour segmenter le cerveau du porcelet. Afin de combiner le contexte général et des informations locales pendant l'entraînement du modèle CNN, nous proposons une approche par patches 3D, comme cela a été récemment proposé pour l'analyse d'IRMs du cerveau humain [103, 153]. L'utilisation d'un modèle U-Net basé sur des patches 3D, combinée à la stratégie de sélection de ces patches (focalisée sur les contours du cerveau), constitue la première originalité de ce travail.

Comme expliqué précédemment, l'étude de l'impact de la lésion cérébrale précoce sur le développement du cerveau nécessite un suivi longitudinal du développement cérébral. Pour cela, il est nécessaire de fournir une solution permettant de segmenter fidèlement le cerveau sur IRM à n'importe quel stade du processus de maturation et ce, malgré les transformations rapides et importantes du cerveau du porcelet décrites auparavant (intensification des contrastes de tissus, formation de sinus, muscles et graisses autour du cerveau). Par ailleurs, la base de données, déjà relativement petite pour des raisons éthiques et financières, peut se réduire au fil du temps du fait des conditions expérimentales (mort possible de porcelets), complexifiant l'entraînement d'un CNN à chaque stade de développement. Pour répondre à ce problème, nous proposons une stratégie d'apprentissage par transfert (section 1.2.4) exploitant les données de jeunes porcelets (âgés de 2 semaines) pour pré-entraîner le réseau U-Net 3D et transférer ensuite les caractéristiques apprises pour segmenter le cerveau de porcelets plus âgés malgré les échantillons réduits à ces stades plus avancés. De cette manière, la segmentation du cerveau à chaque stade de développement est réalisée à l'aide d'un réseau pré-entraîné avec les caractéristiques du cerveau du stade de maturation précédent. L'utilisation de l'apprentissage par transfert pour la segmentation du cerveau a déjà montré sa capacité à transférer de bonnes performances du cerveau humain au cerveau de primates non humains ou de porcs adultes [154]. Cependant, elle n'a encore jamais été testée pour une étude longitudinale de cerveau porcin, ceci constituant notre deuxième originalité.

Nous avons enfin vu que plusieurs travaux s'intéressent à la segmentation du cerveau

animal. On peut citer des travaux relatifs au cochon [154], au macaque (DIKA-Nets [152], atlasBREX [148]) ou encore au rongeur (U-Net 2D [155], PCNN3D [151], RATS [156] et atlasBREX [148]). Notre proposition diffère des techniques existantes car elle considère une approche basée sur des patches 3D (approche 2D seulement dans [155], approche 3D sans patches dans [151, 154]), ne requiert pas d'atlas (contrairement à [148]) et aborde la question du suivi longitudinal (hors du champ d'application de nombreux travaux connexes [151, 154, 155, 156]) en utilisant une approche par transfert (ignorée dans [152]).

Dans la suite de ce chapitre, nous détaillerons la méthode proposée (section 3.2) avant de présenter notre base de données, les expériences menées et les résultats obtenus (section 3.3). Nous achèverons le chapitre par une discussion des avantages et limites de l'approche.

3.2 Méthode proposée

La Figure 3.1 donne un aperçu de la méthode proposée sur un exemple d'IRM T1 de porcelet. Un premier réseau est entraîné à 2 semaines (t_0) à partir des données IRM disponibles (voir Figure 3.1 - bas). Depuis une base de données annotée d'entraînement d'IRMs de porcelets (représentée sur fond bleu), des patches se chevauchant de dimension 32^3 voxels sont extraits selon une stratégie spécifique (section 3.2.2.1). Les patches extraits sont utilisés pour entraîner un réseau de neurones à segmenter le cerveau de porcelets sur IRM. Le modèle considéré est un réseau U-Net (section 3.2.1) recevant en entrée les patches 3D de dimension 32^3 voxels extraits de l'IRM T1 et fournissant en sortie des patches de même taille avec la segmentation du cerveau. Ainsi, à la rencontre d'une nouvelle IRM de porcelet (représentée sur fond rose), nous divisons d'abord celle-ci en patches adjacents de dimension 32^3 voxels qui passeront par le réseau U-Net afin d'être segmentés. Les patches segmentés sont ensuite agrégés pour reconstituer le volume entier et obtenir une première version du masque du cerveau. Ce masque passe ensuite dans une chaîne de post-traitement destinée à corriger ses irrégularités et à éliminer les faux-positifs. Cette chaîne se compose d'une conservation de la plus grande composante connexe du masque (connectivité de 18 considérée) suivie d'un lissage (filtre médian de dimension 5^3 voxels) et d'une régularisation spatiale en 2D par des champs aléatoires conditionnels (CRFs) tels que présentés dans [157] et considérés dans le chapitre 2. Une présentation plus détaillée des CRFs est fournie dans l'article associé à ces travaux [158]. La chaîne de post-traitement s'achève par une dilatation morphologique afin d'obtenir une segmentation du

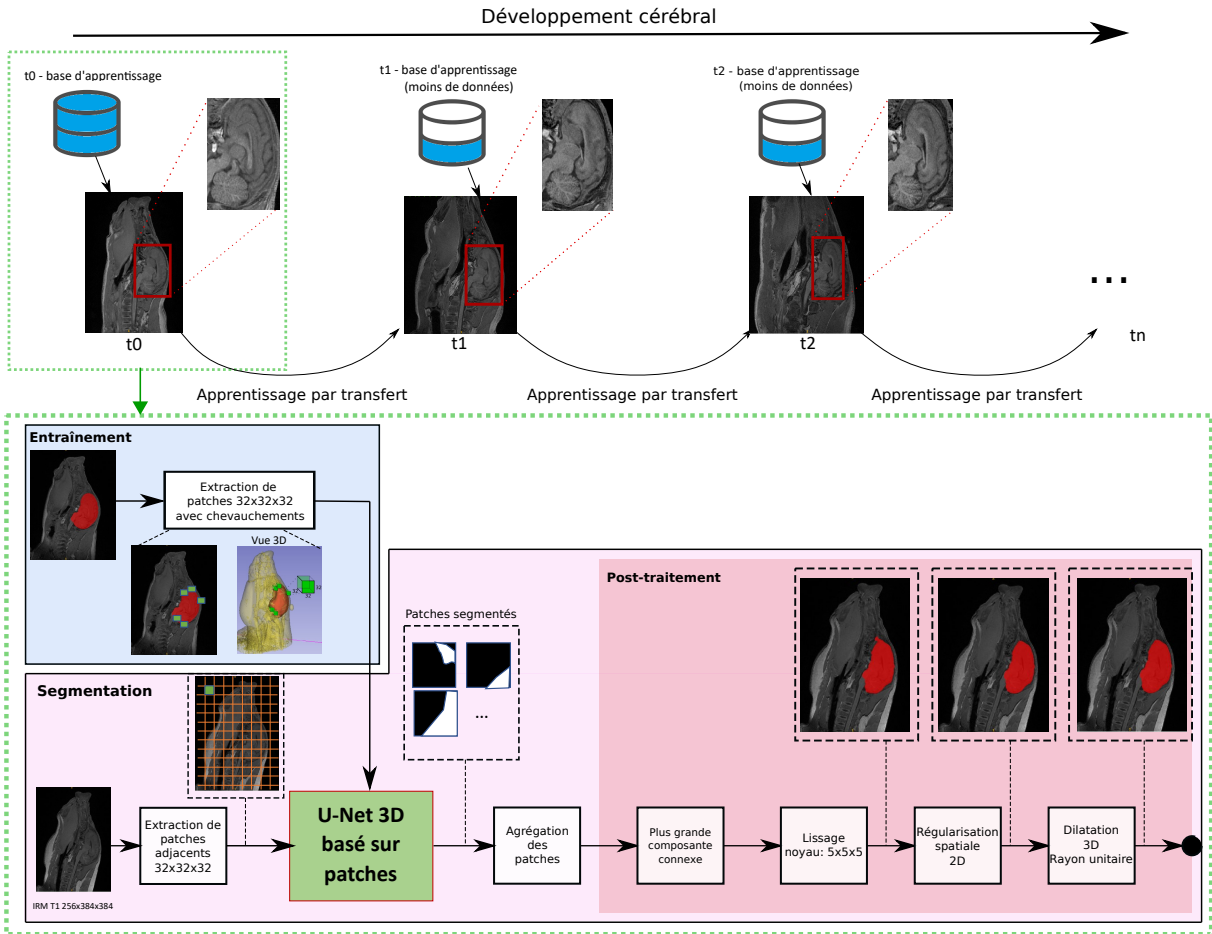


FIGURE 3.1 – Vue d’ensemble de la méthode proposée de segmentation automatique du cerveau du porcelet sur IRM T1. La partie supérieure illustre le principe de l’apprentissage par transfert appliqué au développement du cerveau (intensification des contrastes, formation de graisses, croissance). La partie inférieure détaille la chaîne de segmentation proposée (apprentissage profond et post-traitement).

cerveau avec une marge de précaution pour les étapes ultérieures de segmentation des structures cérébrales internes.

Un deuxième réseau U-Net est entraîné 4 semaines plus tard (t_1) à partir d’une base d’apprentissage plus petite (Figure 3.1 - haut). Pour compenser la petite taille de cette base de données, un mécanisme d’apprentissage par transfert (section 3.2.3) est mis en place pour transférer les caractéristiques apprises à 2 semaines (t_0). Cette logique est répétée à chaque stade de développement étudié pour assurer un suivi longitudinal. Cela signifie que, au stade t_2 (10 semaines), un autre réseau U-Net est entraîné à partir des données disponibles et du transfert des caractéristiques apprises au stade t_1 (6 semaines).

Dans la suite de cette section, nous détaillons l'architecture du réseau U-Net considéré (section 3.2.1), la stratégie d'extraction de patches choisie pour l'entraînement du réseau et pour la phase d'inférence (section 3.2.2) ainsi que l'apprentissage par transfert utilisé pour l'étude longitudinale (section 3.2.3).

3.2.1 Architecture du réseau U-Net

L'architecture U-Net classique est présentée dans la section 1.2.3.

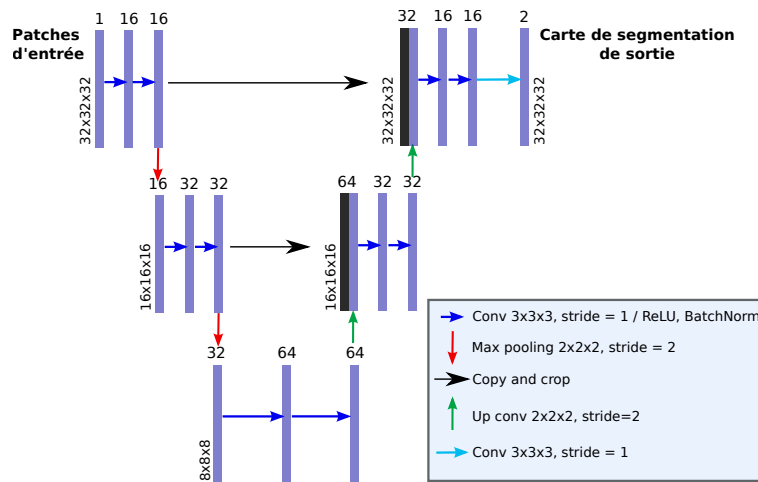


FIGURE 3.2 – Architecture du réseau U-Net implémenté.

La Figure 3.2 représente l'architecture du réseau U-Net implémenté. Il reçoit en entrée des patches de dimension 32^3 voxels et produit 2 cartes de segmentation de dimension 32^3 correspondant aux 2 classes (fond et cerveau) après application d'une fonction *Softmax*.

3.2.2 Stratégie d'extraction de patches 3D

Dans la section 1.2.3, nous avons rappelé que l'architecture U-Net a montré de bonnes performances pour la segmentation d'images. Cependant, plus la taille de l'image d'entrée est importante, plus la capacité de la mémoire du GPU nécessaire pour l'apprentissage est importante. De plus, lorsque l'architecture prend en entrée l'ensemble de l'IRM ou une coupe de l'IRM, le modèle a tendance à manquer de détails dans certaines zones de l'image (problème de classes déséquilibrées [103]). Pour surmonter ce problème, nous proposons une approche basée sur patches qui présente l'avantage d'être plus précise (le réseau peut se concentrer sur les informations locales au patch) tout en nécessitant moins de mémoire pour l'entraînement et l'inférence.

Afin d'enrichir l'information de contexte contenue dans le patch, une bonne solution consiste à utiliser des patches 3D bénéficiant de l'information dans les 3 directions (axiale, coronale et sagittale) [153]. En comparaison, l'approche 2D utilise l'information contextuelle sur 2 directions seulement et requiert un post-traitement additionnel pour fusionner chaque décision 2D. Nous détaillons ci-après les stratégies d'extraction de patches 3D pour les phases d'entraînement et d'inférence.

3.2.2.1 Entraînement

Pour l'entraînement, nous considérons des patches pouvant se chevaucher (Figure 3.1) car le chevauchement de patches aide le CNN à voir à la fois l'information locale et globale [153]. En effet, l'utilisation de petits patches 3D permet de se concentrer sur l'information locale (ex. voisinage local de 32^3 voxels). Mais, la multiplicité des patches, couvrant une large région du cerveau, apporte de l'information globale au réseau. Chaque partie du cerveau est perçue à travers différents patches, chacun apportant son information locale. La prédiction \hat{y}_i de chaque voxel pendant l'entraînement vient de la décision de tous les patches p superposés contenant ce voxel (eq 3.2). Pour chaque voxel dans chaque patch, le U-Net fournit une probabilité d'appartenance au cerveau. Les paramètres du réseau sont ajustés afin de minimiser la fonction de coût d'entropie croisée binaire [159] :

$$L(Y, \hat{y}) = \sum_i -(Y_i \log(\hat{y}_i) + (1 - Y_i) \log(1 - \hat{y}_i)) \quad (3.1)$$

où Y_i est la classe réelle du voxel x_i (0 pour le fond, 1 pour le cerveau) et \hat{y}_i est la probabilité que le voxel x_i appartienne au cerveau. Si des patches se superposent, \hat{y}_i est moyennée sur tous les patches concernés selon la relation :

$$\hat{y}_i = \frac{\sum_{p \in P_i} \text{Softmax}_p(y_i)}{|P_i|} \quad (3.2)$$

avec P_i étant l'ensemble des patches contenant le voxel x_i et $\text{Softmax}_p(y_i)$ étant la prédiction relative à l'appartenance au cerveau du voxel x_i du patch p , définie comme :

$$\text{Softmax}_p(y_i) = \frac{\exp^{y_i}}{\exp^{(1-y_i)} + \exp^{y_i}} \quad (3.3)$$

où y_i signifie que l'on classe le voxel x_i comme appartenant au cerveau et $1 - y_i$ comme appartenant au fond.

De plus, comme le cerveau occupe un volume relativement faible de l'IRM (question de

la représentativité des classes soulignée dans [153]), nous imposons la distribution suivante des N patches d'entraînement (Figure 3.3) :

- 85% des patches contiennent du cerveau.
- 15% des patches contiennent des parties de la tête du porcelet, externes au cerveau.

Afin d'apprendre plus précisément la délimitation du cerveau, 50% des patches du cerveau sont localisés aux contours de celui-ci.

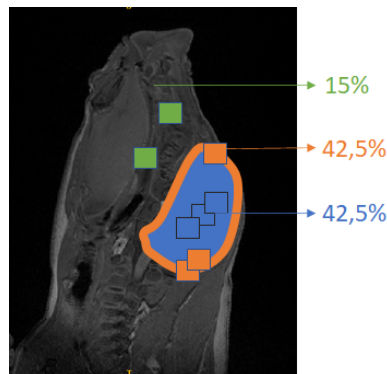


FIGURE 3.3 – Distribution des patches d'entraînement (*vert* : patches à l'extérieur du cerveau et à l'intérieur de la tête du porcelet, *bleu* : patches à l'intérieur du cerveau, *orange* : patches au niveau des contours du cerveau).

La sélection des patches pour chaque région (à partir des images annotées) est automatique et basée sur la stratégie suivante :

- **Patches externes au cerveau** : Pour chaque patch, on sélectionne aléatoirement un voxel central $O(x, y, z)$ dans la tête du porcelet. Un patch de dimension 32^3 est extrait autour de O . Si le patch contient des voxels du cerveau, il est rejeté. La sélection aléatoire est répétée jusqu'à avoir $0.15 \times N$ patches.
- **Patches internes au cerveau** : sélection aléatoire de patches dont le voxel central est situé dans un rayon de 30 voxels du barycentre du cerveau. Le premier patch extrait a comme voxel central le barycentre du cerveau du porcelet. La sélection aléatoire est répétée jusqu'à avoir $0.425 \times N$ patches.
- **Patches de contours** : Le contour du masque cérébral est récupéré sur la vérité terrain afin d'obtenir une image binaire du contour du cerveau. Pour chaque patch, on sélectionne aléatoirement un voxel central $O(x, y, z)$ dans l'image. Un patch de dimension 32^3 est extrait autour de O . Le patch est conservé si la proportion de voxels appartenant au contour du cerveau est supérieure à 1% (c.a.d. si le patch apporte de l'information sur le contour). Le contour du cerveau est large

de seulement 1 voxel, justifiant le faible pourcentage considéré (1%). La sélection aléatoire est répétée jusqu'à avoir $0.425 \times N$ patches.

3.2.2.2 Inférence

Pour l'inférence, nous considérons des patches adjacents afin de réduire les temps de calcul. Beaucoup d'études ont montré la pertinence d'utiliser des patches adjacents (Figure 3.1) à des fins de segmentation [103, 160].

Chaque IRM est divisée en patches adjacents 3D couvrant l'entièreté de l'image comme illustré sur la Figure 3.1-bas. Pour chaque voxel x_i de chaque patch p , le réseau U-Net retourne un vecteur de probabilité v_i d'appartenance à chaque classe c (fond ou cerveau). Chaque voxel est assigné à la classe de plus grande probabilité :

$$\hat{y}_i = \operatorname{argmax}_c v_i \quad (3.4)$$

Les patches 3D adjacents sont ensuite agrégés pour reconstruire le volume qui passe alors dans la chaîne de post-traitement (Figure 3.1-bas).

3.2.3 Apprentissage par transfert et analyse longitudinale

Comme nous l'avons vu dans le chapitre 1 (section 1.2.4), l'apprentissage par transfert est très courant dans l'apprentissage profond [24] et peut être utilisé pour des tâches relatives à la segmentation du cerveau [28, 154]. Nous avons vu précédemment que l'une des principales raisons de son succès est sa capacité à exploiter les connaissances (caractéristiques, poids) de modèles précédemment entraînés pour entraîner de nouveaux modèles et ainsi résoudre des problèmes tels que le manque de données pour réaliser une autre tâche similaire (Figure 1.5) conduisant souvent les modèles d'apprentissage profond à des performances moindres [24].

Cela correspond parfaitement à notre cas puisque nous allons entraîner un modèle U-Net 3D avec des données de porcelets de 15 jours et que nous voulons utiliser ce modèle entraîné pour automatiser la segmentation du cerveau de porcelets plus âgés, population pour laquelle nous disposons de moins de données (section 3.3.1). Considérons que la segmentation du cerveau des porcelets au stade $t-1$ correspond à une tâche T_{t-1} . Appelons T_t la tâche consistant à segmenter le cerveau des porcelets au stade t . En appliquant le modèle entraîné à réaliser T_{t-1} aux porcelets plus âgés, nous devrions faire face à une

dégradation des performances liée au biais du modèle sur les données d'entraînement à $t - 1$ et aux changements rapides du cerveau du porcelet.

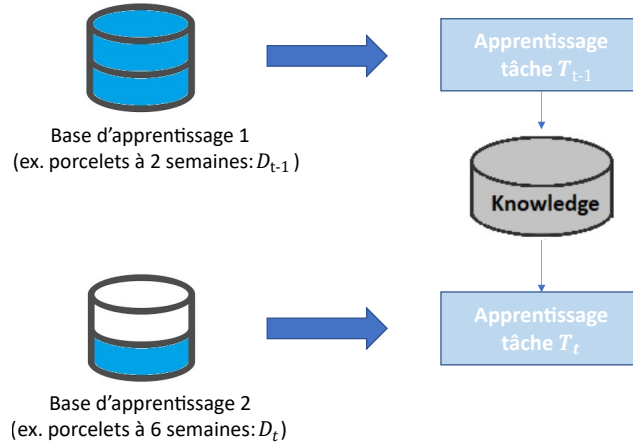


FIGURE 3.4 – Apprentissage par transfert : principe illustré sur l'application relative à l'extraction du cerveau du porcelet. t et $t - 1$ indiquent les stades de développement des porcelets étudiés.

Nous allons utiliser l'apprentissage de T_{t-1} et généraliser ses connaissances (poids du réseau entraîné) pour réaliser T_t , domaine disposant de moins de données d'entraînement, comme illustré par la Figure 3.4. Suivant la notation de [161], on définit un domaine D comme un espace de caractéristique $X = \{x_1, \dots, x_n\}$ (ensemble des niveaux de gris des voxels de l'image). Une tâche T est composée d'un espace d'étiquettes $Y = \{0, 1\}$ (1 pour le cerveau, 0 sinon) et d'une fonction de prédiction objective f_t apprise à partir des données d'entraînement au stade t (ensemble des paires $\{x_i, y_i\}$ avec $x_i \in X$, $y_i \in Y$ tels que l'indice i dénote le voxel i). Le transfert s'effectue d'un domaine et d'une tâche source vers un domaine et une tâche cible. Nous considérons le domaine source D_{t-1} comme étant celui des IRMs T1 de porcelets au stade $t - 1$ (ex. à 2 semaines). Le domaine cible D_t sera composé des IRMs T1 de porcelets au stade t (ex. à 6 semaines). Comme vu dans la section 1.2.4, nous faisons face à de l'apprentissage par transfert transductif car T_{t-1} et T_t sont similaires (segmentation du cerveau) mais D_{t-1} et D_t sont différents (stades de développement différents et donc anatomie du cerveau modifiée).

Dans la section 1.2.4, nous avons présenté plusieurs stratégies de transfert, ré-entraînant plus ou moins de couches dans le réseau. Dans notre cas, pour ne pas avoir à choisir arbitrairement les couches à ré-entraîner (typiquement en divisant le réseau entre les premières couches plus génériques et les dernières couches plus spécifiques [162]), tous les poids Θ_{t-1} sont considérés pendant le transfert : la fonction de prédiction f_t est initialisée avec les

poids Θ_{t-1} . L'apprentissage va ensuite consister à adapter les paramètres du réseau Θ_t pour minimiser la fonction de coût :

$$\Theta_t = \operatorname{argmin}_{\theta} \frac{1}{|D_t|} \sum_{i=1}^{|D_t|} L(f_t(\theta, \Theta_{t-1}, x_{i,t}), y_{i,t}) \quad (3.5)$$

$|D_t|$ est la dimension du domaine D_t et L est la fonction de coût d'entropie croisée binaire (eq 3.1). Seul f_{t_0} est initialisée avec des poids aléatoires car elle ne dispose pas d'un stade antérieur à partir duquel initialiser ses poids. Ensuite, au stade t_1 (6 semaines), la procédure d'apprentissage par transfert présentée précédemment est appliquée depuis les connaissances sur les porcelets de 2 semaines. Pour le stade t_2 , D_{t-1} désignera les porcelets de 6 semaines (t_1) et D_t désignera les porcelets de 10 semaines. Le réseau entraîné à 6 semaines (à partir des données de 6 semaines et des caractéristiques transférées des porcelets de 2 semaines) est chargé et entraîné à nouveau avec les IRMs des porcelets de 10 semaines. Cela signifie que le réseau entraîné à 10 semaines bénéficiera du transfert des caractéristiques des deux stades de développement précédents. Le mécanisme est répété à t_3 (18 semaines), assurant ainsi un suivi longitudinal automatique.

3.3 Expériences

Nous présentons dans cette section la base de données considérée (section 3.3.1), le protocole d'évaluation (section 3.3.2) ainsi que les résultats obtenus (section 3.3.3). Dans la section 3.3.3, nous dévoilons les résultats en terme d'impact de l'usage des patches 3D, de la régularisation spatiale et de l'apprentissage par transfert. Une comparaison avec d'autres techniques de l'état de l'art est également fournie.

3.3.1 Base de données

Notre méthode est évaluée sur 27 porcelets (26 mâles et 1 femelle). Ces porcelets ont été suivis pendant leurs 4 premiers mois de vie dans le cadre du projet REPAR, introduit à l'entame de la partie II. Dans le contexte de ce projet, une lésion cérébrale précoce est provoquée sur ces porcelets à 2 semaines par injection d'endothéline sous stéréotaxie. Pour caractériser l'impact de la lésion au niveau macroscopique, une IRM 3D T1 est acquise à différents stades de développement (à 2, 6, 10 et 18 semaines) pour chaque porcelet. Afin de réduire la souffrance de l'animal liée à l'IRM, une anesthésie générale par injection intramusculaire de 20 mg/kg de kétamine (Imalgene 1000) et de 2 mg/kg de

xylazine (Rompun 2%) est mise en place. En raison de la lésion induite et des sacrifices réalisés pour compléter l'étude au niveau microscopique (coupes histologiques), certains porcelets sont morts au cours de la période d'étude. Par conséquent, la quantité de données IRM disponible diminue avec l'âge des porcelets. Les données utilisées à chaque stade de développement sont détaillées dans la Table 3.1.

TABLE 3.1 – Base de données disponible à chaque stade de développement étudié.

Age (semaines)	Nombre d'IRMs 3D T1	Volume cérébral moyen
2	27	53.25 (\pm 7.20) cm^3
6	17	71.85 (\pm 10.71) cm^3
10	14	92.59 (\pm 5.79) cm^3
18	5	118.5 (\pm 13.6) cm^3

Les IRMs 3D T1 sont de dimension $256 \times 384 \times 384$ voxels avec une résolution spatiale de $0.5 \times 0.5 \times 0.5$ mm³. Le volume cérébral est évalué à partir de l'annotation manuelle du cerveau que nous avons réalisée avec le logiciel ITKSnap [146].

3.3.2 Protocole expérimental

Nos expériences ont été menées avec le même environnement que celles du chapitre 2 (voir section 2.4.1). L'implémentation de l'architecture U-Net a été réalisée avec la librairie *PyTorch* [163]. Les paramètres détaillés de l'architecture sont donnés dans la section 3.2.1. La régularisation spatiale par CRF est effectuée à l'aide de la librairie *pydensecrf*¹.

Pour les expériences relatives au premier stade de développement (c.a.d les 27 porcelets âgés de 2 semaines), nous divisons notre base de données comme suit : 7 porcelets (et donc 7 IRMs) sont utilisés pour l'entraînement, 2 porcelets pour la validation et 18 porcelets pour le test. Le modèle est entraîné par descente stochastique de gradient sur 75 époques à partir de mini-lots de taille 8 afin de minimiser la fonction de coût d'entropie croisée binaire (eq 3.1). Une stratégie de réduction du taux d'apprentissage sur plateau est utilisée avec un taux d'apprentissage initial $\Xi_0 = 0.1$ et un facteur de réduction $\sigma = 0.01$.

Pour entraîner le réseau U-Net et évaluer la qualité des segmentations obtenues, nous prenons comme référence notre segmentation manuelle du cerveau réalisée avec le logiciel ITKSnap [146]. L'évaluation des segmentations se base sur un calcul du coefficient de Dice noté DSC (eq 2.6), de la distance d'Hausdorff notée HD (eq 2.7) et de la distance de surfaces symétriques moyenne notée ASSD [164] entre la vérité terrain et le masque prédit

1. <https://github.com/lucasb-eyer/pydensecrf>

par notre méthode. La distance de surfaces symétriques moyenne entre deux ensembles I et I' est la moyenne de toutes les distances entre les points de I et le point le plus proche de I' (et inversement) comme indiqué dans l'éq 3.6. La distance entre un voxel $a \in I$ et la surface de I' est donnée par $D_{I'}(a)$.

$$\text{ASSD}(I, I') = \frac{1}{|I| + |I'|} \times \left(\sum_{a \in I} D_{I'}(a) + \sum_{b \in I'} D_I(b) \right) \quad (3.6)$$

Notons que, pour certaines expériences, nous séparons le Dice en précision et rappel. En effet, pour segmenter ensuite les structures et tissus cérébraux, un critère important est de ne perdre aucune partie du cerveau. Pour cela, nous veillons à maximiser le rappel tout en assurant une précision acceptable.

La taille des patches utilisés pour entraîner le modèle est importante car elle affecte la quantité d'informations locales prises en compte [153]. Pour cette raison, nous étudions l'impact de la taille des patches considérés en comparant l'utilisation de patches de taille 32^3 et 64^3 voxels. De même, nous comparons notre stratégie de sélection des patches d'entraînement à d'autres afin d'évaluer l'importance de la distribution des patches entre le cerveau et le reste de l'image ainsi que l'impact d'un surentraînement spécifique aux régions correspondant aux contours du cerveau. Trois stratégies sont envisagées :

- **Stratégie d'extraction 1** : Distribution entièrement aléatoire des patches.
- **Stratégie d'extraction 2** : 15% des patches externes au cerveau et 85% des patches uniformément distribués autour du barycentre du cerveau.
- **Stratégie d'extraction 3** : Notre stratégie (section 3.2.2.1) c.a.d. 15% des patches externes au cerveau, 42.5% des patches uniformément distribués autour du barycentre du cerveau et 42.5% des patches focalisés sur le contour du cerveau.

Nous étudions ensuite l'apport des opérations finales de post-traitement (régularisation spatiale avec CRF et dilatation morphologique).

Pour montrer la pertinence de l'approche par patches 3D, nous comparons les performances de notre U-Net 3D basé sur patches à l'utilisation d'un U-Net 2D comme celui considéré pour la segmentation du cerveau du fœtus dans [150] ou des rongeurs dans [155]. Comme pour notre modèle U-Net 3D, le U-Net 2D est entraîné par descente stochastique de gradient sur 75 époques à partir de mini-lots de taille 8. Pour compenser la petite taille de la base d'apprentissage, une augmentation de données (section 1.2.4) est mise en oeuvre (avec champs de biais aléatoires et mouvements aléatoires pour créer des don-

nées "réalistes"). Afin de mettre en évidence uniquement les bénéfices de l'approche par patches, indépendamment de toute régularisation spatiale, cette comparaison est effectuée avec et sans l'utilisation de CRFs.

Nous comparons également nos performances aux résultats obtenus avec l'outil *Brain Extraction Tool* (BET) de FSL². Nous avons choisi BET car il s'agit d'un outil standard pour l'extraction automatique du cerveau humain sur IRM offrant de bons résultats [136]. De plus, l'outil est basé sur des modèles déformables et ne dépend donc pas de la quantité de données disponible, contrairement à d'autres méthodes de segmentation du cerveau basées sur un atlas [148]. Nous complétons notre positionnement avec l'état de l'art en comparant notre méthode à d'autres techniques appliquées à différents modèles animaux : les macaques [148, 152], les rongeurs [148, 151, 155, 156] et les cochons [154]. Cette comparaison aux performances des travaux connexes est présentée à titre indicatif car les jeux de données sont différents et certains codes ne sont pas en libre accès [152, 154, 155].

Enfin, nous abordons la question du rapide développement du cerveau porcin au cours des premières semaines en évaluant l'efficacité de l'apprentissage par transfert pour la segmentation des cerveaux plus développés (6, 10 et 18 semaines). A chaque stade de développement t_i , nous comparons 3 stratégies afin de déterminer la plus efficace pour apprendre les variations du cerveau au fil du temps, malgré le peu d'images à disposition :

- **Stratégie d'entraînement 1** : Application du réseau entraîné à l'étape t_{i-1} (réseau entraîné sur des porcelets de 15 jours appliqué aux données à 6 semaines, réseau entraîné sur des porcelets de 6 semaines appliqué aux données à 10 semaines, etc.). La stratégie consiste à considérer que la tâche demandée est la même et que le réseau entraîné à l'étape t_{i-1} sera plus efficace en raison de la plus grande quantité de données.
- **Stratégie d'entraînement 2** : Entraînement d'un nouveau réseau avec les données disponibles au stade t_i . C'est la stratégie d'entraîner un réseau spécifique à chaque stade de développement pour faire face à la variabilité liée à l'âge, malgré le nombre décroissant de données IRM au fur et à mesure que les porcelets vieillissent.
- **Stratégie d'entraînement 3** : Notre stratégie (section 3.2.3) qui consiste à utiliser l'apprentissage par transfert pour compléter les caractéristiques apprises à t_{i-1} par les données des porcelets au stade t_i .

A chaque stade de développement, la répartition des données (section 3.3.1) utilisées pour l'entraînement, la validation et les tests est indiquée dans la Table 3.2.

2. <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET/UserGuide>

TABLE 3.2 – Répartition des données pour l’entraînement, la validation et les tests relatifs à l’apprentissage profond. Les valeurs numériques correspondent au nombre de porcelets utilisés pour chaque étape.

Âge	Entraînement	Validation	Test
2 semaines	7	2	18
6 semaines	4	2	11
10 semaines	4	2	8
18 semaines	2	1	2

3.3.3 Résultats

Les résultats présentés dans cette section, relatifs à l’utilisation de patches 3D, au post-traitement ainsi que les comparaisons avec l’état de l’art sont obtenus au stade t_0 , c’est-à-dire sur les porcelets de 2 semaines.

3.3.3.1 Taille et distribution des patches d’entraînement

Utiliser de petits patches (32^3) réduit significativement les temps de calcul comparé à l’utilisation de patches plus grands (64^3) comme indiqué dans la Table 3.3 (moyenne sur 18 porcelets). Cependant, les plus petits patches contiennent moins d’information sur le contexte ce qui dégrade les résultats (Table 3.3 et Figure 3.5). Seule la distance d’Hausdorff est améliorée par l’utilisation de petits patches.

TABLE 3.3 – Comparaison du temps d’apprentissage et des performances du réseau U-Net en fonction du nombre et de la taille des patches utilisés (moyennes sur 18 porcelets). N.A. : non disponible.

Patches	DSC	HD	ASSD	Entraînement
300x64 ³	0.918	30.23	2.68	4h30
300x32 ³	0.879	21.24	3.83	1h10
700x64 ³	N.A.	N.A.	N.A.	>7h
700x32 ³	0.951	20.24	1.49	2h30

La Table 3.3 montre également que les résultats sont meilleurs lorsque l’on augmente le nombre de petits patches par IRM à 700 (probablement en raison d’un chevauchement plus fréquent des patches). Les patches de taille 32^3 permettent d’atteindre un Dice moyen de 0.951 contre 0.918 avec les patches de taille 64^3 tout en réduisant presque de moitié le temps d’apprentissage. La distance d’Hausdorff moyenne est réduite à 20.24. Enfin, la distance de surfaces symétriques moyenne est améliorée (1.49 avec les patches de 32^3

contre 2.68 avec les patches de 64^3). Ces résultats sont illustrés par l'image à droite dans la Figure 3.5 sur laquelle le cerveau du porcelet est segmenté plus précisément.

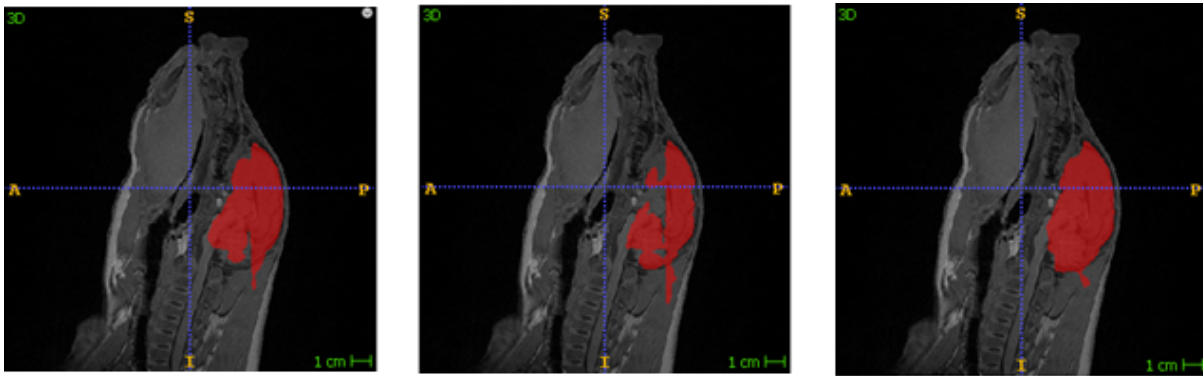


FIGURE 3.5 – Exemples de segmentations obtenues sur un porcelet en fonction de la taille et du nombre de patches utilisés (de gauche à droite : 300×64^3 , 300×32^3 , 700×32^3).

La Table 3.4 compare les performances du réseau selon la stratégie de distribution des 700 patches d'entraînement de taille 32^3 .

TABLE 3.4 – Comparaison des performances du réseau U-Net en fonction de la distribution des patches d'entraînement (moyenne sur 18 porcelets) .

	Précision	Rappel	DSC	HD	ASSD
Stratégie 1	0.861	0.653	0.757	24.42	5.81
Stratégie 2	0.899	0.911	0.905	45.70	4.31
Stratégie 3	0.959	0.943	0.951	20.24	1.49

En s'assurant que 85% des patches d'entraînement contiennent des parties du cerveau (stratégie 2), nous améliorons significativement, par rapport à la stratégie 1, la performance du réseau (+ 0.148 pour DSC, -1.50 pour ASSD). Le rappel (puisque notre objectif principal est de le maximiser) est augmenté de 0.258. Cependant, la distance d'Hausdorff moyenne est significativement dégradée en raison de l'absence de patches d'entraînement localisés en dehors de l'aire cérébrale ou au niveau des contours (manque de représentativité équilibrée). Cette dégradation significative du HD peut s'expliquer par les erreurs grossières liées à deux porcelets spécifiques (les porcelets 15 et 17), comme illustré sur le graphique droit de la Figure A.2.1 en annexe. Les erreurs concernant ces deux porcelets sont également reflétées dans le graphe de gauche par un DSC plus petit.

Selon la Table 3.4, les meilleurs résultats sont obtenus en considérant l'approche par contours (stratégie 3) car elle améliore les performances de segmentation (+0.046 pour

DSC, -2.82 pour ASSD, -25.46 pour HD par rapport à la stratégie 2). Dans l'annexe A.2.1, il est évident que la stratégie 3 (notre stratégie) offre le meilleur Dice pour tous les porcelets de la base de données test. Concernant la distance d'Hausdorff, notre stratégie donne, en moyenne, les meilleurs résultats même si, pour certains porcelets spécifiques, les performances ne sont pas aussi bonnes qu'avec la stratégie 1 (porcelets 2, 5, 7 et 15).

3.3.3.2 Post-traitement : régularisation spatiale et dilatation

Sur la Figure 3.6, nous observons que la régularisation spatiale à l'aide de champs aléatoires conditionnels (CRFs) corrige les irrégularités de la segmentation aux contours du cerveau. Selon les premières lignes de la Table 3.5, il semble que l'utilisation de CRFs soit bénéfique : elle réduit de moitié la distance d'Hausdorff et diminue la distance de surfaces symétriques moyenne. L'ajout de cette opération augmente raisonnablement le temps de traitement de l'IRM. Cependant, la régularisation spatiale dégrade le rappel ce qui diminue légèrement le Dice (0.948 contre 0.951 sans CRFs).

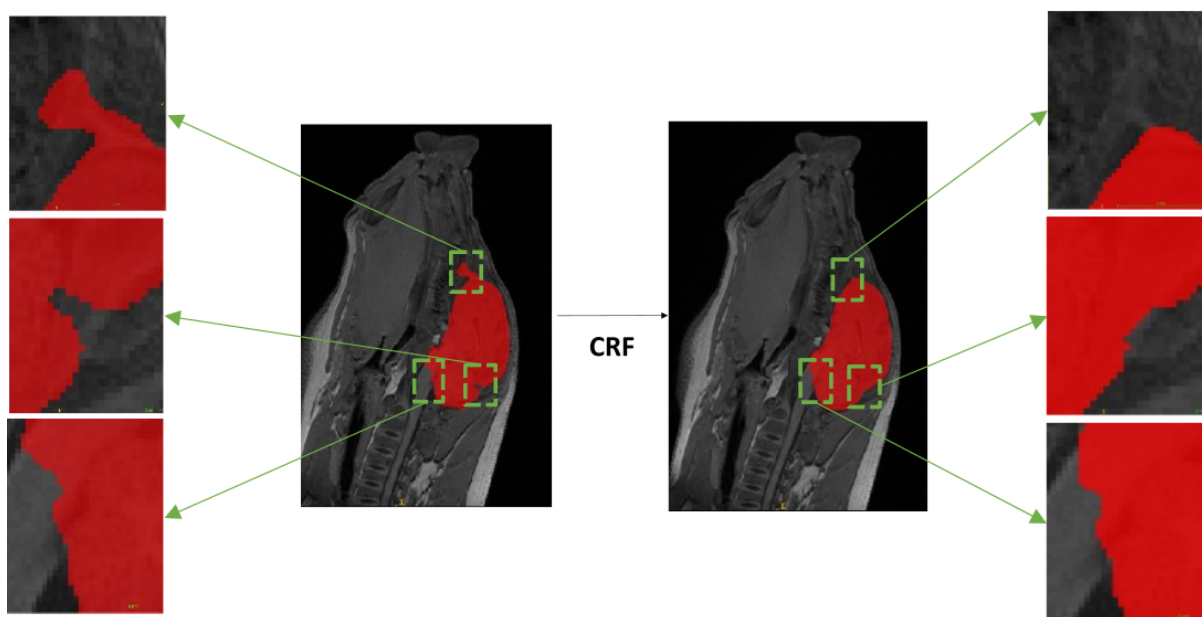


FIGURE 3.6 – Impact de la régularisation spatiale par champs aléatoires conditionnels.

L'amélioration du rappel est l'objectif de la dilatation morphologique appliquée à la fin de la chaîne de traitement. Dans la Table 3.5, nous remarquons que la dilatation augmente le rappel de 0.918 à 0.951 et réduit le HD (8.51) ainsi que l'ASSD (1.34). Ceci reflète une récupération de certaines régions du cerveau initialement perdues. On peut également noter le bon équilibre entre précision et rappel résultant de la dilatation morphologique.

TABLE 3.5 – Comparaison des performances de segmentation du cerveau obtenues avec et sans régularisation spatiale (CRF) et dilatation (Dilat.). PTI : post-traitement initial incluant conservation de la plus grande composante connexe et lissage.

	Precision	Rappel	DSC	HD	ASSD	Temps de traitement
U-Net + PTI	0.959	0.943	0.951	20.24	1.49	1min
U-Net + PTI + CRF	0.978	0.918	0.948	9.33	1.43	1min46s
U-Net + PTI + CRF + Dilat.	0.953	0.951	0.952	8.51	1.34	1min48s

3.3.3.3 Comparaison avec l'état de l'art

Tous les résultats sont reportés dans la Table 3.6. Seuls le réseau U-Net 2D et l'outil BET (2e et 3e ligne de la Table 3.6) ont été appliqués à notre jeu de données (sur les porcelets de 2 semaines) et fournissent donc une réelle comparaison avec notre méthode. Le jeu de données de 27 porcelets comprend les 7 porcelets d'entraînement, les 2 porcelets de validation et les 18 porcelets de test. Les autres résultats (lignes 4 à 9) ont été obtenus sur des jeux de données différents. Les premières lignes du tableau montrent que notre méthode est plus performante que le réseau U-Net 2D et BET pour la segmentation du cerveau de porcelets âgés de 2 semaines. Ces résultats sont détaillés ci-après.

TABLE 3.6 – Comparaison de la méthode proposée avec des algorithmes d'extraction du cerveau de l'état de l'art (T.L. : apprentissage par transfert depuis l'humain).

Animal	Méthode	Taille du jeu de données	DSC	HD
Porcelets	Notre méthode	27	0.952	8.51
Porcelets	BET	27	0.764	25.91
Porcelets	2D U-Net	27	0.919	11.06
Macaques	DIKA-Nets [152]	155	0.964	1.47
Macaques/Rongeurs	atlasBREX [148]	8/6	0.950	-
Rongeurs	PCNN3D [151]	6	0.930	-
Rongeurs	2D U-Net [155]	132	0.940	6.81
Rongeurs	RATS [156]	22	0.920	13.6
Cochons	T.L. [154]	3	0.930	-

La Table 3.7 rapporte les résultats moyens obtenus sur les 18 porcelets de test avec notre stratégie et avec un U-Net 2D classique tel qu'il est utilisé dans [150, 155]. L'approche 2D sans post-traitement produit une meilleure distance d'Hausdorff (13.48) qu'un U-Net 3D basé sur des patches (20.24). Cependant, l'approche 3D offre un Dice plus élevé (0.951 contre 0.919) et une distance de surfaces symétriques moyenne plus faible (1.49 contre 2.16). Ces constatations sont visibles également sur la Figure 3.7. En effet, nous observons avec le porcelet 1 que l'approche 2D peut perdre certaines parties du cerveau (ici la partie supérieure) contrairement à l'approche 3D basée sur des patches. Néanmoins, les deux

porcelets illustrent les irrégularités produites par l'approche 3D (à l'avant du cerveau en particulier et également à l'arrière pour le porcelet 1) qui peuvent considérablement altérer le HD comme l'indiquent les résultats numériques.

TABLE 3.7 – Comparaison des performances du réseau U-Net 2D et de notre réseau U-Net 3D basé sur des patches (moyenne calculée sur les 18 porcelets de test). PTF : post-traitement final incluant régularisation spatiale et dilatation. La conservation de la plus grande composante connexe et le lissage sont appliqués dans tous les cas.

	DSC	HD	ASSD
2D U-Net	0.919	13.48	2.16
3D U-Net avec patches	0.951	20.24	1.49
2D U-Net avec PTF	0.919	11.06	2.14
3D U-Net avec patches et PTF	0.952	8.51	1.34

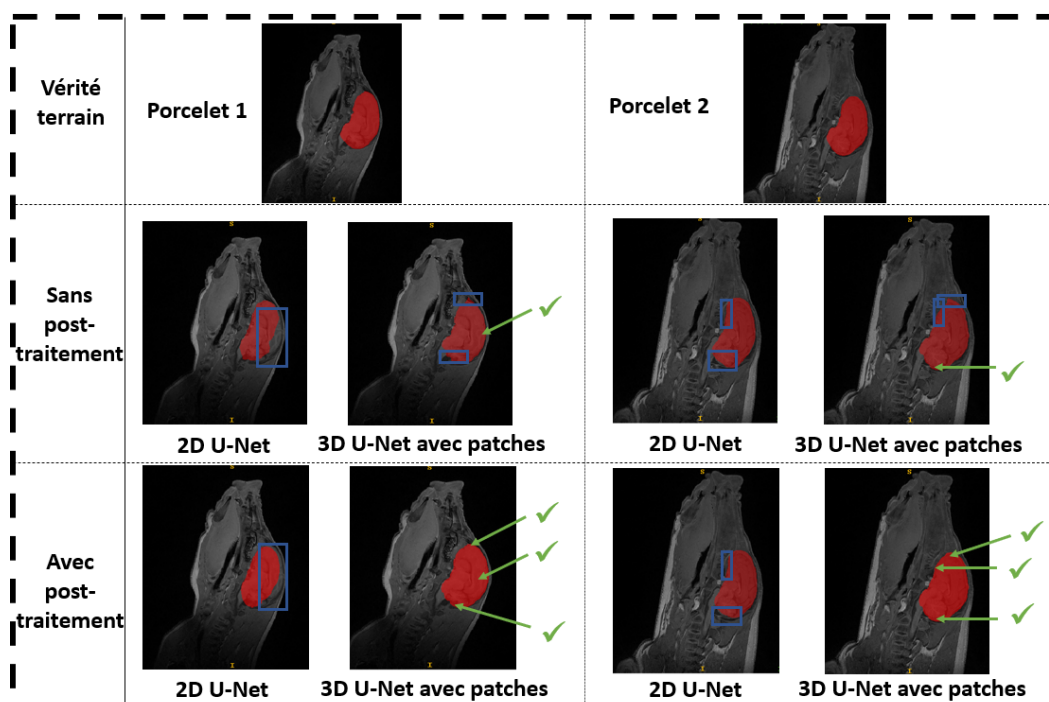


FIGURE 3.7 – Exemples de segmentations obtenues sur deux porcelets à l'aide d'un U-Net 2D et d'un U-Net 3D à base de patches. Le post-traitement inclut la régularisation spatiale et la dilatation.

L'ajout des dernières étapes du post-traitement (régularisation spatiale par CRF et dilatation) ne modifie que très légèrement les performances de la stratégie en 2D avec une réduction de la distance d'Hausdorff à 11.06. A l'inverse, l'application de ce post-traitement à l'approche 3D basée sur patches divise par 2 la distance d'Hausdorff moyenne

et réduit l'ASSD. Sur la Figure 3.7, on peut observer sur la dernière ligne les effets du post-traitement. Ce dernier ne corrige pas les erreurs de segmentation de l'approche 2D sur le porcelet 1. Par contre, il lisse les irrégularités de la segmentation obtenue par notre méthode, entraînant une nette réduction du HD. De cette manière, notre proposition surpasse nettement l'approche 2D, considérée pour les rongeurs (ligne 7 de la Table 3.6).

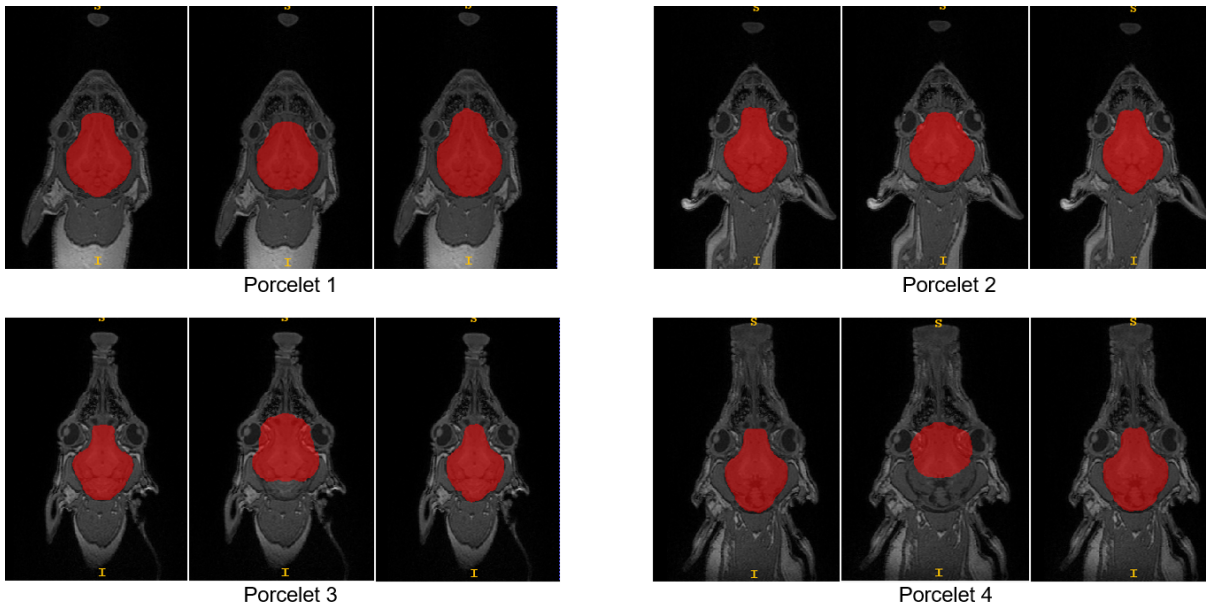


FIGURE 3.8 – Exemples de segmentations en vue coronale pour 4 porcelets de la base de données. De gauche à droite : vérité terrain, BET, notre méthode.

La Table A.2.3 fournie en annexe donne les résultats obtenus avec notre méthode et avec BET sur les 18 porcelets de test. Notre méthode apparaît plus performante que BET pour la plupart des porcelets, avec une amélioration de toutes les mesures considérées (DSC, HD et ASSD). BET obtient une meilleure distance d'Hausdorff que notre méthode pour un seul porcelet (porcelet 13) avec une légère différence (13.96 avec BET contre 14.32 avec notre méthode). Par ailleurs, à l'exception du porcelet 3, le volume cérébral estimé par notre méthode de segmentation est plus proche du volume cérébral réel (mesuré à partir de la vérité terrain). La Figure 3.8 donne des exemples de segmentations obtenues. Nous constatons que notre méthode parvient à segmenter le cerveau du porcelet avec plus de précision que BET, qui a tendance à perdre des parties ou à incorporer des zones externes dans le masque du cerveau (porcelets 3 et 4).

Concernant les autres méthodes dont les performances indiquées concernent d'autres jeux de données (lignes 4 à 9 de la Table 3.6), seul DIKA-Nets, développé pour le macaque,

surpasse significativement notre méthode (DSC de 0.964 et HD de 1.47). Néanmoins, le jeu de données utilisé est beaucoup plus important (155 IRMs de macaques) que celui disponible pour les porcelets (27 IRMs seulement). La seule autre méthode surpassant notre proposition, en termes de HD, est le réseau U-Net 2D appliqué aux rongeurs (HD de 6.81). Cependant, cette étude dispose également d'un plus grand jeu de données (132 IRMs) et nous avons précédemment montré que l'approche 2D est moins efficace sur les porcelets que l'approche 3D basée sur patches. Notre approche atteint des performances équivalentes à celles basées sur un atlas (atlasBREX) et dépasse RATS avec un ensemble de données comparable (22 IRMs). En ce qui concerne l'application sur les porcs, notre méthode offre de meilleures performances que la solution basée sur le transfert des caractéristiques du cerveau humain [154].

3.3.3.4 Apprentissage par transfert pour l'étude longitudinale

TABLE 3.8 – Qualité des segmentations du cerveau à chaque stade de développement en fonction de la stratégie d'entraînement utilisée (Stratégie 1 : U-Net entraîné au stade t_{i-1} , Stratégie 2 : U-Net entraîné seulement au stade t_i , Stratégie 3 : transfert des caractéristiques du réseau entraîné à t_{i-1} et complété par les données disponibles au stade t_i). Les valeurs moyennes sont calculées sur les porcelets de test (c.a.d. 18 porcelets à t_0 , 11 porcelets à t_1 , 8 porcelets à t_2 et 2 porcelets à t_3). La stratégie 2 n'est pas appliquée au stade t_3 à cause du trop petit nombre de données disponibles.

	Stratégie 1			Stratégie 2			Stratégie 3		
	DSC	HD	ASSD	DSC	HD	ASSD	DSC	HD	ASSD
t_0 : 2 sem.	-	-	-	0.952	8.51	1.34	-	-	-
t_1 : 6 sem.	0.891	15.42	3.14	0.636	82.58	42.34	0.934	10.72	1.99
t_2 : 10 sem.	0.894	31.09	3.60	0.907	44.76	4.15	0.956	16.17	1.53
t_3 : 18 sem.	0.954	12.86	1.62	-	-	-	0.958	12.20	1.47

La Table 3.8 met en avant les performances obtenues par les 3 stratégies d'apprentissage (présentées dans la section 3.2.3) pour segmenter le cerveau des porcelets à chaque stade de développement, c'est-à-dire à 6, 10 et 18 semaines. La première ligne, correspondant au stade t_0 (2 semaines), est donnée comme référence. Comme il s'agit du premier stade étudié, les stratégies 1 et 3 ne sont pas appliquées. La première colonne (Stratégie 1) montre les résultats moyens obtenus à chaque stade de développement t_i en appliquant le réseau U-Net 3D entraîné à l'étape t_{i-1} . Pour les stades t_1 et t_2 , on observe une altération des performances de segmentation par rapport aux résultats de référence obtenus à t_0 (première ligne). En effet, toutes les métriques sont dégradées (-0.061 pour DSC, +6.91

pour HD, +1.8 pour ASSD à t_1 et -0.058 pour DSC, +22.58 pour HD, +2.26 pour ASSD à t_2). Cette baisse de performance est confirmée par la Figure 3.9 appliquant le réseau U-Net entraîné à t_0 au même porcelet à t_0 (15 jours) et à t_1 (6 semaines). Notre réseau est toujours capable de localiser le cerveau dans l'IRM à 6 semaines mais, il perd des parties importantes de celui-ci (en particulier l'arrière du cervelet ici, entouré dans la Figure 3.9).

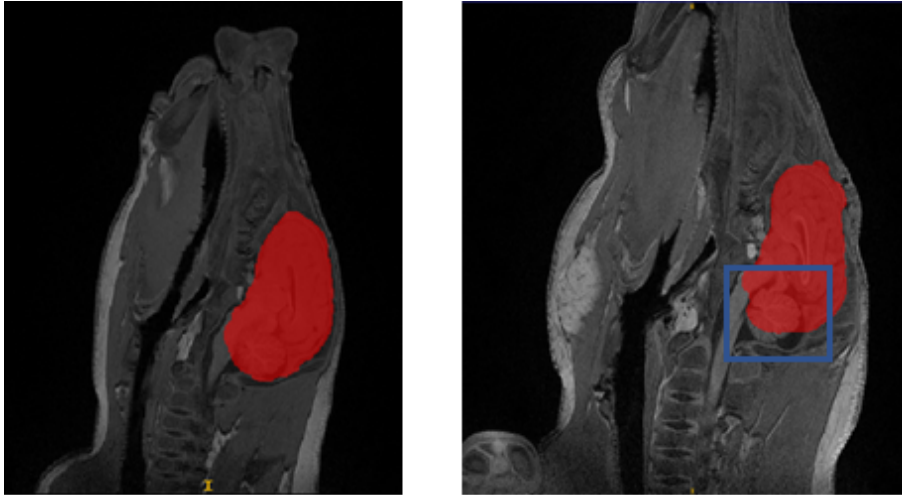


FIGURE 3.9 – Exemples de segmentations obtenues pour le même porcelet à 15 jours (à gauche) et à 6 semaines (à droite) avec le réseau U-Net 3D basé sur patches entraîné sur les données de 15 jours.

Ces résultats confirment notre hypothèse selon laquelle les caractéristiques apprises à un stade (intensité des tissus, taille du cerveau, sinus périphériques) varient rapidement pour un même porcelet au cours de son développement cérébral réduisant les performances du réseau. Cette variation en termes de contraste est visible (notamment dans la région du cervelet) dans la partie supérieure de la Figure 3.1 à travers les agrandissements du cerveau fournis à chaque stade de développement. On peut cependant noter que les résultats obtenus à 18 semaines (stade t_3) avec la stratégie 1 sont équivalents aux performances obtenues à t_0 . Ces bons résultats peuvent traduire un ralentissement du développement cérébral du porcelet entre 10 et 18 semaines.

Si l'on considère maintenant uniquement les données disponibles au stade t_i pour entraîner le modèle U-Net (stratégie 2 - Table 3.8), on observe que les résultats se dégradent considérablement (-0.255 pour DSC, +67.16 pour HD, +39.20 pour ASSD à t_1 et -0.045 pour DSC, +36.25 pour HD, +2.81 pour ASSD à t_2). Les très mauvais résultats à 6 semaines s'expliquent par une mauvaise identification de la région cérébrale sur 3 des 11 porcelets de test (voir annexe A.2.2). En effet, ces 3 porcelets ont un Dice nul et un ASSD

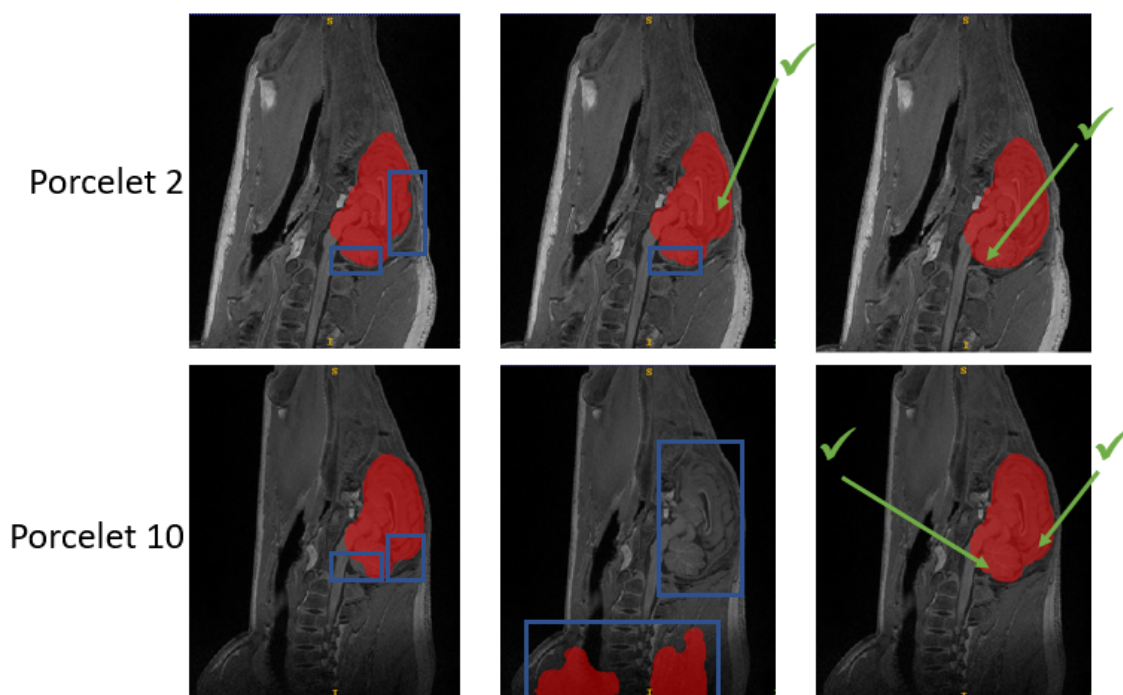


FIGURE 3.10 – Comparaison de segmentations obtenues sur deux porcelets de 6 semaines en fonction de la stratégie d’apprentissage considérée (de gauche à droite : réseau entraîné à 2 semaines, réseau entraîné à 6 semaines, apprentissage par transfert).

supérieur à 100, ce qui aggrave les résultats moyens. On peut voir sur la 2e ligne de la Figure 3.10, correspondant au porcelet 10, que le réseau entraîné à 6 semaines détecte un cerveau dans le bas de l’image en raison de la présence de contrastes similaires. Ces erreurs grossières sont très probablement dues à un manque de données d’entraînement (seulement 4 porcelets). Néanmoins, lorsque le réseau entraîné avec les données de 6 semaines ne fait pas de grosse erreur pour localiser le cerveau, les résultats sont la moitié du temps améliorés (porcelets 1, 2, 6 et 8) comme illustré dans l’annexe A.2.2. Sur la Figure 3.10, on observe que le réseau entraîné à 6 semaines récupère certaines régions du cerveau du porcelet 2 par rapport au réseau entraîné à 15 jours. Ces résultats confirment l’importance d’exploiter le petit nombre de données à 6 semaines pour tenir compte des variabilités morphologiques liées à l’âge.

Ainsi, la Table 3.8 montre que l’apprentissage par transfert améliore la moyenne de toutes les métriques par rapport aux stratégies 1 et 2. Cette amélioration est valable pour tous les stades de développement et fournit un suivi longitudinal satisfaisant pour segmenter fidèlement le cerveau du porcelet au cours de son développement (Figure 3.11).

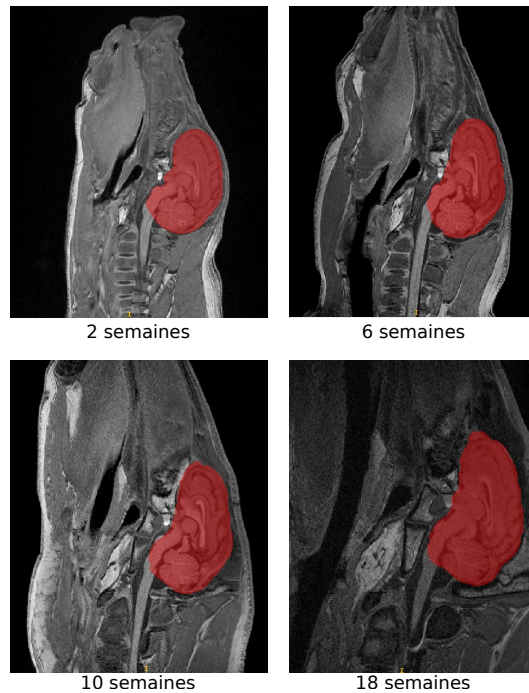


FIGURE 3.11 – Segmentation du cerveau du porcelet à chaque stade de développement obtenue avec l'apprentissage par transfert (stratégie 3).

Prenons le stade t_1 (6 semaines) pour illustrer ces améliorations. L'annexe A.2.2 met en évidence l'amélioration de toutes les métriques sur les porcelets de test. Nous pouvons également constater la nette amélioration des segmentations du cerveau avec la Figure 3.10 où, pour les deux porcelets considérés, la segmentation obtenue par apprentissage par transfert (image de droite) est celle qui englobe le cerveau avec le plus de précision.

Nous pouvons terminer cette étude en soulignant les résultats légèrement inférieurs obtenus sur les porcelets plus âgés (malgré l'utilisation de l'apprentissage par transfert) par rapport à ceux obtenus à 2 semaines. Pour chaque stade de développement, la distance d'Hausdorff et la distance de surfaces symétriques moyenne sont légèrement dégradées (HD de 10.72 à t_1 , 16.17 à t_2 et 12.20 à t_3 contre 8.51 à t_0 , ASSD de 1.99 à t_1 , 1.53 à t_2 et 1.47 à t_3 contre 1.34 à t_0). Il s'agit toutefois de résultats satisfaisants qui pourraient être améliorés avec plus de données IRM à chaque stade de développement. On peut également noter la très légère amélioration du Dice à t_2 et t_3 .

3.4 Discussion

A travers nos expériences, nous avons montré la capacité d'un réseau U-Net basé sur des patches 3D à segmenter précisément le cerveau de porcelets sur IRM T1, en concentrant les patches d'entraînement sur les contours du cerveau. La méthode proposée a l'avantage de fonctionner à partir d'une base d'apprentissage relativement petite (seulement 7 porcelets d'entraînement à 2 semaines), ceci apparaissant comme un avantage considérable pour les études relatives aux modèles de gros animaux pour lesquels peu de données sont disponibles. Nous avons également trouvé un équilibre dans la taille des patches afin de combiner les informations liées au contexte global (l'ensemble des patches couvre une large partie de l'IRM et surtout de la tête du porcelet) et aux informations locales (propres à chaque patch) dans un temps de calcul acceptable. Le grand nombre de patches d'entraînement extraits par IRM (700) compense la taille modeste du jeu de données tout en garantissant des chevauchements plus systématiques, rendant la prédiction plus précise. Contrairement à d'autres travaux, aucune étape de pré-traitement n'est nécessaire (ex. champ de biais dans [154]) et aucune information *a priori* n'est requise (ex. carte de distances au barycentre et carte de distances signées dans [152]).

Par ailleurs, notre méthode parvient à corriger les erreurs de segmentation du modèle U-Net liées au petit nombre de données grâce à une chaîne de post-traitement incluant des opérations de traitement d'images et de régularisation spatiale par CRF. Il convient de noter que les CRFs pourraient être entièrement intégrés dans le modèle U-Net, comme proposé dans [165], afin de bénéficier de la rétro-propagation pendant l'apprentissage. Le post-traitement serait ainsi directement appliqué sur un masque régularisé. Il serait aussi envisageable d'utiliser des patches se chevauchant plutôt que des patches adjacents lors de l'inférence afin de compenser l'incertitude de la prédiction liée aux effets de bord et ainsi réduire les irrégularités de la segmentation [153].

Nous avons montré que l'association du réseau U-Net basé sur des patches 3D avec le post-traitement final donne de meilleurs résultats que les approches 2D classiques [150, 155]. Par l'ensemble de ces propriétés, notre solution dépasse incontestablement, dans le cas du porcelet, des outils d'extraction de cerveau spécifiques à l'homme tels que BET. Plus généralement, nous avons vu que notre approche semble faire mieux que d'autres méthodes sans atlas proposées pour divers animaux (PCNN3D, RATS). Elle offre même des performances légèrement supérieures à celles de techniques basées sur atlas comme atlasBREX. Cependant, la différence de résultats avec DIKA-Nets, princi-

pablement due, selon nous, à la plus petite taille de notre base de données, suggère de possibles améliorations (ex. augmenter la taille de la base d'apprentissage) pour réduire la distance d'Hausdorff.

En outre, nous proposons une solution pour l'étude longitudinale du cerveau porcin grâce à l'apprentissage par transfert. Les performances réduites du réseau entraîné au stade $t - 1$ sur des porcelets au stade de développement t (stratégie 1 de la section 3.2.3) confirment l'évolution rapide du cerveau du porcelet au cours des premières semaines de vie [131]. Notons que cette dynamique du développement cérébral est déjà observable sur les porcelets de 2 semaines étudiés dont les âges fluctuent de 13 à 16 jours avec des variations de volumes cérébraux allant de 46.05cm^3 à 59.64cm^3 (annexe A.2.3). Nous obtenons, en moyenne, de meilleurs résultats sur les porcelets dont le stade de développement est plus proche des données d'entraînement (volume cérébral moyen de 53.25cm^3). Évidemment, ces variations dans le développement sont également propres à chaque individu. Malgré une base d'apprentissage très petite pour les porcelets plus âgés (seulement 4 porcelets à 6 semaines et 10 semaines, 2 porcelets d'entraînement à 18 semaines), nous parvenons à obtenir de bonnes performances de segmentation à partir du réseau entraîné au stade de développement précédent. Le succès de l'apprentissage par transfert proposé peut s'expliquer par la similarité de la structure globale du cerveau et de la proportion de tissus dans le cerveau des porcelets, quel que soit leur âge. Les caractéristiques apprises à l'état $t - 1$ restent en partie valables. Les variabilités morphologiques liées à la croissance (intensités, volume) sont apprises à partir du petit nombre de porcelets d'entraînement disponibles au stade de développement étudié. Les résultats de la Table 3.8 au stade t_3 (18 semaines) suggèrent un ralentissement du développement du cerveau du porcelet entre 10 et 18 semaines. En effet, l'application du réseau entraîné à 10 semaines sur les porcelets de 18 semaines offre de bonnes performances (DSC : 0.954, HD : 12.86, ASSD : 1.62), presque équivalentes à celles obtenues avec l'apprentissage par transfert. Il serait intéressant de poursuivre l'étude à des stades de développement ultérieurs afin de confirmer cette hypothèse et de définir plus précisément l'âge limite à partir duquel le développement cérébral du porcelet se ralentit et ne requiert plus nécessairement de nouvelles données d'entraînement.

Il reste à discuter certains aspects de nos expériences. Tout d'abord, les performances ont été évaluées sur un ensemble de données de test relativement restreint de 39 IRMs (18 à 2 semaines, 11 à 6 semaines, 8 à 10 semaines et 2 à 18 semaines). Néanmoins, notre jeu de données complet est du même ordre de grandeur que d'autres travaux portant sur les

porcelets. L'étude de [131] implique 40 IRMs de porcelets, celle de [135] utilise seulement 15 porcelets tandis que nous exploitons au total 27 IRMs à 2 semaines, 17 IRMs à 6 semaines, 14 IRMs à 10 semaines et 5 IRMs à 18 semaines. Les études néonatales sur le modèle de gros animal étant restrictives, comme souligné précédemment, nous pouvons considérer que notre base de données est satisfaisante.

Enfin, il est à noter que notre méthode est appliquée uniquement sur le porcelet. Dans de futurs travaux, il serait intéressant de l'évaluer sur d'autres modèles animaux d'intérêt mais difficiles à analyser tels que le mouton, le furet ou le chien.

3.5 Conclusion

Dans ce chapitre, nous avons proposé une méthode originale pour extraire par segmentation le cerveau du porcelet sur IRM T1 à l'aide d'une architecture U-Net considérant des patches 3D, extraits selon une stratégie appropriée. Les résultats moyens de segmentation obtenus sont prometteurs (DSC de 0.953, HD de 8.51 et ASSD de 1.34), surpassant des approches 2D ou d'autres techniques implémentées pour des espèces différentes (ex. humains (BET), rongeurs (PCNN3D, RATS)). Nous avons également traité la question de l'évolution rapide du cerveau dans les premiers mois de vie en utilisant une stratégie d'apprentissage par transfert assurant un suivi longitudinal du cerveau. La méthode a été évaluée à 4 stades de développement du porcelet (2, 6, 10 et 18 semaines). Il est à souligner que ces travaux ont fait l'objet d'une communication à la conférence nationale *Orasis* [166] et d'un article dans le journal *Computer Methods and Programs in Biomedicine* [158].

Avec cette méthode d'extraction du cerveau du porcelet, transposable nous l'espérons à d'autres modèles de gros animaux, nous progressons dans la chaîne de traitement relative à l'utilisation de modèles animaux pour comprendre les mécanismes *in vivo* de la lésion cérébrale précoce (Figure II.1). Les étapes suivantes consisteront à développer des algorithmes automatisant la segmentation fidèle des structures et tissus cérébraux du porcelet (amélioration des performances obtenues avec les atlas existants [131, 135]). Cette segmentation est indispensable à l'analyse des caractéristiques cérébrales après lésion, après traitement, etc. Du fait de la difficulté à distinguer les structures dans le cerveau porcin et pour d'autres motifs externes à notre responsabilité, nous sommes passés à des études sur l'humain pour comprendre les conséquences de la lésion cérébrale précoce sur la morphométrie cérébrale et sur la motricité. C'est ce que nous abordons dans les chapitres suivants en nous appuyant sur le modèle de l'AVC néonatal chez l'enfant.

À retenir

- L'extraction du cerveau est une opération complexe sur le porcelet, modèle animal adapté à l'étude du développement neurologique précoce. Pourtant, il s'agit d'une étape cruciale à la segmentation des structures et tissus cérébraux nécessaire pour caractériser les mécanismes de la lésion cérébrale précoce.
- La segmentation du cerveau du porcelet sur IRM avec un modèle U-Net 3D basé sur des patches extraits de façon spécifique surpasse d'autres solutions telles que l'apprentissage profond en 2D ou des outils standards spécifiques au cerveau humain tels que BET.
- L'utilisation de patches 3D se chevauchant pendant l'apprentissage enrichit les informations contextuelles des données d'apprentissage tout en réduisant la capacité de mémoire requise par rapport à une approche 3D traditionnelle.
- L'apprentissage par transfert permet de transférer les performances d'un réseau entraîné sur de jeunes porcelets à un stade plus avancé du développement cérébral, facilitant ainsi une étude longitudinale automatique, même avec des données moins représentatives.

SEGMENTATION DES NOYAUX GRIS CENTRAUX DE L'ENFANT APRÈS AVC NÉONATAL : EXPLOITATION DES CONNAISSANCES STRUCTURELLES

4.1 Introduction

Nous avons vu dans l'introduction de la partie II que l'analyse des noyaux gris centraux chez l'enfant après AVC néonatal, et donc avec lésion cérébrale précoce, est nécessaire de par leur rôle majeur dans les fonctions sensori-motrices [115] et leur forte croissance durant les premières années de vie [116]. Ces structures sont d'ailleurs souvent étudiées dans le contexte de la lésion cérébrale précoce [117, 118, 119]. Dans notre chaîne de traitement, relative à l'utilisation du modèle de l'AVC néonatal pour étudier la lésion cérébrale précoce à partir d'IRMs (Figure II.1), nous observons qu'une étape préliminaire et cruciale à l'analyse des noyaux gris centraux consiste à les segmenter. Or, nous avons vu qu'il s'agit d'une opération difficile, spécialement dans le cas de l'enfant après AVC néonatal. Nous rappelons ici brièvement les raisons de cette complexité.

Tout d'abord, les noyaux gris centraux sont des structures petites et profondes localisées dans une région du cerveau avec de faibles contrastes tissulaires sur IRM [22]. Ensuite, la segmentation de ces structures est encore plus complexe chez le jeune enfant. En effet, à ce stade précoce, le cerveau de l'enfant évolue très rapidement ce qui modifie à la fois la morphologie des structures cérébrales mais également le contraste tissulaire sur IRM (contraste très peu prononcé avant l'âge de 2 ans). En plus des problèmes mentionnés précédemment, on trouve des complications liées à la faible quantité de données à disposition. En effet, les cohortes d'enfants sont souvent plus réduites que les cohortes d'adultes pour des raisons éthiques combinées à la difficulté à mettre en place des protocoles d'acqui-

tion d’IRM. La rareté de la pathologie étudiée (prévalence de l’AVC néonatal présentée en introduction de partie II) rend les données encore moins fréquentes. Enfin, les enfants ont davantage tendance à bouger durant l’acquisition IRM, ce qui réduit d’autant plus la quantité d’images de bonne qualité exploitables pour l’analyse.

Dans le chapitre 1 de cette thèse, nous avons montré que les techniques d’apprentissage profond avaient permis d’atteindre une efficacité remarquable ces dernières années pour différentes tâches de segmentation d’images [11], notamment grâce à la montée en puissance des CNNs (section 1.2.2) et des mécanismes d’attention (section 1.2.3). Ainsi, de nombreuses architectures CNN [22, 123, 124] et logiciels [121, 122] ont été développés pour la segmentation des structures subcorticales du cerveau. Si ces techniques ont permis d’améliorer la précision de segmentation des noyaux gris centraux, elles se montrent moins efficaces dans le cas de l’enfant avec lésion cérébrale précoce. D’une part, les petites bases de données à disposition empêchent un apprentissage correct des CNNs (tendance au sur-entraînement). D’autre part, les modifications du cerveau induites par la lésion (variation d’intensité, modification de la structure générale du cerveau), illustrées sur la Figure II.2, perturbent ces méthodes qui produisent alors des erreurs de classification. Certaines techniques parviennent à gérer la segmentation en présence de lésion de la substance blanche [122], mais ce n’est pas encore le cas pour les lésions corticales (type AVC néonatal). Par conséquent, la segmentation des noyaux gris centraux dans les cerveaux avec lésion cérébrale précoce (particulièrement après AVC néonatal), se fait manuellement [118], ceci étant une tâche fastidieuse et chronophage, nécessitant une expertise humaine.

Dans ce chapitre, nous proposons d’appliquer la méthode de post-traitement introduite dans le chapitre 2, complémentaire aux CNNs existants, pour améliorer la segmentation des noyaux gris centraux dans le cerveau de l’enfant, y compris en présence d’une lésion corticale. Nous prenons l’exemple d’une population d’enfants de 7 ans, parmi lesquels certains ont eu un infarctus cérébral artériel néonatal (NAIS), car il s’agit du type d’AVC néonatal le plus courant [167]. Nous espérons que la méthode sera transposable à d’autres types de lésions cérébrales précoces. Nous pensons que la méthode de segmentation sémantique précédemment introduite, basée sur les informations structurelles des régions de l’image, sera efficace du fait de l’organisation spatiale cohérente connue des noyaux gris centraux [115], non explicitement exploitée par les CNNs restant limités au champ réceptif (chapitre 1). Nous souhaitons ainsi corriger les incohérences structurelles de segmentation fournies par les logiciels et algorithmes existants grâce aux connaissances structurelles de haut-niveau représentées sous forme de graphes. Comme dans le chapitre 2, les noeuds

représentent les régions détectées par le CNN (et portent des informations relatives à la carte de segmentation produite par le CNN) tandis que les arêtes indiquent des relations spatiales entre ces régions. Ces graphes sont ensuite exploités par un réseau de neurones sur graphes (GNN) pour réaliser une tâche de classification de noeuds. Pour les raisons mentionnées dans le chapitre 2, l’opérateur de convolution sur graphe ECConv [68] est considéré. Concernant l’importance de la topologie des graphes, soulignée également dans les chapitres 1 et 2, nous proposons une solution alternative. Contrairement au chapitre 2, nous choisissons de travailler avec des graphes complets (c.a.d. où tous les noeuds sont reliés entre eux par une arête) et d’ajuster les poids de chaque attribut d’arêtes lors de l’entraînement du modèle GNN par une fonction supplémentaire dans l’opérateur de convolution lui-même (section 1.4.3). Ce choix se justifie par une volonté de ne pas élaguer les arêtes sur un critère spatial. En effet, pour segmenter le putamen droit par exemple (voir Figure II.2), la localisation du putamen gauche, bien que plus éloignée, est aussi importante (car normalement en face) que celle du pallidum droit collé au putamen. Enfin, nous avons vu dans le chapitre 2 que notre méthode était robuste aux petits jeux de données. Dans ce contexte d’étude de la lésion cérébrale précoce, où les données sont rares comme expliqué précédemment, nous étudions l’évolution des performances de la méthode avec la réduction de la taille de la base d’apprentissage. L’idée est de se projeter sur des applications médicales pour lesquelles la quantité de données disponibles serait encore plus faible.

Dans la suite de ce chapitre, nous expliquerons les modifications faites à la méthode du chapitre 2 (section 4.2) avant de présenter notre base de données, les expériences menées et les résultats obtenus (section 4.3). Nous achèverons le chapitre par une discussion sur les performances atteintes.

4.2 Adaptation de la méthode de segmentation sémantique

La Figure 4.1 fournit une vue générale de la méthode proposée (très similaire à celle présentée dans le chapitre 2). En effet, comme dans le chapitre 2, un CNN est entraîné à partir d’une base de données annotée (Figure 4.1-A) à segmenter les noyaux gris centraux de cerveaux sains et de cerveaux présentant une lésion cérébrale précoce (corticale). Un GNN est également entraîné avec le jeu de données annoté à classer les noeuds des graphes construits à partir de l’image segmentée par le CNN. La seule différence repose

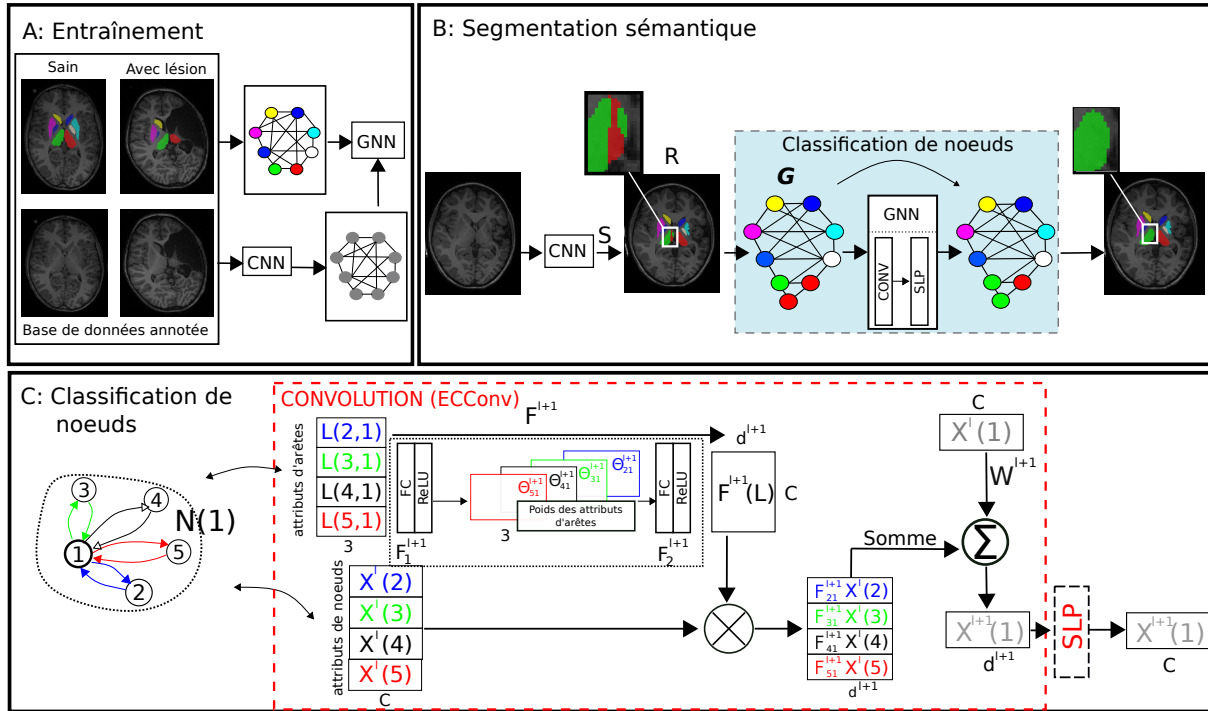


FIGURE 4.1 – Vue générale de la méthode proposée. Par souci de clarté, toutes les arêtes des graphes complets ne sont pas représentées.

sur le fait que les graphes construits sont complets. Ainsi, aucun apprentissage de rayon d'élagage (paramètre ρ du chapitre 2) n'est requis. La disparition du rayon d'élagage est compensée par l'ajout d'une couche supplémentaire dans l'opérateur de convolution (fonction F_1 sur la Figure 4.1) détaillée ultérieurement. A la rencontre d'une nouvelle IRM (Figure 4.1-B), le CNN entraîné fournit une première segmentation des noyaux gris centraux pouvant contenir des erreurs (ex. thalamus gauche détecté dans le thalamus droit). A partir de cette proposition de segmentation, un graphe complet est construit. Le GNN entraîné est alors utilisé pour classifier les noeuds du graphe en s'appuyant sur les attributs des noeuds et des arêtes (Figure 4.1-C). Une fois les noeuds classifiés, les régions sont finalement identifiées.

La méthode étant similaire à celle du chapitre 2, nous nous concentrons sur les nouveautés apportées concernant la construction des graphes et l'architecture du GNN.

La construction du graphe valué $G = (V, E, X, L)$ à partir de l'image segmentée par le CNN ne change pas. Ainsi, chaque noeud $v \in V$ correspond à une région $R_v \in R$ (Figure 4.1) détectée par le CNN et la fonction $X \rightarrow \mathbb{R}^C$ de G attribue à chaque noeud v le vecteur

S_v (eq 2.1) de probabilité d'appartenance moyenne des voxels de R_v à chaque classe $c \in C$. La différence dans la construction du graphe est liée aux arêtes puisque nous considérons des graphes complets, c'est-à-dire que tous les noeuds sont connectés entre eux. Pour cette application, la fonction $L : E \rightarrow \mathfrak{R}^3$ assigne à chaque arête (u, v) la position relative des barycentres des 2 régions connectées R_u et R_v selon les 3 directions (gauche-droite, antérieure-supérieure et inférieure-supérieure), normalisée par la plus grande distance D observée dans l'IRM (la diagonale de l'IRM) :

$$L((u, v)) = \left[\frac{d_x^{R_v, R_u}}{D}, \frac{d_y^{R_v, R_u}}{D}, \frac{d_z^{R_v, R_u}}{D} \right] \quad (4.1)$$

L'architecture du GNN considérée est similaire à celle du chapitre 2 et se décompose en deux couches : une couche de convolution et un perceptron à couche unique, comme illustré dans la Figure 4.1. La principale modification apportée réside dans l'opérateur de convolution ($\mathfrak{R}^C \rightarrow \mathfrak{R}^{d^{l+1}}$) destiné à agréger les informations de voisinage. Tout d'abord, en raison des bonnes performances observées dans le chapitre 2, nous avons considéré l'opérateur de convolution ECCConv [68] capable de considérer des attributs d'arêtes multi-dimensionnels lors du *message passing*. Le comportement de cet opérateur a été formalisé dans le chapitre 1 (eq 1.7) et s'appuie notamment sur la fonction différentiable F^{l+1} . Pour cette application, la fonction $F^{l+1} : \mathfrak{R}^3 \rightarrow \mathfrak{R}^{d^{l+1} \times C}$ (perceptron multi-couches) est composée de deux couches entièrement connectées suivies d'une fonction d'activation ReLU, comme détaillé dans la Figure 4.1-C. Dans le chapitre 2, F^{l+1} était limitée à la deuxième couche entièrement connectée $F_2^{l+1} : \mathfrak{R}^3 \rightarrow \mathfrak{R}^{d^{l+1} \times C}$ dont l'objectif est de mettre en correspondance l'espace des arêtes avec l'espace des noeuds (fonction dite de *mapping*) afin de pouvoir combiner par la suite les deux sources d'informations structurelles à travers le produit $F^{l+1}(L(u, v))X^l(u)$. Dans ces travaux, afin de manipuler des graphes complets, nous rajoutons la première couche $F_1^{l+1} : \mathfrak{R}^3 \rightarrow \mathfrak{R}^3$ dont le rôle est d'ajuster le poids de chaque attribut de chaque arête dans le calcul de la nouvelle représentation $X^{l+1}(v)$ de chaque noeud $v \in V$. Le but est de réduire l'impact, durant le *mapping* (F_2), des arêtes (ou plutôt des attributs d'arêtes) non pertinentes. Ainsi, la pertinence des arêtes n'est plus définie par un paramètre spatial (rayon d'élagage) mais est apprise par le GNN. $X^{l+1}(v)$ est ensuite calculé à partir de $X^l(v)$ et de la combinaison des informations de noeuds voisins et des arêtes connexes (produit $F^{l+1}(L(u, v))X^l(u)$, $u \in \mathcal{N}(v)$) par addition, opérateur invariant par permutation comme requis (voir fonction Ψ - eq 1.5).

Le reste de l'architecture GNN reste inchangé, c'est-à-dire que la sortie de l'opérateur de convolution, dont la dimension d^{l+1} est un hyperparamètre de la méthode, passe

à travers un perceptron à couche unique (SLP : $\mathfrak{R}^{d+1} \rightarrow \mathfrak{R}^C$) assignant un vecteur de probabilité d’appartenance à chaque classe $c \in C$ pour chaque noeud $v \in V$ du graphe. De même, tous les paramètres du réseau sont optimisés avec la base d’entraînement pour minimiser la fonction de coût du logarithme négatif de la vraisemblance (eq 2.2).

4.3 Expériences

Nous présentons dans cette section la base de données considérée (section 4.3.1), le protocole d’évaluation (section 4.3.2), les CNNs utilisés ainsi que l’obtention des graphes à partir de leur sortie (section 4.3.3) et les résultats obtenus (section 4.3.4).

4.3.1 Base de données

Notre méthode est évaluée sur la cohorte *AVCnn* [120, 140] de 69 enfants âgés de 7 ans. Ces enfants incluent des enfants sains (contrôles) et des enfants après NAIS présentant une lésion de l’hémisphère droit ou gauche. Chaque enfant de la cohorte a passé une IRM sur un scanner 3.0 Tesla (volumes 3D T1 de 256x256x176 voxels avec une résolution spatiale de 1x1x1mm³) sur le site de Neurospin, CEA Saclay, France. Les séquences d’imagerie comprenaient également une séquence 3D FLAIR à haute résolution. Pour construire les vérités terrain de la segmentation des noyaux gris centraux sur IRM T1, nous combinons une segmentation manuelle et une segmentation basée sur atlas obtenue à l’aide de l’atlas *Hammersmith* [168]. Plus précisément, les noyaux gris centraux segmentés avec l’atlas d’*Hammersmith* sont recalés, à l’aide de CAT12, sur un atlas du cerveau d’enfants de 7 ans construit avec l’outil Template-O-Matic [169] à partir des IRMs des 31 enfants sains de la base de données. A partir de la segmentation des noyaux gris centraux sur l’atlas des enfants de 7 ans, nous obtenons, par recalage, une segmentation dans l’espace natif de chaque IRM. La segmentation est ensuite corrigée manuellement coupe par coupe avec le logiciel ITKSnap [146] à l’aide de la séquence FLAIR. Chaque segmentation est vérifiée visuellement et validée par consensus avec un spécialiste. Neuf classes sont considérées dans la segmentation : le thalamus, le noyau caudé, le putamen, le pallidum (gauche et droit pour chaque structure) et le fond.

Un enfant a été exclu de l’étude car il avait trop bougé au moment de l’acquisition IRM : l’image acquise n’offrait pas une qualité suffisante pour la segmentation. Ainsi, la base de données finale est composée de 68 IRMs T1 segmentées d’enfants de 7 ans (31

enfants contrôles et 37 avec lésion cérébrale précoce comme indiqué dans la Table 4.1).

4.3.2 Protocole expérimental

Les expériences de ce chapitre sont menées sur le même environnement que celles du chapitre 2. Le modèle GNN est entraîné avec Adam sur 200 époques, avec un *dropout* de 0.5. Une stratégie de réduction du taux d'apprentissage sur plateau est utilisée avec un taux d'apprentissage initial $\Xi_0 = 0.01$ et un facteur de réduction $\sigma = 5e^{-4}$. La dimension d^{l+1} des attributs de noeuds en sortie de l'opérateur de convolution est fixée à 12.

Une fois encore, la qualité des segmentations est mesurée à l'aide du score de Dice noté DSC (eq 2.6) et de la distance d'Hausdorff notée HD (eq 2.7). Comme dans le chapitre 2, nous comparons nos performances de segmentation avec celles obtenues en sortie de CNN. Afin d'estimer l'efficacité de notre méthode pour segmenter correctement les noyaux gris centraux des cerveaux sains et des cerveaux lésés, nous distinguons les résultats obtenus sur les deux populations et évaluons la segmentation par hémisphère pour les enfants lésés (comparaison de l'hémisphère lésé et de l'hémisphère contra-lésé). En effet, la lésion liée à l'AVC néonatale étant unilatérale, on s'attend à ce qu'un hémisphère soit moins destructuré que l'autre et donc moins sujet à des erreurs de segmentation par les CNNs. Il est important de vérifier la segmentation dans chaque hémisphère séparément chez les enfants avec lésion cérébrale précoce pour pouvoir ensuite effectuer, à partir de la segmentation, des études inter-hémisphériques fiables [118, 170].

TABLE 4.1 – Taille de la base de données et configurations étudiées de la répartition en données d'entraînement (validation) et de test. Le nombre de sujets contrôles et lésés dans la base d'apprentissage (validation) pour les deux tirages aléatoires est indiqué.

Configuration	#données	entraînement (validation)				test
		A	B	C	D	
Total	68	35 (10)	17 (6)	8 (4)	4 (2)	23
Contrôles	31	18 (2)	6 (3)	3 (1)	3 (0)	11
		16 (4)	10 (1)	1 (2)	2 (2)	
Lésés	37	17 (8)	11 (3)	5 (3)	1 (2)	12
		19 (6)	7 (5)	7 (2)	2 (0)	

Pour être capable de segmenter tous types de cerveaux, nous entraînons et testons les modèles (CNNs et GNNs) avec des sujets des deux populations (contrôles et lésés) comme indiqué dans la Table 4.1. Pour évaluer la robustesse de notre méthode face à de petits jeux de données, nous considérons différentes tailles de la base d'apprentissage (c.a.d. de la

base servant à l’entraînement du CNN et du GNN). Cette stratégie se justifie par le désir de développer une méthode efficace, même avec très peu de données, ce qui constitue un défi dans le domaine médical [171]. Ainsi, nous travaillons avec les 4 configurations A, B, C et D présentées dans la Table 4.1. Pour chaque configuration, l’ensemble d’images utilisé à des fins de test est toujours le même (composé de 11 sujets contrôles et de 12 sujets ayant subi un AVC néonatal) et les résultats sont moyennés selon 2 tirages aléatoires de la base d’apprentissage, donnant lieu à 2 CNNs entraînés différemment. Dans nos expériences, un U-Net et deux architectures récentes incluant des mécanismes d’attention (section 1.2.3) sont considérés. Ces architectures, ci-après détaillées, ont été sélectionnées parmi celles proposant des implémentations libres d’accès pour les images médicales 3D. Cela conduit donc à 6 CNNs par configuration d’entraînement (3 CNNs différents avec 2 tirages aléatoires des images d’entraînement dans chaque cas). Nous comparons également nos performances avec deux méthodes basées sur atlas : l’algorithme entièrement automatique *vol2Brain* basé sur la technologie de fusion d’étiquettes multi-atlas [122] et l’atlas des enfants de 7 ans construit à partir des enfants contrôles de notre base de données.

4.3.3 CNNs et construction des graphes

Nous appliquons notre méthode de post-traitement basée sur GNN à différents CNNs. Ainsi, nous considérons un réseau U-Net 3D combiné à des CRFs ainsi que le récent UNETr [21] basé sur des mécanismes d’attention détaillés dans le chapitre 2 pour la base IBSR. Nous appliquons également notre méthode au réseau MSGSE-Net [22], basé sur des contextes d’image multi-échelles et des mécanismes d’attention, et dédié à la segmentation des noyaux gris centraux sur des cerveaux adultes sains. Nous avons choisi ce dernier modèle car il semble surpasser de nombreuses autres méthodes (FreeSurfer [121], BrainSegNet [123], Ψ -Net [124]) pour la segmentation des noyaux gris centraux [22].

Concernant le modèle U-Net avec CRFs, nous avons utilisé la même architecture et mis en oeuvre le même apprentissage que pour la base IBSR dans le chapitre 2 (section 2.4.3). Pour l’apprentissage du *Visual Transformer* UNETr, nous avons fait quelques modifications. Les hyperparamètres du réseau sont fixés comme dans [21] mais cette fois avec des patches de taille 32^3 . Par ailleurs, afin de réduire les temps de calcul et la capacité de mémoire nécessaire, les IRMs de taille $256 \times 256 \times 176$ voxels sont sous-échantillonnées en images de $128 \times 128 \times 128$ voxels. Après segmentation par notre méthode, les images sont sur-échantillonnées (interpolation par spline cubique) pour revenir à leur taille initiale. Enfin, le modèle MSGSE-Net est implémenté uniquement, pour l’instant, pour des

patches 2D. Il combine des contextes d'image multi-échelles et des mécanismes d'attention pour apprendre plus précisément des caractéristiques discriminantes dans l'image. Nous entraînons le modèle avec Adam sur 60 époques en considérant les hyperparamètres et la fonction de coût (fonction de perte de Dice pondérée par entropie) de [22] à partir des coupes 2D de taille 256^2 pixels des IRMs.

A partir de la sortie du CNN, un graphe G est construit comme détaillé dans le chapitre 2 et rappelé dans la section 4.2. Une connectivité de 26 est considérée pour extraire les composantes connexes de la carte de segmentation fournie par le CNN. Ainsi, chaque noeud a un attribut de dimension 9 correspondant aux 9 classes considérées ($X : V \rightarrow \mathfrak{R}^9$) et les arêtes portent des informations de dimension 3 correspondant aux positions relatives des barycentres des régions connectées selon les 3 directions ($L : E \rightarrow \mathfrak{R}^3$).

4.3.4 Résultats

La Table 4.2 rapporte les tailles des graphes résultants des différentes architectures de réseaux de neurones avec les 4 configurations de la base d'apprentissage. La taille des graphes impliqués est, ici encore, arbitraire. Le nombre de noeuds augmente à mesure que la taille de la base d'apprentissage diminue car le CNN, moins efficace, implique plus de régions candidates (c.a.d. de noeuds) par classe.

TABLE 4.2 – Nombre moyen de noeuds pour les différents CNNs et configurations d'entraînement (valeurs moyennes sur les 2 tirages aléatoires de la base d'apprentissage).

	A	B	C	D
U-Net+CRF	9.8	11.1	15.3	13.1
UNETr	9.1	10.1	9.5	12.4
MSGSE-Net	9.9	11.3	11.7	13.7

La Table 4.3 compare les performances de notre méthode avec celles obtenues en sortie de CNN. Les valeurs en gras indiquent la meilleure des deux approches (CNN seulement ou CNN combiné à notre méthode). Par exemple, notre proposition avec U-Net (configuration B) est meilleure que le U-Net seul (valeur en gras pour notre proposition). Les résultats obtenus avec les 3 réseaux neuronaux et les 4 configurations sont rapportés.

Quels que soient le CNN et la configuration d'entraînement, notre proposition améliore le Dice médian et la distance d'Hausdorff médiane. Les améliorations fournies par le GNN, illustrées sur la Figure 4.2, engendrent une légère amélioration du Dice médian (jusqu'à

TABLE 4.3 – Comparaison de la segmentation des noyaux gris centraux sur des enfants contrôles et lésés avec un CNN uniquement (U-Net+CRF, UNETr ou MSGSE-Net) et avec la combinaison proposée de CNN et GNN (P+CNN). Les résultats (moyenne, médiane (Méd.) et écart-types) avec les 4 configurations sont présentés.

	Configuration A				Configuration B			
	DSC		HD		DSC		HD	
Contrôles	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.
U-Net+CRF	0.96 (± 0.01)	0.96	8.14 (± 6.43)	5.72	0.95 (± 0.01)	0.95	17.72 (± 6.50)	17.77
P+U-Net+CRF	0.96 (± 0.01)	0.96	2.04 (± 0.19)	2.05	0.95 (± 0.01)	0.95	2.14 (± 0.23)	2.16
UNETr	0.92 (± 0.01)	0.92	3.18 (± 1.04)	2.86	0.89 (± 0.02)	0.90	3.95 (± 1.56)	3.35
P+UNETr	0.92 (± 0.01)	0.92	2.73 (± 0.33)	2.61	0.89 (± 0.02)	0.90	3.18 (± 0.43)	3.07
MSGSE-Net	0.95 (± 0.01)	0.95	4.25 (± 3.60)	2.64	0.91 (± 0.03)	0.92	7.20 (± 3.83)	6.62
P+MSGSE-Net	0.95 (± 0.01)	0.95	3.26 (± 2.60)	2.51	0.94 (± 0.01)	0.95	2.86 (± 0.30)	2.84
Lésés	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.
U-Net+CRF	0.89 (± 0.11)	0.94	11.11 (± 11.03)	7.30	0.87 (± 0.12)	0.93	20.52 (± 7.80)	21.87
P+U-Net+CRF	0.89 (± 0.12)	0.95	11.38 (± 21.48)	2.52	0.87 (± 0.12)	0.94	9.37 (± 19.40)	2.92
UNETr	0.85 (± 0.10)	0.89	5.21 (± 2.78)	3.86	0.82 (± 0.09)	0.86	6.84 (± 4.78)	4.66
P+UNETr	0.85 (± 0.10)	0.89	6.47 (± 10.53)	3.16	0.82 (± 0.08)	0.86	4.61 (± 2.26)	3.85
MSGSE-Net	0.86 (± 0.09)	0.87	8.40 (± 6.05)	5.85	0.79 (± 0.13)	0.87	16.64 (± 12.93)	13.28
P+MSGSE-Net	0.85 (± 0.11)	0.90	14.04 (± 23.47)	4.16	0.84 (± 0.12)	0.91	18.07 (± 32.75)	4.47
	Configuration C				Configuration D			
	DSC		HD		DSC		HD	
Contrôles	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.
U-Net+CRF	0.92 (± 0.01)	0.92	35.33 (± 8.30)	36.45	0.92 (± 0.01)	0.92	33.72 (± 8.00)	33.47
P+U-Net+CRF	0.94 (± 0.01)	0.94	2.47 (± 0.38)	2.51	0.94 (± 0.01)	0.94	3.29 (± 1.99)	2.73
UNETr	0.80 (± 0.03)	0.80	8.12 (± 3.15)	7.5	0.74 (± 0.07)	0.75	19.69 (± 7.35)	16.67
P+UNETr	0.80 (± 0.03)	0.80	5.04 (± 0.51)	5.22	0.75 (± 0.07)	0.76	6.08 (± 1.49)	5.84
MSGSE-Net	0.88 (± 0.03)	0.88	10.13 (± 3.09)	9.26	0.85 (± 0.05)	0.86	17.33 (± 6.23)	15.47
P+MSGSE-Net	0.89 (± 0.04)	0.90	5.87 (± 3.05)	4.67	0.89 (± 0.02)	0.90	6.08 (± 3.72)	4.87
Lésés	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.	Moyenne	Méd.
U-Net+CRF	0.85 (± 0.12)	0.91	35.80 (± 10.27)	34.13	0.84 (± 0.12)	0.90	37.52 (± 11.60)	40.64
P+U-Net+CRF	0.85 (± 0.13)	0.93	14.87 (± 22.23)	2.87	0.85 (± 0.12)	0.91	16.44 (± 23.51)	3.74
UNETr	0.72 (± 0.10)	0.73	14.71 (± 12.79)	10.15	0.67 (± 0.07)	0.68	27.20 (± 13.77)	25.23
P+UNETr	0.72 (± 0.10)	0.73	9.42 (± 13.58)	5.43	0.68 (± 0.07)	0.69	12.73 (± 13.99)	7.15
MSGSE-Net	0.75 (± 0.16)	0.80	22.42 (± 25.64)	14.42	0.74 (± 0.11)	0.72	18.27 (± 5.23)	18.25
P+MSGSE-Net	0.76 (± 0.19)	0.84	31.77 (± 44.16)	8.45	0.76 (± 0.15)	0.80	24.87 (± 41.06)	7.76

8% - MSGSE-Net avec la configuration D sur les enfants lésés) et sont particulièrement significatives en termes de distance d’Hausdorff. Cela souligne la capacité de notre méthode à éliminer les artefacts (générés par le CNN) situés loin des noyaux gris centraux. Notre approche semble également corriger les erreurs de segmentation relativement proches des régions d’intérêt. En effet, même avec un *Visual Transformer* comme UNETr, fournissant moins d’artefacts (c.a.d. moins de noeuds - Table 4.2) grâce à ses mécanismes d’attention, notre proposition réduit de manière significative le HD moyen de 5.39 (moyenne des enfants sains et lésés) à 3.89 (réduction de 28%) pour la configuration d’entraînement B, de 11.41 à 7.23 (réduction de 37%) pour la configuration d’entraînement C et de 23.44 à 9.40 (réduction de 60%) pour la configuration d’entraînement D. Cependant, dans quelques cas

détaillés plus loin, notre méthode ne parvient pas à segmenter correctement les noyaux gris centraux. Par exemple, sur la Figure 4.3-droite, nous perdons avec notre méthode le putamen droit et une partie du noyau caudé droit. De même, sur la Figure 4.3-gauche, notre méthode ne parvient pas à récupérer les structures de l'hémisphère lésé (gauche), excepté une partie du thalamus.

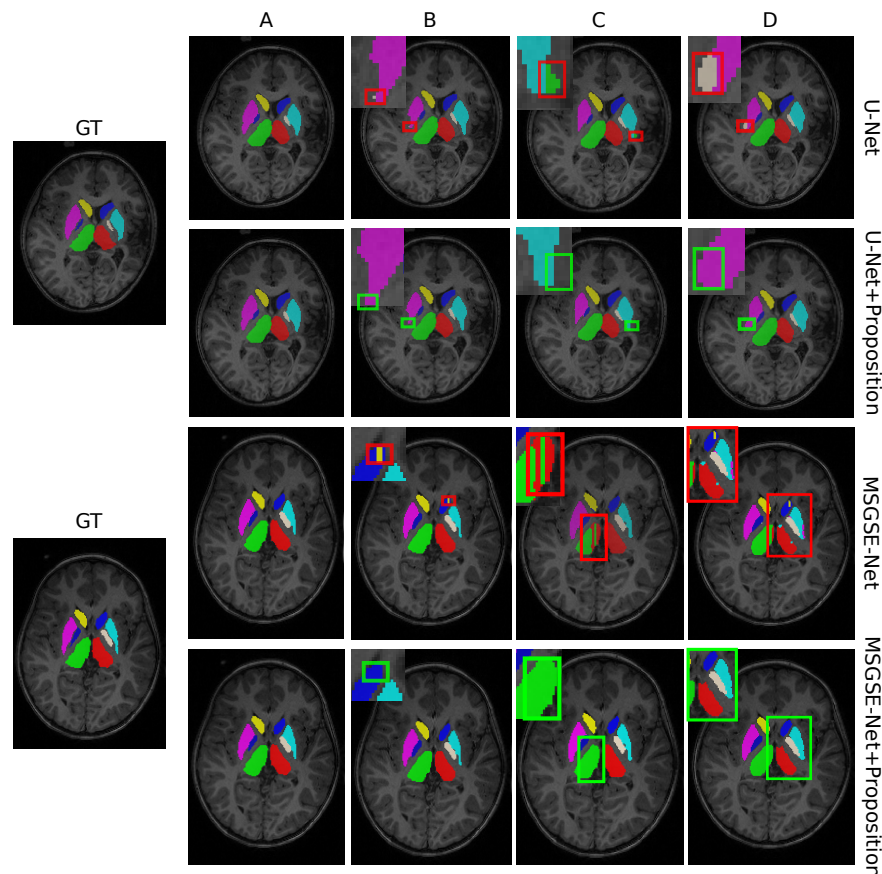


FIGURE 4.2 – Exemples de segmentations des noyaux gris centraux sur 2 enfants lésés considérant différents réseaux de neurones profonds (U-Net+CRF et MSGSE-Net) et l'ensemble des configurations considérées (A, B, C et D). Les résultats avec et sans notre méthode sont illustrés. GT : vérité terrain.

On peut observer dans la Table 4.3 que les performances des CNNs se dégradent avec la réduction du nombre de données d'entraînement en termes de Dice et de distance d'Hausdorff (observation mise en évidence par la Figure 4.2). Néanmoins, l'ajout de notre post-traitement basé sur GNN semble compenser la réduction de la base d'apprentissage en maintenant, pour toutes les configurations, une distance d'Hausdorff médiane relativement faible et constante, comme le montre la Figure 4.4.

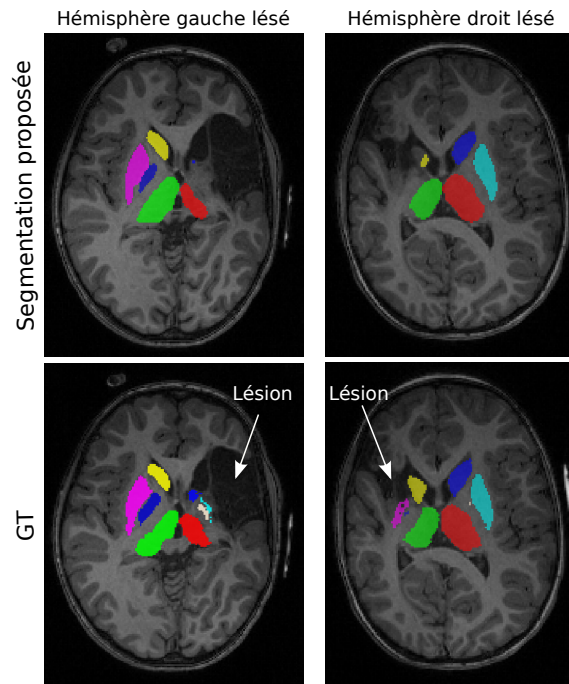


FIGURE 4.3 – Exemples de segmentations sur des sujets présentant une lésion importante à proximité des noyaux gris centraux. Les résultats sont obtenus dans la configuration B avec U-Net+CRF. GT : vérité terrain.

Au fur et à mesure que le nombre de données d’entraînement diminue, l’amélioration apportée par notre méthode augmente (voir Figure 4.2). Cela illustre la capacité de notre méthode à compenser de manière significative le manque de représentativité lorsque la base d’apprentissage est petite : les relations apprises au cours de l’apprentissage corrigent les erreurs structurelles commises par les CNNs.

A partir de la Table 4.3, on constate également que les performances obtenues sur les enfants avec lésion sont inférieures et plus hétérogènes que celles obtenues sur les enfants contrôles, y compris avec notre méthode de post-traitement (ex. MSGSE-Net en configuration A : DSC de $0.85 (\pm 0.11)$ contre $0.95 (\pm 0.01)$ pour les sujets sains, HD de $14.04 (\pm 23.47)$ contre $3.26 (\pm 2.60)$). Cela peut s’expliquer par des variations très importantes du cerveau, parfois observées dû à la lésion qui, en modifiant l’organisation structurelle du cerveau, perturbe le CNN et les GNNs basés sur les relations spatiales entre structures (Figure 4.3). Il en résulte un écart-type beaucoup plus élevé chez les enfants avec lésion cérébrale précoce pour les 2 métriques (DSC et HD). Pour quelques sujets lésés, en raison d’une lésion importante, la performance est si faible qu’elle dégrade significativement la performance moyenne sur l’ensemble des données de test (ex. notre méthode

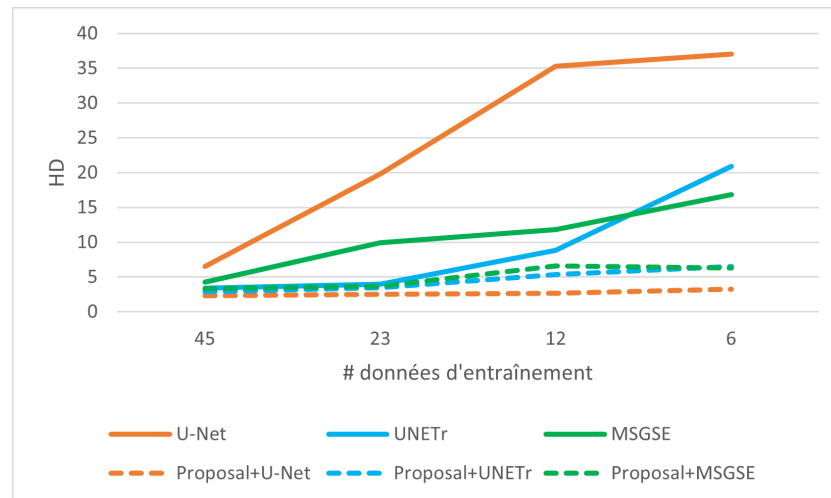


FIGURE 4.4 – Évolution de la distance d’Hausdorff médiane en fonction du nombre de données d’entraînement. L’évolution est illustrée pour les 3 CNNs (U-Net, UNETr et MSGSE-Net) avec et sans notre méthode de post-traitement.

appliquée sur MSGSE-Net dans toutes les configurations). C’est pourquoi nous indiquons les valeurs médianes afin d’être moins sensibles à ces valeurs aberrantes. Même si, dans certains cas particuliers, un CNN seul obtient, en moyenne, de meilleurs résultats que notre approche du fait des valeurs aberrantes, notre méthode permet toujours d’obtenir les meilleures performances médianes (DSC et HD). En effet, elle améliore jusqu’à 93% la distance d’Hausdorff pour les enfants contrôles (réduction de 36.45 à 2.51 dans la configuration C avec U-Net+CRF) et jusqu’à 91% pour les enfants avec lésion cérébrale précoce (réduction de 40.64 à 3.74 dans la configuration D avec U-Net+CRF). Cela démontre la capacité de notre post-traitement à segmenter les noyaux gris centraux à la fois dans des cerveaux sains et dans des cerveaux avec lésion corticale (Figure 4.2), à condition que cette dernière ne perturbe pas trop l’organisation spatiale des noyaux gris centraux.

Afin de mieux comprendre les moins bonnes performances sur les enfants ayant eu un AVC néonatal, nous détaillons dans la Table 4.4 les résultats par hémisphère obtenus sur les enfants après NAIS, pour les 4 configurations, en considérant le réseau U-Net (la tendance est similaire pour les autres CNNs mais les résultats ne sont pas rapportés par souci de clarté). On note clairement que c’est la segmentation de l’hémisphère lésé qui dégrade les résultats, y compris dans la configuration A avec plus de données d’entraînement (DSC de 0.83 pour l’hémisphère lésé contre 0.96 pour l’hémisphère contra-lésé, HD de 16.08 contre 2.19). Ce constat est bien illustré dans la Figure 4.3. En effet, on observe que pour l’enfant présentant une lésion dans l’hémisphère gauche, les erreurs de segmentation

TABLE 4.4 – Performances moyennes (médianes) par hémisphère (Hem.) de la segmentation des noyaux gris centraux des enfants avec lésion cérébrale précoce (résultats obtenus avec le réseau U-Net en considérant les 4 configurations d’entraînement). Une comparaison sans notre post-traitement basé sur GNN est fournie.

	DSC				HD			
	A	B	C	D	A	B	C	D
Hem. contra-lésé sans GNN	0.95 (0.96)	0.95 (0.95)	0.94 (0.94)	0.93 (0.93)	3.73 (2.31)	6.53 (4.66)	13.38 (13.42)	16.34 (21.90)
Hem. contra-lésé	0.96 (0.96)	0.95 (0.95)	0.94 (0.94)	0.94 (0.94)	2.19 (2.34)	2.37 (2.43)	2.77 (2.77)	2.92 (2.9)
Hem. lésé sans GNN	0.82 (0.88)	0.81 (0.91)	0.79 (0.86)	0.79 (0.89)	11.01 (7.39)	8.97 (9.04)	24.16 (15.17)	29.98 (23.42)
Hem. lésé	0.83 (0.95)	0.83 (0.92)	0.80 (0.89)	0.81 (0.89)	16.08 (2.67)	12.58 (2.90)	20.86 (2.98)	21.97 (3.49)

se situent dans l’hémisphère gauche (perte du noyau caudé, du putamen et du pallidum) et inversement pour l’enfant dont l’hémisphère droit est lésé (perte du putamen droit et d’une partie du noyau caudé droit). Cette dégradation des résultats dans l’hémisphère lésé de certains sujets peut conduire à une moins bonne performance moyenne de notre proposition par rapport au CNN (HD configurations A et B - hémisphère lésé). Les hémisphères contra-lésés sont, eux, correctement segmentés : la performance moyenne de l’hémisphère contra-lésé est toujours améliorée avec la proposition. Néanmoins, on constate dans la Table 4.4 que notre méthode rééquilibre, en considérant les valeurs médianes, les performances inter-hémisphériques de segmentation par rapport au CNN seul. Par exemple, si l’on considère la configuration B, notre méthode améliore légèrement le HD médian de l’hémisphère contra-lésé (4.66 à 2.43) mais plus significativement le HD médian de l’hémisphère lésé (9.04 à 2.90). Quelle que soit la configuration, notre méthode réduit la différence de performance médiane entre les hémisphères à moins de 1 pour le HD et à moins de 5% pour le Dice. Nous remarquons encore que l’apport de notre proposition dans chaque hémisphère augmente avec la réduction du nombre de données d’entraînement.

L’un des principaux défis de la recherche médicale en neuroimagerie est de développer des méthodes d’analyse objectives, robustes et efficaces avec très peu de données, en raison de la faible quantité de données disponibles, particulièrement dans l’étude de pathologies rares. Dans la Table 4.5, nous comparons les résultats obtenus avec notre proposition appliquée à la sortie d’un U-Net (avec CRF), en considérant seulement 6 images dans la base d’apprentissage (4 pour l’entraînement et 2 pour la validation - configuration

TABLE 4.5 – Comparaison de notre méthode appliqué à la sortie d’un U-Net (avec CRF) avec une approche basée sur atlas ainsi que d’autres méthodes (logiciels ou CNNs) pour la segmentation des noyaux gris centraux du cerveau des enfants sains et des enfants avec lésion cérébrale précoce (le nombre de données pour les CNNs correspond à la configuration avec les meilleurs résultats). M : million, N.A. : non disponible.

	# données	# param. entraînaibles	DSC (médiane)	HD (médiane)
Référence (atlas)	31	-	0.99	3.09
<i>vol2Brain</i> [122]	N.A.	N.A.	0.84	5.02
U-Net (CRF) [98]	45 (35+10)	15M	0.95	6.51
UNETr [21]	45 (35+10)	94M	0.91	3.36
MSGSE-Net [22]	45 (35+10)	33M	0.91	4.24
U-Net (CRF) + Proposition	6 (4+2)	15M	0.93	3.23

D), avec la meilleure performance obtenue avec les 3 CNNs considérés, correspondant pour tous à la configuration A (35 images pour l’entraînement et 10 pour la validation). Logiquement, UNETr et MSGSE-Net sont plus performants que U-Net seul en termes de HD même si le Dice est légèrement meilleur avec U-Net. Notre méthode de post-traitement, qui ajoute un nombre négligeable de paramètres (seulement 345 paramètres entraînaibles pour le GNN), améliore le HD médian du U-Net, devenant même inférieur à celui de UNETr et MSGSE-Net seuls, avec un ensemble de données d’entraînement réduit et un nombre de paramètres entraînaibles nettement plus petit (15M contre 94M pour UNETr et 33M pour MSGSE-Net). Dans la Table 4.5, on observe également que notre méthode fournit des performances équivalentes à celles basées sur l’atlas des enfants de 7 ans, en termes de HD, avec seulement 6 sujets d’entraînement. Il s’agit d’une bonne performance puisque la construction de l’atlas nécessite davantage de données. De plus, l’atlas de 7 ans a été construit à partir de sujets contrôles dont certains sont ici utilisés pour l’évaluation. Par conséquent, il est raisonnable de supposer que les valeurs présentées (DSC et HD) sont trop optimistes par rapport aux CNNs qui n’ont pas été entraînés avec ces sujets contrôles. Ainsi, même si la distance d’Hausdorff avec notre méthode est légèrement moins bonne qu’avec l’atlas, elle semble convaincante. Pour le confirmer, nous pouvons également constater dans la Table 4.5 que notre méthode surpasse, en considérant les résultats médians, le récent algorithme en ligne multi-atlas *vol2Brain* [122]. Cela illustre la pertinence de tenir compte des informations structurelles de haut-niveau, non seulement en termes d’efficacité de la segmentation, mais aussi en termes de quantité de données et de paramètres entraînaibles, ce qui constitue un défi [81].

4.4 Discussion

Dans ces travaux, nous proposons d’exploiter une adaptation de la méthode de post-traitement basée sur GNN pour la segmentation sémantique d’images, introduite dans le chapitre 2, afin d’améliorer la segmentation des noyaux gris centraux fournie par les CNNs (section 1.2.2). Notre approche a pour objectif d’automatiser la segmentation fidèle des noyaux gris centraux, dans des cerveaux considérés comme sains tout comme dans des cerveaux présentant une lésion cérébrale précoce, en exploitant les prédictions de CNNs ainsi que les informations structurelles entre régions détectées par ces derniers. Nous étudions la pertinence de la méthode sur des enfants ayant eu un infarctus cérébral artériel néonatal (NAIS), se traduisant par une lésion corticale comme expliqué en introduction de la partie II. Il s’agit d’un exemple de lésion cérébrale précoce relativement complexe (du fait de la localisation de la lésion) et représentatif puisque nous avons vu que le NAIS était le plus courant des accidents vasculaires cérébraux périnataux.

Ainsi, nous avons entraîné les réseaux (CNNs et GNNs) à partir d’un jeu de données composé d’enfants sains et d’enfants ayant eu un NAIS. Nous tentons de corriger les erreurs de segmentation du CNN (artefacts et incohérences spatiales souvent liées à une configuration différente due à la lésion) en s’appuyant sur les relations structurelles connues entre les régions. Nous avons montré que notre post-traitement améliore la performance de segmentation (DSC et HD) sur les sujets sains mais aussi, en considérant la valeur médiane, sur les sujets avec lésion cérébrale précoce (Table 4.3). Plus précisément, notre méthode élimine, dans la plupart des cas, les erreurs de segmentation dans l’hémisphère lésé produites par le CNN. Cela permet d’obtenir une qualité de segmentation équivalente dans les deux hémisphères, en considérant les performances médianes. Cette amélioration est observée avec différents CNNs adaptés à la segmentation des noyaux gris centraux sur IRM, en particulier les réseaux UNETr et MSGSE-Net récemment proposés.

Nous avons exploité l’ensemble des informations spatiales entre les régions, grâce à l’utilisation de graphes complets, afin d’enrichir la prédiction de chaque noeud. L’ajout de la fonction F_1^{l+1} dans la fonction différentiable F^{l+1} de ECCConv (eq 1.7) a permis de distinguer l’impact de chaque noeud voisin et de chaque attribut d’arête lors de l’agrégation des informations de voisinage. L’application de l’architecture du chapitre 2, considérant uniquement la fonction de *mapping* F_2^{l+1} , a donné de moins bons résultats. Il serait toutefois intéressant d’étendre cette comparaison à l’utilisation de la technique d’élagage présentée dans le chapitre 2 pour évaluer l’apport de la fonction F_1 et des graphes complets.

Une fois encore, nous avons montré la robustesse de notre approche face au manque de données d’entraînement, un problème récurrent dans le domaine médical surtout dans les études pédiatriques relatives à des pathologies rares. Pour mettre cela en évidence, nous avons pris l’exemple d’une population d’enfants présentant une pathologie non fréquente (NAIS [111]) conduisant à une petite base de données, bien que du même ordre de grandeur que d’autres travaux récents portant sur l’étude de l’AVC néonatal (31 enfants sains et 37 enfants lésés contre 20 enfants sains et 44 enfants lésés dans [118]). Nous avons étudié l’influence de la taille de la base d’apprentissage sur les performances (configurations A à D). Ainsi, nous parvenons à segmenter avec précision (HD médian ≤ 8) les noyaux gris centraux des enfants sains et lésés dans toutes les configurations (c.a.d. avec une base d’apprentissage allant de 45 à seulement 6 images - Table 4.1 et Figure 4.4). L’application de notre post-traitement basé sur GNN améliore considérablement les résultats de segmentation obtenus avec tous les CNNs considérés et entraînés sur peu de données (Table 4.3, Figure 4.4 - Configuration D), en corrigeant les incohérences structurelles induites par le CNN (Figure 4.2). De plus, dans la Table 4.5, nous montrons que l’utilisation de notre post-traitement à la sortie d’un U-Net entraîné avec seulement 6 images, conduit à de meilleurs résultats sur l’ensemble de la base de données de test (enfants sains et lésés) que la solution multi-atlas *vol2Brain* et que des réseaux d’apprentissage profonds plus complexes tels que UNETr et MSGSE-Net entraînés sur 45 images.

Cependant, notre méthode n’est pas encore assez robuste en présence d’une grande lésion (Figure 4.3), spécifiquement située près des structures cérébrales sous-corticales, car celle-ci entraîne un changement significatif de la structure spatiale dans l’hémisphère lésé, qui devient alors trop différente de la structure apprise par le GNN. C’est pourquoi, nos performances de segmentation moyennes sont inférieures dans l’hémisphère lésé (Table 4.4). Pour améliorer l’efficacité de notre méthode dans le cas de lésions larges et proches des voies motrices, une alternative pourrait être de considérer, uniquement dans ce cas, des techniques tierces axées sur la segmentation des lésions (masquage des lésions) [172] qui permettraient de guider la segmentation des noyaux gris centraux, au prix d’un processus de segmentation plus complexe. Sinon, il serait envisageable, ici aussi, d’exploiter l’incertitude du CNN [105] afin de subdiviser plus finement les régions produites par ce dernier. Il est à noter qu’une telle situation pourrait également être gérée par une correction manuelle (segmentation semi-automatique) puisque, comme nous l’avons observé dans les expériences, cela ne se produit que dans quelques cas (sur notre jeu de données).

Bien que la taille du jeu de données considéré soit similaire à celle d’autres travaux

relatifs à l’AVC néonatal, ces résultats doivent être confirmés sur un plus grand nombre de données. En effet, les résultats ont été obtenus sur une cohorte homogène en termes de caractéristiques des patients (nouveau-nés à terme aujourd’hui âgés de 7 ans), de caractéristiques des lésions (type NAIS) et de conditions d’imagerie (acquisition monocentrique des données à l’aide d’un scanner IRM 3T). Par conséquent, ils ne sont pas brouillés par les conséquences développementales de la prématurité ou la variabilité de l’acquisition expérimentale multicentrique. On peut envisager d’étendre l’approche à des populations adultes (supposées plus simples car présentant un contraste tissulaire plus marqué sur l’IRM). Toutefois, il serait intéressant de la tester à des stades plus précoces pour garantir une étude longitudinale du développement des noyaux gris centraux, en particulier en présence d’une lésion cérébrale précoce. En outre, la proposition devra être confirmée sur d’autres types de lésions cérébrales précoces (lésions de la substance blanche ou autres types d’AVC néonataux, lésions unilatérales ou bilatérales).

4.5 Conclusion

Pour traiter la tâche complexe mais cruciale de segmentation des noyaux gris centraux d’enfants sur IRM, en particulier en présence d’une lésion cérébrale précoce, nous avons proposé dans ce chapitre d’appliquer une adaptation de la méthode de segmentation sémantique présentée dans le chapitre 2. L’application de cette méthode de post-traitement, basée sur l’exploitation des informations structurelles sous forme de graphe par les GNNs, repose sur l’organisation spatiale connue et relativement constante des noyaux gris centraux. L’adaptation proposée s’appuie sur des graphes complets et cherche à ajuster les poids des attributs des arêtes pour la classification de noeuds à l’aide d’une fonction intégrée dans l’opérateur de convolution sur graphe ECCConv. Notre proposition s’est avérée robuste au manque de grands ensembles de données d’entraînement représentatifs (particulièrement présent dans les études pédiatriques reliées à des pathologies peu fréquentes), surpassant des méthodes multi-atlas et des architectures neuronales complexes (UNETr, MSGSE-Net) entraînées avec davantage de données. La méthode a été évaluée sur une cohorte d’enfants sains et d’enfants après NAIS (type le plus courant d’AVC néonatal). Nous espérons qu’elle pourra être généralisée à d’autres types de lésions cérébrales unilatérales précoces. Nos recherches futures se concentreront sur l’amélioration de la segmentation dans l’hémisphère lésé des sujets présentant une lésion importante proche des voies motrices (basée sur l’incertitude du CNN et sur la segmentation de la lésion elle-même). Ces

travaux sont dans la continuité de ceux présentés lors de la conférence *IPTA 2022* [108].

Avec cette méthode de segmentation des noyaux gris centraux chez l'enfant avec lésion cérébrale précoce, nous progressons dans la chaîne de traitement (Figure II.1) pour comprendre l'impact de la lésion cérébrale précoce sur le développement cérébral et sur la motricité des enfants. Une fois ces structures cérébrales d'intérêt segmentées, nous pouvons les exploiter afin de caractériser le cerveau des enfants après AVC néonatal. Dans le chapitre suivant, nous nous interrogeons sur une possible corrélation entre l'organisation structurelle des noyaux gris centraux et l'apparition de troubles moteurs permanents chez les enfants, pouvant être diagnostiquée comme paralysie cérébrale unilatérale.

À retenir

- En raison de leur rôle clé dans les fonctions cérébrales, les noyaux gris centraux sont souvent étudiés chez les enfants. Cependant, leur analyse, en particulier après une lésion cérébrale précoce, est difficile car les CNNs actuels ne parviennent pas à les segmenter avec précision sur IRM.
- De par l'organisation spatiale connue et stable des noyaux gris centraux, nous appliquons notre post-traitement exploitant les informations structurelles par un GNN pour améliorer la segmentation sémantique de ces derniers.
- Les informations structurelles portées par les arêtes (relations spatiales entre les régions détectées par le CNN) et les noeuds du graphe (probabilités des régions issues du CNN) corrigent, dans une certaine mesure, les incohérences spatiales produites par le CNN.
- La méthode s'appuie sur des graphes complets et le poids des attributs d'arêtes est appris par le GNN, grâce à une fonction appropriée intégrée dans l'opérateur de convolution.
- L'approche reste robuste au manque de données et peut surpasser, notamment pour les enfants avec lésion cérébrale précoce, des méthodes récentes d'apprentissage profond pour la segmentation des noyaux gris centraux.

ÉTUDE DE LA MOTRICITÉ DE L'ENFANT APRÈS AVC NÉONATAL : GNNs ET ORGANISATION MACROSTRUCTURELLE DES NOYAUX GRIS CENTRAUX

5.1 Introduction

A l'entame de la partie II de ce manuscrit, nous avons vu que l'AVC néonatal s'avère être un bon modèle d'étude de la lésion cérébrale précoce car il correspond à une lésion bien caractérisée (lésion corticale ou sous-corticale), non progressive et intervenant toujours à la même période chez des enfants nés à terme [109]. Nous avons également constaté que l'infarctus cérébral artériel néonatal (NAIS) est le type d'AVC néonatal le plus commun avec une prévalence estimée à environ 1/4000 naissances à terme [111]. NAIS apparaît comme l'accident vasculaire cérébral périnatal le plus étudié pour comprendre l'impact de la lésion cérébrale unilatérale précoce sur la plasticité du cerveau en développement [109], à l'origine de conséquences sur le long terme, notamment sur le plan moteur. En effet, concernant la déficience motrice à long terme, des données récentes indiquent que le taux de paralysie cérébrale (PC) unilatérale [112] chez les enfants nés à terme avec NAIS est proche de 30%, faisant de la PC la déficience motrice la plus fréquente chez les enfants [173]. Nous avons aussi précédemment expliqué que les nouveaux-nés présentant les lésions cérébrales les plus étendues ont le plus grand risque de souffrir de handicaps moteurs à long terme [113] bien que certains enfants chez lesquels la lésion cérébrale est relativement petite mais spécifiquement localisée le long des voies motrices, atteignant le tractus cortico-spinal, peuvent également développer une PC [114, 174]. Ainsi, il est difficile de prédire, à partir d'une IRM, la trajectoire du développement moteur des enfants après AVC néonatal et l'apparition possible de la PC. Pourtant, ce diagnostic précoce est

souhaitable pour mettre en oeuvre des interventions thérapeutiques adaptées (Figure II.1).

Au-delà de diagnostiquer la présence d'une paralysie cérébrale, déterminer la sévérité des atteintes motrices chez l'enfant à partir d'une IRM est encore plus complexe. Si tous les enfants atteints de PC unilatérale après NAIS marchent, les difficultés motrices sont principalement liées à une déficience de la main affectée (dite contra-lésionnelle), mais aussi, étonnamment, de la main ipsilésionnelle (théoriquement "non-affectée"). Il est bien établi que la fonction de cette dernière est souvent anormale chez ces enfants [175], sans qu'il y ait d'explication convaincante à cela. Ainsi, la prédiction avec précision de la fonction motrice des mains (main affectée et "non-affectée") à partir de l'IRM morphométrique est difficile, conduisant à des résultats contradictoires [117, 118, 119, 174].

Par ailleurs, nous avons vu dans l'introduction de la partie II et dans le chapitre 4, l'importance d'étudier les noyaux gris centraux (thalamus, noyau caudé, putamen et pallidum) de par leur rôle clé dans les fonctions sensori-motrices [115]. Ces derniers apparaissent alors comme des candidats idéaux pour expliquer la fonction motrice après lésion cérébrale précoce. Jusqu'à récemment, ces structures ont souvent été explorées dans le cadre d'études en neuroimagerie afin d'évaluer leur relation avec la fonction motrice à la suite d'une lésion cérébrale précoce, avec des résultats controversés. Certains travaux affirment que les enfants après NAIS ont un thalamus et un putamen ipsilésionnels plus petits, ceci étant corrélé à l'altération de la fonction de la main [119]. D'autres ne trouvent pas cette association pour le thalamus [117]. De façon intéressante, des recherches ont également trouvé une relation entre le volume du putamen ipsilésionnel et la fonction de la main affectée sans analyser le volume du thalamus, pourtant très lié à la fonction motrice du cerveau [118]. Des résultats encore plus controversés existent concernant le lien entre le volume des noyaux gris centraux et la motricité de la main dite "non-affectée" [117, 119]. À ce jour, il n'existe pas de relation claire entre les résultats moteurs cliniques et les volumes des noyaux gris centraux, que ce soit pour la main affectée ou la main "non-affectée" (c.a.d. pour l'hémisphère lésé ou contra-lésé).

De manière générale, ces études s'intéressent uniquement au lien entre le volume des structures cérébrales et les fonctions motrices. Néanmoins, compte tenu de l'anatomie complexe des noyaux gris centraux ainsi que des relations complexes entre ces structures cérébrales profondes et le cortex (réseau cortico-basallien-thalamo-cortical), l'analyse de l'association entre la fonction motrice et les seuls volumes des structures semble être une approche trop simpliste. C'est probablement la raison pour laquelle les différents travaux n'ont pas trouvé les mêmes résultats relatifs à la fonction motrice à long terme

après NAIS. Nous émettons donc l’hypothèse de l’implication d’autres paramètres que le volume, dans la relation entre ces structures anatomiques et la fonction motrice. Ainsi, une étude sur des sujets contrôles a démontré, basé sur une analyse de forme, que les changements morphométriques locaux sont plus fortement liés au fonctionnement cognitif que les changements de volume globaux [176]. Dans ce sens, nous proposons d’étudier la motricité des enfants après AVC néonatal en exploitant l’organisation macrostructurelle générale des noyaux gris centraux incluant les volumes, les allongements et les distances entre structures. Dans le chapitre 1, nous avons vu que ce type d’informations structurelles peut être représenté sous forme de graphe où chaque noeud correspond à une structure (c.a.d. qu’il porte des informations spécifiques aux structures : volumes, allongements, etc.) et les arêtes indiquent les informations spatiales entre structures (ex. distances).

Dans un premier temps, nous proposons d’exploiter ces graphes représentant l’organisation macrostructurelle des noyaux gris centraux par un GNN (section 1.4.2) entraîné à réaliser de la classification de graphes pour détecter la présence de paralysie cérébrale chez les enfants ayant eu un AVC néonatal. L’utilisation de graphes permet de représenter plus explicitement les relations spatiales entre structures qu’avec des classifieurs traditionnels d’apprentissage machine qui considéreraient les différents attributs comme des covariables. De plus, la pertinence des graphes pour diagnostiquer des pathologies à partir d’IRMs a été récemment démontrée (classification de troubles neuropsychiatriques [177], diagnostic des troubles du spectre autistique [178]). Ainsi, nous cherchons à voir si à partir de l’organisation structurelle des noyaux gris centraux, les GNNs arrivent à discerner d’une part les enfants ayant eu un AVC néonatal et d’autre part, parmi ces derniers, les enfants ayant développé une paralysie cérébrale.

Dans un second temps, nous exploitons à nouveau ces mêmes graphes par un GNN pour traiter un problème plus difficile que la simple détection de PC, à savoir l’estimation des fonctions motrices de chacune des deux mains (évaluées à l’aide d’un test de motricité présenté plus tard). La prédiction de ce score de motricité peut être vue comme un problème de régression sur graphes. Dans notre contexte médical, ce type d’approche, basé sur l’usage des GNNs pour des problèmes de régression, a été récemment considéré pour évaluer l’intelligence en prédisant le score de QI à partir de graphes représentant la connectivité cérébrale [179]. En tenant compte des différents éléments décrits ci-dessus (rôle clé des noyaux gris centraux dans le système moteur, impact d’autres paramètres morphologiques et structurels que les volumes sur la fonction clinique), nous émettons l’hypothèse que des enfants, à 7 ans, après un AVC néonatal, vont présenter des change-

ments au niveau de l'organisation structurale des noyaux gris centraux (modélisée par un graphe), liés à la fonction motrice des deux mains (main affectée et main "non-affectée").

Dans la suite de ce chapitre, nous détaillerons la méthode de construction des graphes à partir des images IRM commune aux deux applications (section 5.2) avant de présenter la base de données utilisée (images et méthode d'évaluation des fonctions motrices) dans la section 5.3. Nous distinguerons ensuite les expériences associées à la détection de la paralysie cérébrale chez les enfants ayant eu un AVC néonatal (section 5.4) et celles visant à évaluer les fonctions motrices manuelles (section 5.5). Nous achèverons le chapitre par une discussion sur les résultats obtenus ainsi qu'une conclusion.

5.2 Méthode

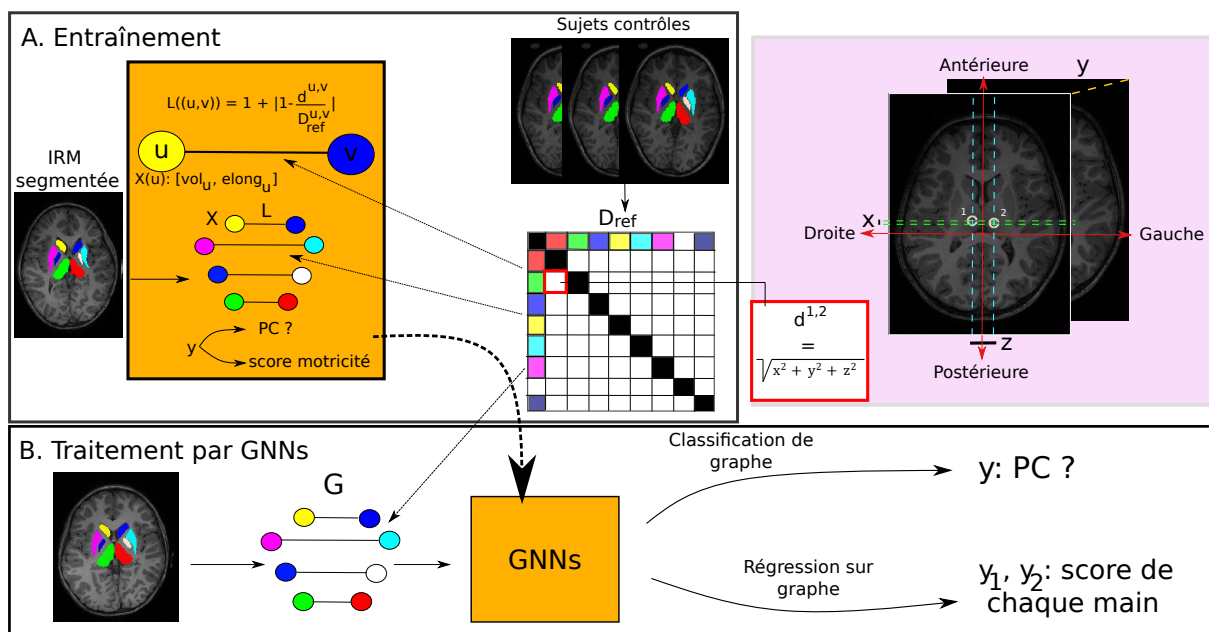


FIGURE 5.1 – Vue générale de la méthode proposée relative à l'étude de la motricité chez les enfants après AVC néonatal. La partie supérieure illustre la stratégie de construction des graphes. Les graphes issus des données d'apprentissage sont destinés à entraîner deux types de GNN. Le premier fait de la classification de graphes pour détecter les enfants avec paralysie cérébrale. Le deuxième cherche à résoudre un problème de régression pour évaluer un score de motricité de chacune des deux mains.

La Figure 5.1 fournit une vue d'ensemble de la méthode proposée. A partir de la segmentation des noyaux gris centraux sur une IRM d'enfant ayant eu un AVC néonatal

(segmentation obtenue, par exemple, par la technique présentée dans le chapitre 4), nous construisons un graphe $G = (V, E, X, L)$. Chaque noeud $u \in V$ correspond à une structure des noyaux gris centraux. Dans une volonté d'étudier les relations structurelles inter-hémisphériques entre des structures supposées symétriques, nous connectons uniquement les structures deux à deux (c.a.d. le thalamus gauche avec le thalamus droit, etc.). Les structures inter et intra-hémisphériques de natures différentes ne sont pas reliées par une arête afin de ne pas fournir des informations qui pourraient perturber le GNN (ex. le volume du thalamus droit n'est pas comparable à celui du noyau caudé droit). La structure des graphes construits est illustrée sur la Figure 5.1. La fonction $X : V \rightarrow \mathfrak{R}^2$ du graphe G attribue à chaque noeud u des attributs relatifs au volume (normalisé par le volume cérébral total) et à l'allongement (normalisé par le périmètre crânien) de la structure associée à u ($X(u) = [\text{vol}_u, \text{elong}_u]$ - Figure 5.1). Plus précisément, l'allongement est mesuré comme la longueur de l'axe majeur de l'ellipsoïde s'ajustant à la forme de la structure (même second moment centré et normalisé que la structure). La fonction $L : E \rightarrow \mathfrak{R}$ du graphe G attribue aux arêtes une valeur liée à la distance entre les deux structures connectées. Les arêtes portent plus exactement une information relative au rapport de cette distance avec la distance moyenne chez les individus contrôles :

$$L((u, v)) = 1 + \left| 1 - \frac{d^{u,v}}{D_{\text{ref}}^{u,v}} \right| \quad (5.1)$$

tel que $d^{u,v}$ indique la distance euclidienne entre les barycentres des structures u et v dans l'espace tri-dimensionnel et $D_{\text{ref}}^{u,v}$ indique la distance euclidienne moyenne entre les barycentres des structures u et v chez des sujets contrôles, supposés sains. Ainsi, la matrice D_{ref} est préalablement construite, à partir des IRMs d'enfants contrôles, comme la moyenne sur tous ces enfants de la distance euclidienne entre les barycentres de chaque structure segmentée. Par exemple, la case encadrée en rouge de D_{ref} dans la Figure 5.1 contient la distance euclidienne moyenne sur tous les sujets contrôles entre les barycentres du thalamus droit (1) et du thalamus gauche (2) dans l'espace 3D. Les attributs des arêtes sont calculés de manière à être supérieurs à 1 afin d'éviter des valeurs trop faibles (proches de 0) qui annuleraient les informations portées par les noeuds lors de l'opération de convolution du GNN (eq 1.5). Les architectures des GNNs implémentés dans chaque cas sont détaillées dans les sections 5.4 et 5.5.

En fonction du problème traité (classification ou régression), chaque graphe est associé à une classe y (1 si présence de paralysie cérébrale, 0 sinon) ou à un score moteur (y_1 pour la main affectée et y_2 pour la main "non-affectée"). Les graphes construits à partir

des données d'entraînement sont utilisés pour entraîner deux architectures de GNN. La première se focalise sur une classification de graphes pour détecter les enfants présentant une paralysie cérébrale (section 5.4). La seconde traite un problème de régression pour prédire le score moteur manuel des enfants ayant eu un AVC néonatal (section 5.5).

5.3 Base de données

Notre méthode est évaluée à partir de la cohorte *AVCnn* de 69 enfants de 7 ans [120, 140]. Cette cohorte inclut 31 enfants sains considérés comme contrôles et 38 enfants ayant eu un infarctus cérébral artériel néonatal (NAIS). Sur les 38 enfants avec lésion cérébrale précoce, 22 présentent une lésion dans l'hémisphère gauche et 16 dans l'hémisphère droit. Tous les enfants ayant eu un NAIS ont été examinés à 7 ans par un clinicien expérimenté (neurologue ou spécialiste de la rééducation en pédiatrie). Cet examen comportait une évaluation du comportement dont l'objectif était de diagnostiquer la présence d'une paralysie cérébrale (PC). La définition de la paralysie cérébrale fournie par le réseau de surveillance de la paralysie cérébrale en Europe (SCPE) [180] a été utilisée. Sur les 38 enfants avec lésion cérébrale précoce, 13 ont été diagnostiqués comme ayant une PC.

Durant cet examen médical, la dextérité manuelle grossière des enfants a été évaluée à travers un test commun en études cliniques : le *Box and Blocks Test (BBT)* [181] (Figure 5.2). Le BBT consiste en une boîte avec deux compartiments séparés en leur milieu. Au départ, 100 cubes se trouvent dans l'un des compartiments, du même côté que la main testée. Les enfants déplacent autant de cubes qu'ils le peuvent d'un compartiment à l'autre. Le résultat est le nombre de cubes déplacés dans le deuxième compartiment en 1 minute. Plus le score est élevé, meilleure est la fonction globale de la main testée. Il est à noter que 2 enfants n'ont pas passé le BBT. Ils ne seront donc pas considérés pour la seconde partie expérimentale traitant de l'estimation de la fonction motrice des enfants. Les enfants contrôles n'ayant pas passé le test, ils seront uniquement considérés dans la section 5.5 pour construire la matrice D_{ref} .

Les performances motrices étant influencées notamment par la motivation du sujet, nous avons calculé pour chaque enfant et chaque main un rapport du score au BBT (rapport décrit dans la section 5.5) afin de réduire l'impact d'autres facteurs externes à la fonction motrice et propres à l'individu (comportement, cognition, etc.) pouvant entraîner des variations dans le score brut au BBT. En générant un tel rapport de score au BBT, nous faisons l'hypothèse que, pour chaque enfant, l'impact potentiel de la motivation,



FIGURE 5.2 – Illustration¹du *Box and Blocks Test (BBT)* considéré pour évaluer la motricité manuelle des enfants avec lésion cérébrale précoce. L'objectif est de déplacer en un temps donné (1 minute) le maximum de cubes d'un compartiment à l'autre.

du comportement ou des troubles cognitifs est le même pour les deux mains. Ainsi, en calculant ce ratio, nous cherchons à réduire l'impact de ces paramètres pour évaluer plus précisément la motricité.

Les 69 enfants de la cohorte ont également passé un examen IRM (volumes 3D T1 de dimension 256x256x176 voxels avec une résolution spatiale de 1x1x1mm³). Une séquence 3D FLAIR (volumes 3D de dimension 230x230x160 voxels avec une résolution de 0.9x0.9x1mm³) a également été acquise. L'IRM d'un enfant avec lésion cérébrale précoce a été exclue de notre base de données en raison d'artefacts de mouvements trop importants, n'offrant pas une qualité suffisante pour la segmentation des noyaux gris centraux. Ainsi, sur les 38 enfants ayant eu un NAIS, nous en considérons 37 (comme dans le chapitre 4) pour l'étude relative à la détection de la PC et 35 pour l'estimation du score au BBT. La construction de la matrice D_{ref} s'appuie sur les IRMs des 31 enfants contrôles.

La segmentation de référence des noyaux gris centraux (thalamus, noyau caudé, putamen et pallidum) à partir des IRMs a été réalisée comme décrit dans le chapitre 4 (section 4.3.1), c'est-à-dire en combinant une segmentation manuelle avec l'outil ITKSnap [146] et une segmentation basée sur atlas (atlas *Hammersmith* [168]). La séquence FLAIR recalée a été utilisée comme aide visuelle pour affiner la segmentation sur IRM T1. Notons qu'il serait possible d'automatiser la méthode en considérant notre approche du chapitre 4 pour la segmentation des noyaux gris centraux.

1. <https://airgovie.com/>

5.4 Détection de la paralysie cérébrale chez les enfants après AVC néonatal

Dans cette section, nous présentons l’application de notre approche pour détecter la paralysie cérébrale chez des enfants ayant eu un infarctus cérébral artériel néonatal (NAIS) à partir de leur IRM cérébrale. Dans la section 5.4.1, nous complétons la méthode introduite en section 5.2 en détaillant l’architecture du GNN considéré pour réaliser la classification de graphes. Nous décrivons ensuite les expériences menées ainsi que les résultats obtenus dans la section 5.4.2.

5.4.1 Architecture du GNN considéré

Comme illustré par la Figure 5.3, l’architecture du GNN utilisé se compose de deux couches de convolution sur graphe suivies chacune d’une fonction d’activation ReLU et s’achève par une couche de *pooling* global (section 1.4.2), regroupant tous les noeuds du graphe G en un unique noeud dont l’attribut est la valeur maximum des attributs des noeuds $v \in V$, et une couche entièrement connectée pour la classification finale.

L’agrégation de l’information de voisinage se déroule durant les opérations de convolution, comme expliqué dans le chapitre 1. Comme dans les chapitres 2 et 4, nous avons considéré l’opérateur ECConv [68] du fait de sa capacité à gérer les attributs multidimensionnels sur les noeuds comme sur les arêtes (bien que dans ce cas, les attributs d’arêtes soient scalaires). La Figure 5.3 donne un exemple d’actualisation de la représentation d’état du noeud 1 (noeud cyan représentant le putamen gauche) lors de la première couche de convolution. Soit $X^0(1)$ la représentation du noeud 1 en entrée de GNN ($X^0(1)$ est donc de dimension 2 car il est constitué de 2 attributs correspondant au volume et à l’allongement) et soit $L((2, 1))$ l’attribut de l’arête allant du noeud 2 (voisin du noeud 1) au noeud 1. Les noeuds étant reliés deux à deux, chaque noeud possède un unique voisin. La nouvelle représentation d’état $X^1(1)$ du noeud 1 est calculée à partir de $X^0(1)$, de $X^0(2)$ et de $L((2, 1))$ selon la formule de l’eq 1.7. Pour cette application, nous avons considéré dans la fonction différentiable $F : \mathfrak{R} \rightarrow \mathfrak{R}^{d^l \times d^{l+1}}$ uniquement les couches nécessaires au *mapping* permettant de passer de l’espace de représentation des arêtes à l’espace de représentation des noeuds (voir fonction F du chapitre 2) car ceci s’est avéré suffisant pour cette tâche. La dimension d^1 des noeuds (en sortie de la première couche de convolution) est un hyperparamètre tandis que d^2 est fixée au nombre total de classe

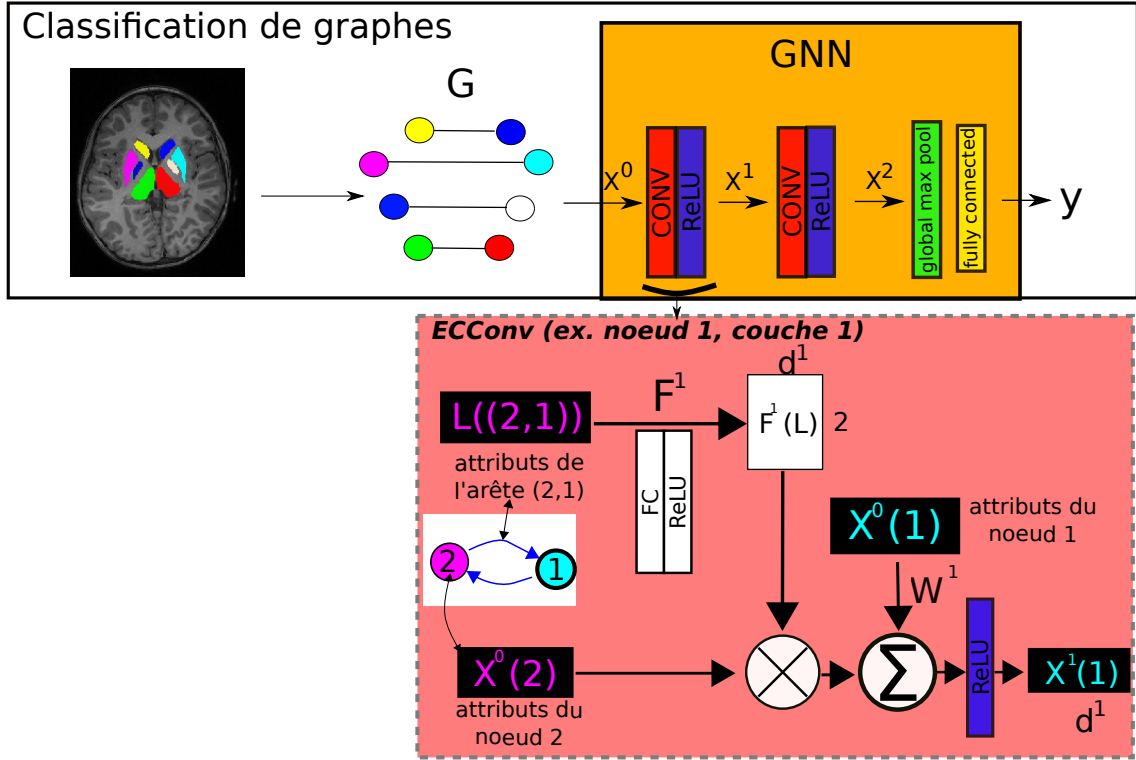


FIGURE 5.3 – Architecture du GNN considéré pour la classification de graphes dans le contexte de la détection de la paralysie cérébrale chez des enfants après AVC néonatal. Un exemple de la mise à jour de l’attribut $X(1)$ du noeud 1 (en cyan) lors de la première couche de convolution avec l’opérateur ECCConv est représenté.

C soit $d^2 = 2$ dans ce travail. En sortie de la deuxième couche de convolution, l’opérateur de sous-échantillonnage global par maximum fournit une sortie unique à l’échelle du graphe en considérant la valeur maximum des attributs de noeuds $v \in V$ dans chaque dimension : $y_g = \max_{v \in V} X^2(v)$. y_g (vecteur de dimension 2) passe ensuite par une couche linéaire (FC) dont la sortie \hat{y} est aussi de dimension 2. La classe y du graphe G est calculée comme : $y = \operatorname{argmax} \hat{y}$

5.4.2 Expériences

Avant d’appliquer notre méthode pour détecter la paralysie cérébrale chez des enfants ayant eu un NAIS, nous considérons une application plus simple consistant à distinguer les enfants après NAIS des enfants contrôles. Cette tâche simple (car faisable visuellement à partir de l’IRM dans la plupart des cas) est considérée pour évaluer la capacité de notre approche à distinguer deux populations visuellement distinctes.

5.4.2.1 Protocole expérimental

Les expériences sont une fois de plus menées avec le même environnement et avec les mêmes bibliothèques que celles du chapitre 2 et 4. Le modèle GNN, introduit dans la section précédente, est entraîné avec Adam sur 300 époques afin de minimiser la fonction de coût d’entropie croisée binaire. Une stratégie de réduction du taux d’apprentissage sur plateau est utilisée avec un taux d’apprentissage initial $\Xi_0 = 0.001$ et un facteur de réduction $\sigma = 5e^{-3}$. La dimension d^l des attributs de noeuds en sortie de la première couche de convolution est fixée à 5.

	Entraînement		Test	
Détection NAIS	Contrôles	NAIS	Contrôles	NAIS
Distribution	20	25	11	12
Détection PC	Pas de PC	PC	Pas de PC	PC
Distribution	16	9	8	4

TABLE 5.1 – Tailles des ensembles de données d’entraînement et de test pour les deux expériences, c’est-à-dire la détection de la lésion cérébrale précoce liée au NAIS et la détection de la paralysie cérébrale (PC). Pour les deux applications, 3 tirages stratifiés avec la distribution indiquée sont considérés.

La distribution de l’ensemble de données d’entraînement et de test pour les deux expériences de classification est présentée dans la Table 5.1. Une stratégie de validation croisée est mise en oeuvre pour tester notre méthode sur tous les enfants tout en ayant suffisamment de données d’entraînement. Pour faire face au déséquilibre de la base de données, notamment pour la détection de la paralysie cérébrale (13 enfants avec PC contre 24 sans), nous réalisons 3 échantillonnages aléatoires stratifiés (tous avec la distribution indiquée dans la Table 5.1). Les résultats seront moyennés sur ces 3 tirages. Pour la détection de la lésion cérébrale précoce parmi tous les enfants, seuls 20 enfants contrôles (contre 31 pour la détection de la PC et les estimations de score moteur) sont utilisés pour calculer la matrice de distances de référence D_{ref} (Figure 5.1) de manière à appliquer notre approche sur des sujets contrôles n’ayant pas servi à construire cette référence.

Pour évaluer la performance de notre méthode, nous mesurons la précision équilibrée de classification [182], définie comme la moyenne de la précision obtenue dans la classe majoritaire et dans la classe minoritaire. Nous étudions également l’apport des attributs de noeuds et d’arêtes pour vérifier la complémentarité de chaque information structurelle.

5.4.2.2 Résultats

	Pas d'attribut	Arêtes	Noeuds	Noeuds+Arêtes
NAIS	0.5	0.63	0.82	0.86
PC	0.5	0.67	0.85	0.89

TABLE 5.2 – Complémentarité des attributs de noeuds et d'arêtes pour la détection de la lésion cérébrale et de la PC (la valeur indiquée est la précision équilibrée).

Reconnaissance des enfants ayant eu un NAIS

La Table 5.2 (1^e ligne) rapporte les performances de classification obtenues. Nous observons que notre méthode atteint une précision équilibrée de 86%.

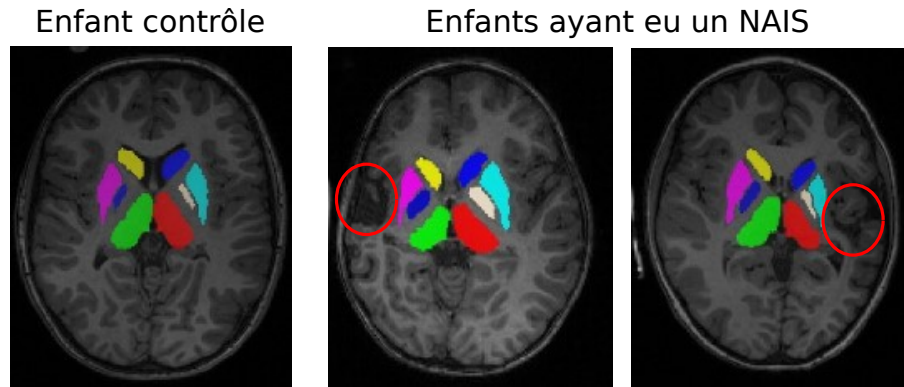


FIGURE 5.4 – Détection des enfants ayant eu un NAIS : exemples d'IRMs des deux classes. À gauche : cerveau contrôle avec une bonne symétrie entre les hémisphères. À droite : cerveau lésé (la lésion est entourée en rouge).

Il s'agit d'une tâche simple puisque la différence entre les deux classes peut être faite visuellement avec une IRM en détectant la lésion et en évaluant la symétrie entre chaque hémisphère (Figure 5.4). Néanmoins, elle valide la capacité de la méthode à classer les populations en s'appuyant sur les informations structurelles des noyaux gris centraux. On constate également la complémentarité de chaque attribut. En effet, dans la Table 5.2, nous observons que les attributs de noeuds augmentent la précision équilibrée de 50% à 82%. Même si la combinaison des attributs d'arêtes et de noeuds améliore peu les résultats (0.82 à 0.86), nous observons que les attributs d'arêtes relatifs aux distances aident à la classification (précision équilibrée de 0.63 contre 0.5 sans aucun attribut).

Détection de la paralysie cérébrale

La Table 5.2 (2^e ligne) rapporte les performances de classification obtenues. Nous observons que notre méthode atteint une précision équilibrée de 89% sur les 37 enfants ayant eu un NAIS. Les détails de classification sont indiqués dans la matrice de confusion de la Table 5.3.

Vérité \ Prédiction	Pas de PC	PC
	Pas de PC	22
PC	2	11

TABLE 5.3 – Matrice de confusion pour la classification des enfants ayant eu un NAIS comme atteint ou non de paralysie cérébrale (PC).

Sur les 24 enfants ne présentant pas de paralysie cérébrale, 22 sont correctement classifiés. De la même manière, notre méthode permet de détecter 11 enfants atteints de PC sur les 13 de la cohorte. Au total, notre approche se trompe pour 4 enfants, ce qui est un assez bon résultat compte tenu de la difficulté du problème mentionné précédemment et illustré dans la Figure 5.5. En effet, il est difficile de distinguer les populations puisque toutes deux présentent une lésion. La proximité de la lésion avec les noyaux gris centraux joue un rôle mais cela reste visuellement difficile de trancher sur la présence d'une PC.

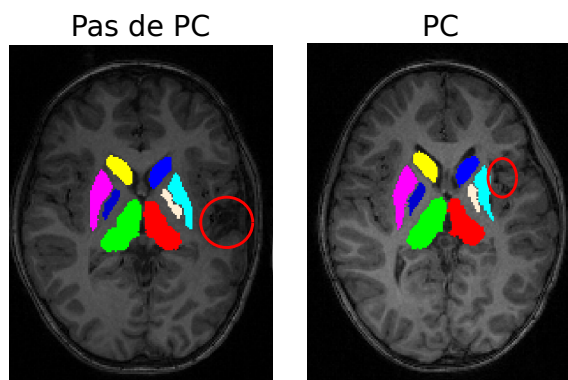


FIGURE 5.5 – Détection de la paralysie cérébrale (PC) : exemples d'IRMs des deux classes. À gauche : cerveau lésé sans paralysie cérébrale. À droite : cerveau lésé d'un enfant atteint de paralysie cérébrale. La lésion est entourée en rouge.

Nous pouvons encore une fois constater dans la Table 5.2 la complémentarité des attributs de noeuds et d'arêtes, améliorant la précision équilibrée de classification (précision de 67% avec les attributs d'arêtes, de 85% avec les attributs de noeuds et de 89% avec les

deux). Cette complémentarité des attributs est confirmée par la courbe ROC (Figure 5.6) correspondant aux performances de classification sur l'un des 3 tirages aléatoires stratifiés.

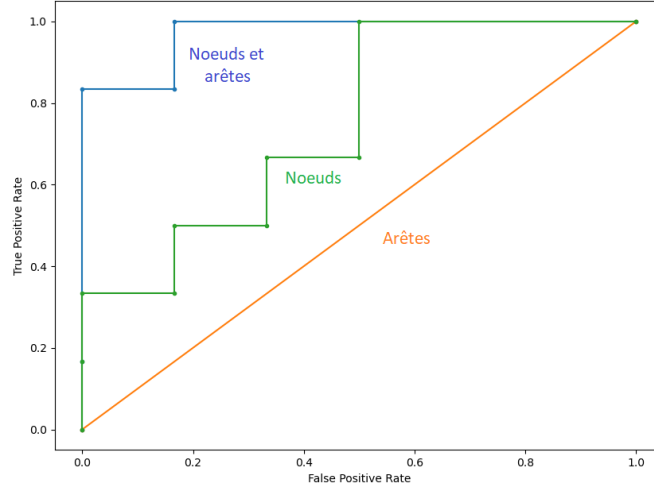


FIGURE 5.6 – Courbe ROC des performances de classification des enfants ayant eu un AVC néonatal (sur 1 tirage aléatoire stratifié) comme présentant ou non une paralysie cérébrale en fonction des attributs considérés dans la modélisation sous forme de graphes. Orange : attributs d'arêtes seuls (distances), vert : attributs de noeuds seuls (volumes et allongements), bleu : combinaison des attributs de noeuds et d'arêtes.

5.5 Estimation de la motricité manuelle des enfants après AVC néonatal

Dans cette section, nous prolongeons notre étude en cherchant à utiliser notre méthode pour estimer le score obtenu par les enfants ayant eu un AVC néonatal au *Box and Blocks Test* (BBT) pour les deux mains (la main affectée et la main dite "non-affectée") à partir de leur IRM cérébrale. Dans la section 5.5.1, nous complétons la méthode introduite en section 5.2 en détaillant l'architecture du GNN considéré pour résoudre le problème de régression sur graphes. Nous décrivons ensuite les expériences menées ainsi que les résultats obtenus dans la section 5.5.2.

5.5.1 Architecture du GNN considéré

Comme illustré par la Figure 5.7, de façon similaire à l'application de classification précédente, l'architecture du GNN proposé se compose de deux couches de convolution sur

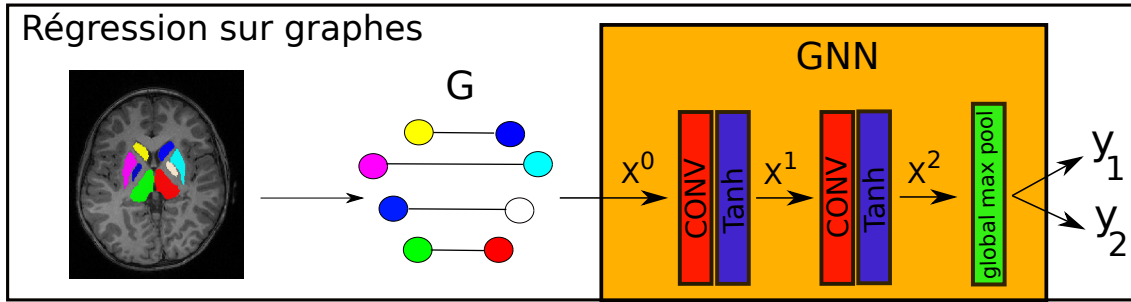


FIGURE 5.7 – Architecture du GNN considéré pour la régression sur graphes destinée à estimer le score de motricité manuelle obtenu au *Box and Blocks Test* par les enfants ayant eu un NAIS. y_1 correspond au score de la main affectée et y_2 au score de la main dite "non-affectée".

graphe suivies chacune d'une fonction d'activation tangente hyperbolique (observée plus pertinente expérimentalement pour estimer le rapport de score décrit en section 5.5.2) et s'achève par une couche de *pooling* global basé sur la valeur maximum des attributs des noeuds du graphe G pour fournir le score final y_1 ou y_2 . L'agrégation des informations de voisinage se fait une fois encore avec l'opérateur ECCConv [68]. L'actualisation de la représentation d'état des noeuds, illustrée sur la Figure 5.3, reste identique.

5.5.2 Expériences

Dans cette section, nous présentons nos expériences destinées à évaluer la pertinence de notre méthode pour estimer le score (ou plus précisément un rapport de score) obtenu par les enfants au BBT à la fois pour la main affectée et pour la main "non-affectée".

5.5.2.1 Protocole expérimental

Les expériences sont menées dans le même environnement et avec les mêmes bibliothèques que celles des chapitres 2 et 4. Le modèle GNN, introduit dans la section précédente, est entraîné avec Adam sur 250 époques afin de minimiser la fonction de coût d'erreur quadratique moyenne. Une stratégie de réduction du taux d'apprentissage sur plateau est utilisée avec un taux d'apprentissage initial $\Xi_0 = 0.001$ et un facteur de réduction $\sigma = 5e^{-3}$. Les dimensions d^1 et d^2 des attributs de noeuds en sortie de la première et de la deuxième couche de convolution sont fixées respectivement à 5 et 1. Deux GNNs sont entraînés pour prédire le score de chacune des deux mains au *Box and Blocks Test* (BBT). Dans les deux cas, une stratégie de validation croisée "leave-one-out" est mise en oeuvre

pour tester notre méthode sur tous les enfants tout en considérant la petite taille de la base de données disponible. Ainsi, à chaque fois, le modèle est entraîné avec 34 graphes et testé sur le graphe restant.

Soient LH le score brut de la main affectée au BBT et CLH le score brut de la main "non-affectée" au BBT. Nous avons dans un premier temps cherché à estimer le score brut obtenu au BBT par chacune des deux mains (architecture du GNN inchangée, exceptée la dernière fonction d'activation remplacée par un ReLU). Ce problème correspondait alors à la prédiction de $y_1 = LH$ (main affectée) et $y_2 = CLH$ (main "non-affectée"). Néanmoins, comme nous supposons que les performances motrices des enfants sont influencées par des facteurs externes aux fonctions motrices (ex. comportement, cognition), nous avons dans un deuxième temps chercher à prédire un rapport de score au BBT entre les deux mains. Pour chaque enfant, les valeurs de y_1 et y_2 à prédire étaient alors définies comme : $y_1 = \frac{LH}{LH+CLH}$ pour la main affectée et $y_2 = \frac{CLH}{LH+CLH}$ pour la main "non-affectée".

Pour évaluer les performances de la méthode, nous mesurons la distance L1 ou distance de Manhattan définie comme la différence absolue entre le score prédit et le score réel. La distance L1 permet de mettre l'accent sur une prédiction très éloignée du score réel. Pour mieux analyser la précision de la prédiction, nous mesurons la moyenne, la médiane et le maximum de la distance L1. Comme nous supposons qu'il existe une relation linéaire entre le score (rapport de score) prédit et le score (rapport de score) réel, nous mesurons le coefficient de corrélation de Pearson R [183] défini comme :

$$R = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5.2)$$

tel que \hat{y} est le score prédit, y est le score réel et n est le nombre de mesures (35 dans notre cas, correspondant aux 35 enfants avec lésion cérébrale précoce ayant passé un BBT). De plus, en raison du petit nombre n de données, nous indiquons dans chaque cas la *p-value* [184] associée afin de rejeter l'hypothèse nulle au niveau de significativité de 5%.

Enfin, nous nous sommes intéressés à la construction des graphes. Pour évaluer l'influence de chaque information (volumes, allongements, distances) dans la prédiction, nous mesurons les métriques précédentes pour différentes configurations de graphes obtenues en retirant certaines informations sur les noeuds ou sur les arêtes. Enfin, pour étudier l'impact de la topologie des graphes, mentionnée dans le chapitre 1 et discutée dans le chapitre 2, nous comparons nos performances à celles obtenues en considérant des graphes complets. Afin de compenser l'utilisation de graphes complets, l'opérateur de convolution

ECCConv est modifié pour ajouter la fonction F_1 introduite dans le chapitre 4 servant à attribuer un poids aux attributs des arêtes.

5.5.2.2 Résultats

Étude de la main affectée

La Figure 5.8 représente le score brut au BBT prédit par notre méthode en fonction du score brut réel obtenu par chaque enfant ayant eu un NAIS avec sa main affectée. On observe une corrélation positive ($R=0.645$, $p=2.94e^{-5}$) entre les deux variables. Néanmoins, si notre modèle parvient à distinguer les enfants avec une importante déficience motrice (score brut au BBT ≤ 15), il ne parvient pas à différencier les enfants ayant une meilleure motricité manuelle (c.a.d. avec un BBT ≥ 15). En effet, il prédit un score entre 25 et 30 pour tous les autres enfants de la base. Nous faisons l'hypothèse que cette difficulté à estimer le score brut au BBT des enfants n'ayant pas une atteinte motrice trop importante peut s'expliquer par des facteurs externes aux fonctions motrices (ex. comportement, cognition des enfants). Nous étudions donc le rapport de score y_1 associé à la main affectée prédit par notre modèle, censé être moins impacté par des facteurs externes car lié à la main "non-affectée". Les enfants ont passé le BBT pour les deux mains sur le même temps évitant ainsi une variable externe entre les deux mains liée à l'instant de test.

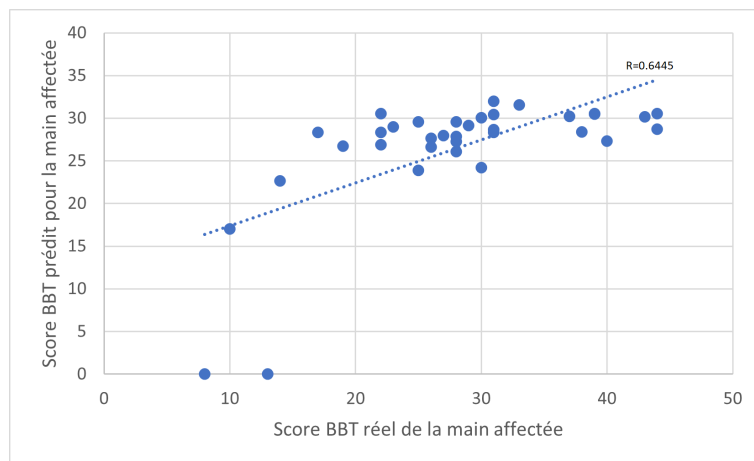


FIGURE 5.8 – Score brut au BBT prédit en fonction du score brut réel obtenu avec la main affectée de chaque enfant de la base.

La Figure 5.9 représente le rapport de score associé à la main affectée des enfants (y_1) prédit par notre modèle en fonction du rapport de score y_1 réel. Nous constatons que le rapport prédit par notre méthode est fortement et positivement corrélé avec le rapport

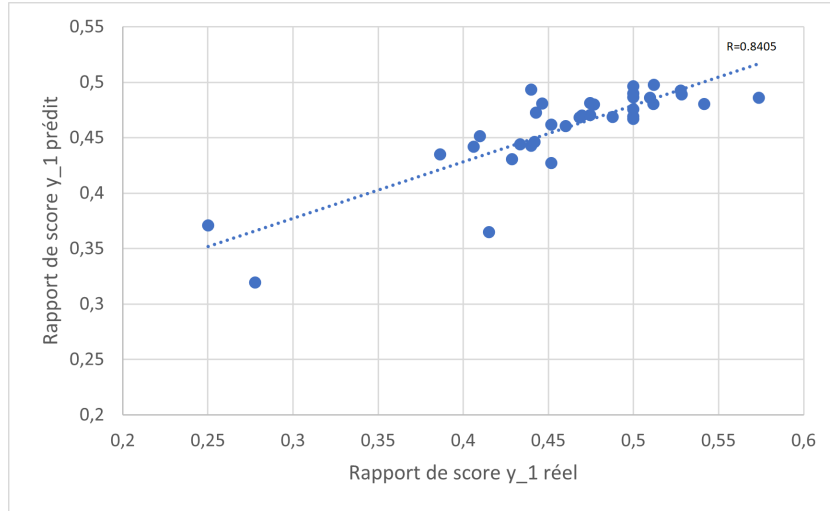


FIGURE 5.9 – Corrélation positive du rapport de score y_1 prédit associé à la main affectée de chaque enfant de la base en fonction du rapport de score y_1 réel.

de score réel ($R = 0.841$, $p = 2.66e^{-10}$). Le modèle parvient à prédire le rapport de score de la main affectée par rapport à la main "non-affectée" avec une distance L1 moyenne de 0.027 (Table 5.4).

	L1 moyenne	L1 médiane	L1 max	R (p-value)
Pas d'attribut	0.046	0.037	0.221	-0.560 ($4.62e^{-4}$)
Distances	0.047	0.035	0.222	-0.638 ($3.75e^{-5}$)
Allongements	0.040	0.031	0.112	0.700 ($2.88e^{-6}$)
Volumes	0.031	0.027	0.095	0.810 ($3.66e^{-9}$)
Volumes+Distances	0.033	0.028	0.117	0.744 ($3.03e^{-7}$)
Allongements+Distances	0.036	0.030	0.108	0.751 ($2.03e^{-7}$)
Volumes+Allongements	0.037	0.033	0.103	0.710 ($1.78e^{-6}$)
Notre proposition	0.027	0.024	0.121	0.841 ($2.66e^{-10}$)

TABLE 5.4 – Performance de prédiction du rapport de score y_1 associé à la main affectée en fonction des informations structurelles considérées. L1 : distance de Manhattan, R : coefficient de corrélation de Pearson.

Dans la Table 5.4, nous étudions l'impact de chaque information structurelle dans la prédiction du rapport de score associé à la main affectée. Nous constatons que le rapport y_1 prédit à partir de graphes intégrant uniquement des informations de volumes est fortement corrélé au rapport de score réel au BBT associé à la main affectée ($R = 0.810$, $p = 3.66e^{-9}$) et réduit la distance L1 maximum de la prédiction à 0.095. La combinaison des informations de volumes avec les distances uniquement ou les allongements uniquement

diminue le coefficient de corrélation ($R = 0.744$, $p = 3.03e^{-7}$ pour les volumes associés aux distances et $R = 0.710$, $p = 1.78e^{-6}$ pour les volumes associés aux allongements) ainsi que la précision de la prédiction (distance L1 moyenne de 0.033 et 0.037 respectivement). La considération des distances seules fournit une prédiction plus éloignée du rapport de score réel (distance L1 moyenne de 0.047), proche de celle obtenue en ne considérant aucune information structurelle (c.a.d. pas d'attribut). Cependant, la combinaison des allongements et des distances donne des résultats corrélés au rapport de score réel ($R = 0.751$, $p = 2.03e^{-7}$) et relativement précis (distance L1 moyenne de 0.036). Toutefois, les meilleures performances sont obtenues avec notre méthode, c'est-à-dire en combinant les 3 types d'informations structurelles.

La Table 5.5 permet de comparer les performances de notre méthode à celles obtenues en considérant des graphes complets. On observe que notre approche surpasse l'utilisation de graphes complets à tout point de vue (distance L1 moyenne de 0.027 contre 0.054 avec des graphes complets, $R=0.841$ contre $R=0.695$), et ce quelles que soient les informations intégrées dans les graphes complets. La Table A.3.1, fournie en annexe, complète l'étude en indiquant les performances obtenues par les graphes complets en fonction des informations structurelles considérées.

	L1 moyenne	L1 médiane	L1 max	R (p-value)
Graphes complets	0.054	0.057	0.162	0.695 ($3.58e^{-6}$)
Notre proposition	0.027	0.024	0.121	0.841 ($2.66e^{-10}$)

TABLE 5.5 – Comparaison des performances de prédiction associées à la main affectée de notre méthode aux résultats obtenus en considérant des graphes complets.

Ces résultats confirment l'importance de la topologie des graphes. Il semble que, pour cette étude, comparer toutes les structures entre elles ne soit pas adapté. Les très faibles résultats obtenus avec les graphes complets en considérant les volumes notamment (Table A.3.1 en annexe) confirment notre intuition selon laquelle il n'est pas cohérent de comparer les volumes de structures de natures distinctes.

Étude de la main "non-affectée"

La Figure 5.10 représente le score brut au BBT prédit par notre méthode en fonction du score brut réel obtenu par chaque enfant ayant eu un NAIS avec sa main "non-affectée". On n'observe pas de corrélation entre les deux variables ($R=0.232$, $p=0.18$). Notre modèle ne parvient pas à estimer le score brut obtenu au BBT avec la main considérée comme "non-affectée". On peut faire l'hypothèse que ce score est fortement impacté par des facteurs

externes (ex. comportement, cognition des enfants). Nous étudions donc le rapport de score y_2 associé à la main "non-affectée" prédit par notre modèle, censé être moins impacté par des facteurs externes car comparé au score de la main affectée. En effet, le rapport estimé caractérise le score au BBT de la main "non-affectée" normalisé par les scores des deux mains de l'enfant, réalisés sur le même temps de test.

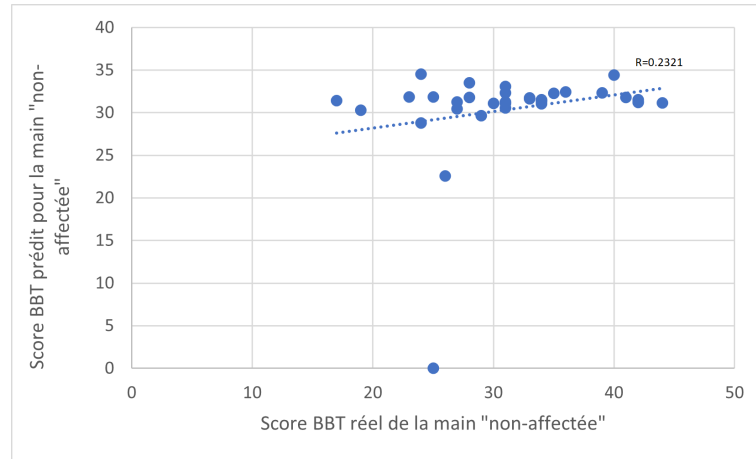


FIGURE 5.10 – Score brut au BBT prédit en fonction du score réel obtenu avec la main "non-affectée" de chaque enfant de la base.

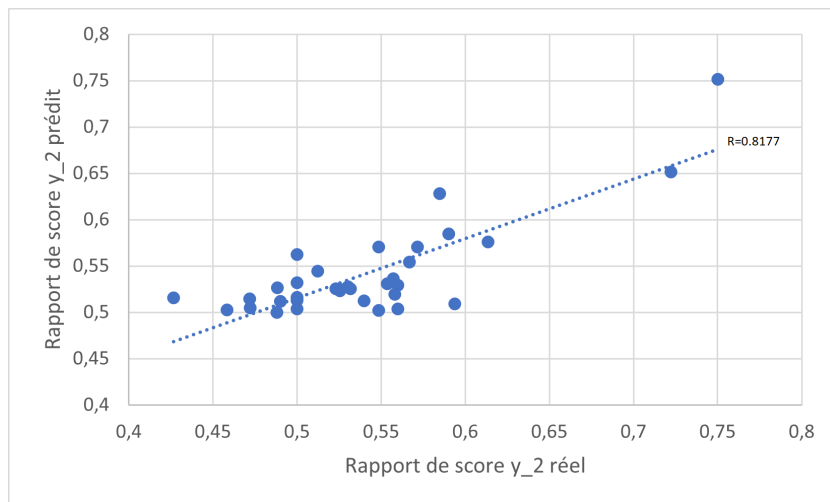


FIGURE 5.11 – Corrélation positive du rapport de score y_2 prédit associée à la main "non-affectée" de chaque enfant de la base en fonction du rapport de score y_2 réel.

La Figure 5.11 représente le rapport de score au BBT de la main "non-affectée" (y_2) prédit par notre modèle en fonction du rapport de score réel. Nous observons que le

rapport prédit par notre méthode est fortement et positivement corrélé avec le rapport de score réel ($R = 0.818$, $p = 2.02e^{-9}$). Notre modèle parvient à prédire le rapport de score associé à la main "non-affectée" avec une distance L1 moyenne de 0.028 (Table 5.6), c'est-à-dire avec la même précision que pour la main affectée.

	L1 moyenne	L1 médiane	L1 max	R (p-value)
Pas d'attribut	0.046	0.035	0.223	-0.810 ($3.70e^{-9}$)
Distances	0.039	0.028	0.155	0.596 ($1.61e^{-4}$)
Allongements	0.030	0.025	0.086	0.830 ($6.82e^{-10}$)
Volumes	0.044	0.036	0.210	0.404 (0.02)
Volumes+Distances	0.041	0.039	0.191	0.532 ($1.00e^{-3}$)
Allongements+Distances	0.033	0.028	0.104	0.755 ($1.64e^{-7}$)
Volumes+Allongements	0.031	0.022	0.115	0.806 ($4.97e^{-9}$)
Notre proposition	0.028	0.023	0.089	0.818 ($2.02e^{-9}$)

TABLE 5.6 – Performance de prédiction du rapport de score y_2 associé à la main "non-affectée" en fonction des informations structurelles considérées. L1 : distance de Manhattan, R : coefficient de corrélation de Pearson.

Dans la Table 5.6, nous étudions l'impact de chaque information structurelle dans la prédiction du rapport de score associé à la main "non-affectée". On constate que le rapport y_2 prédit à partir de graphes intégrant uniquement des informations d'allongements est très fortement corrélé au rapport de score réel au BBT associé à la main "non-affectée" ($R = 0.830$, $p = 6.82e^{-10}$) et réduit la distance L1 maximum de la prédiction à 0.086. Ainsi, la corrélation obtenue avec l'allongement uniquement est plus forte que si l'on tient compte de toutes les informations structurelles. Les prédictions obtenues en considérant les volumes seuls et les distances seules sont faiblement corrélées avec le rapport de score réel de la main "non-affectée" ($R = 0.404$ et $R = 0.596$ respectivement) et n'améliorent pas significativement la distance L1 par rapport au modèle sans information structurelle (Table 5.6) même lorsque les deux informations structurelles sont combinées (Volumes+Distances). Néanmoins, la combinaison de ces informations avec celles relatives aux allongements améliore la précision de la prédiction du modèle. Ainsi, la meilleure distance L1 médiane est obtenue en combinant volumes et allongements (distance L1 médiane de 0.022) et la meilleure distance L1 moyenne est atteinte par combinaison des 3 informations structurelles (distance L1 moyenne de 0.028). Notons que la forte corrélation négative apparente entre la prédiction de score sans aucun attribut et le rapport de score réel ($R = -0.810$, $p = 3.70e^{-9}$) reflète une très légère variation des scores prédits (le modèle prédit toujours un rapport autour de 0.530, étant le rapport de score moyen des données d'entraînement).

La prédiction n'est pas précise, comme le confirment les résultats relatifs à la distance L1.

	L1 moyenne	L1 médiane	L1 max	R (p-value)
Graphes complets	0.063	0.065	0.188	0.054 (0.76)
Notre proposition	0.028	0.023	0.089	0.818 (2.02e⁻⁹)

TABLE 5.7 – Comparaison des performances de prédiction associée à la main "non-affectée" de notre méthode aux résultats obtenus en considérant des graphes complets.

La Table 5.7 permet de comparer les performances de notre méthode à celles obtenues en considérant des graphes complets. On observe, ici encore, que notre approche surpasse l'utilisation de graphes complets à tout point de vue (distance L1 moyenne de 0.028 contre 0.063 avec des graphes complets, R=0.818 contre R=0.054), et ce quelles que soient les informations intégrées dans les graphes complets. La Table A.3.2, fournie en annexe, complète l'étude en indiquant les performances obtenues par les graphes complets en fonction des informations structurelles considérées. Ces résultats montrent l'importance de la topologie des graphes et confirment que, pour cette étude, il n'est pas cohérent de comparer toutes les structures entre elles. Comme pour la main affectée, les très faibles résultats obtenus avec les graphes complets en considérant les informations de volumes (Table A.3.2 en annexe) confirment notre intuition selon laquelle cela n'a pas de sens de comparer les volumes de structures de natures distinctes.

5.6 Discussion

Dans ces travaux, nous avons montré une relation existante entre l'organisation macrostructurelle des noyaux gris centraux et les fonctions motrices des enfants après AVC néonatal. En effet, à partir d'une IRM cérébrale, nous proposons de modéliser sous forme de graphe les informations structurelles (volumes, allongements et distances) spécifiques aux structures des noyaux gris centraux avant d'exploiter ces graphes par un GNN. L'objectif de la méthode est d'aider à étudier la motricité des enfants avec lésion cérébrale précoce. Ainsi, nous avons montré dans la section 5.4 que notre approche permet de détecter dans une cohorte d'enfants de 7 ans les sujets ayant eu un infarctus cérébral artériel néonatal (NAIS) avec une précision équilibrée de 86% mais aussi de reconnaître dans la plupart des cas (89%), parmi les enfants avec NAIS, ceux présentant une paralysie cérébrale. Nous avons montré dans la section 5.5 la capacité de notre méthode à prédire la fonction motrice (évaluée par *Box and Blocks Test*) de chacune des deux mains des enfants

ayant eu un NAIS (main affectée et main "non-affectée"). Les résultats montrent une forte corrélation entre les rapports de score prédits par notre méthode et les rapports de score réels des deux mains ($R > 0.81$), ainsi qu'une précision relativement élevée de la prédiction (distance moyenne L1 < 0.03). L'une des principales conclusions de cette étude repose sur l'importance de considérer l'ensemble de l'organisation macrostructurelle des noyaux gris centraux (incluant certaines propriétés des noyaux gris centraux ainsi que leurs relations spatiales) pour étudier la motricité des enfants. En effet, pour la détection de la PC tout comme pour la prédiction de la fonction motrice manuelle, nous avons montré l'intérêt de combiner les 3 informations structurelles. C'est la première fois, à notre connaissance, que nous pouvons montrer une relation solide entre les caractéristiques structurelles des noyaux gris centraux et la fonction motrice de la main "non-affectée".

Les résultats relatifs à l'estimation des rapports de score au BBT (section 5.5), considérant les diverses configurations de graphes (avec plus ou moins d'informations structurelles), nous interrogent sur certains aspects. Nous avons observé que le volume des noyaux gris centraux est fortement corrélé à la fonction motrice de la main affectée, mesurée par le rapport de score y_1 au BBT. Ce résultat confirme des études antérieures [119, 174] montrant que la perte de volume des noyaux gris centraux est associée à une altération de la fonction de la main après un AVC néonatal. La réduction du volume des noyaux gris centraux après NAIS pourrait être liée à la perte neuronale (due à la dégénérescence Wallérienne) ou à une arborisation dendritique réduite. En plus de l'influence du volume, nous avons également trouvé une relation entre la fonction motrice de la main affectée et les caractéristiques structurelles des noyaux gris centraux : leur forme (représentée par l'allongement de la structure) et les distances entre structures supposées symétriques. Même si les distances ne suffisent pas à elles seules à prédire la fonction motrice des deux mains, leur combinaison avec d'autres informations structurelles (volumes et allongements) permet d'améliorer la prédiction du rapport de scores. Ainsi, les graphes intégrant tous les attributs donnent toujours de meilleurs résultats que les graphes ne comportant que des attributs sur les noeuds dans les Tables 5.4 et 5.6. Nous pouvons supposer que la façon dont les distances sont considérées, c'est-à-dire comme un rapport avec les distances moyennes D_{ref} chez les enfants contrôles, n'est pas assez discriminante pour distinguer les enfants. De plus, comme ces rapports de distances sont de petites valeurs (souvent < 1.5), ils ont peu d'impact dans la prédiction du GNN. Cependant, les informations de distances sont utiles, combinées aux autres informations structurelles, car elles peuvent faire la différence pour distinguer des enfants ayant des structures cérébrales avec des volumes

et des allongements similaires mais une organisation spatiale différente (ex. asymétrie). Avec ces résultats, nous soulignons l'importance de préserver la forme globale des noyaux gris centraux pour préserver la fonction motrice après une lésion cérébrale précoce. Cela confirme indirectement la vulnérabilité spécifique de ces structures profondes aux lésions cérébrales précoces [185] et leur rôle dans la modulation des fonctions motrices.

Il semble que les changements morphométriques (forme et distance) des noyaux gris centraux impactent fortement les fonctions motrices des 2 mains, en plus des variations de volume. Dans le cas de la main "non-affectée", ces caractéristiques sont davantage corrélées à la fonction motrice que le volume de ces structures. Cette observation, bien que évidente dans nos résultats (Table 5.6), est difficile à expliquer. Nous pouvons imaginer qu'elle peut être liée à des variations du développement cérébral dans les deux hémisphères après apparition de la lésion unilatérale. De plus, cela suggère que même s'il n'y a pas de lésion (par définition) dans l'hémisphère contra-lésé ("sain"), il souffre toujours des conséquences de l'AVC néonatal. D'une certaine manière, nous confirmons que « l'AVC néonatal ne semble pas être une lésion focale mais une lésion qui affecte l'ensemble du cerveau en développement » [186]. Enfin, nous avons constaté la complémentarité des 3 informations structurelles pour les deux applications considérées dans ce chapitre. En effet, dans certains cas, la suppression de l'une de ces informations conduit à de moins bons résultats qu'avec une seule information (ex. main "non-affectée" : meilleure performance de l'allongement seul que de l'allongement combiné aux distances, main affectée : meilleure performance du volume seul que du volume combiné aux distances). Cela soulève la question d'un probable "*principe de Boucles d'or*" [187] relatif à l'information structurelle à utiliser pour établir une relation avec les fonctions motrices du cerveau. Ce principe fait référence au fait de trouver le "juste milieu" et se rencontre dans de nombreuses disciplines (ex. zone habitable autour d'une étoile en astrobiologie, dosage des médicaments en médecine, fixation des prix en économie, dosage de l'information (quantité, difficulté) en psychologie du développement). Dans notre contexte, on s'aperçoit qu'il est nécessaire de bien identifier l'information structurelle pertinente pour que notre modèle puisse faire une prédiction suffisamment précise. En effet, les performances se dégradent s'il n'y a pas assez d'informations (Tables 5.2, 5.4 et 5.6) ou s'il y a trop d'informations dans les graphes (graphes complets - Tables A.3.1 et A.3.2 fournies en annexe).

Toutefois, la méthode présentée dans ce chapitre présente des limites dont nous sommes conscients. Tout d'abord, notre approche est basée uniquement sur l'organisation structurelle des structures supposées symétriques puisque les graphes sont connectés deux à

deux. Nous avons vu dans la section 5.5 que l'utilisation de graphes complets portant ces informations structurelles ne fonctionne pas (Tables 5.5 et 5.7). Néanmoins, il serait intéressant de réussir à intégrer l'organisation intra-hémisphérique des noyaux gris centraux (distances, angles entre structures). Pour cela, il faudrait complexifier l'architecture du GNN pour ne pas comparer les volumes et les allongements entre structures intra-hémisphériques (ce qui n'a pas de sens et prête à confusion pour le GNN, comme vu lors des expériences), même si elles sont reliées par une arête. Pour ce faire, nous pourrions envisager de considérer différents types d'arêtes pendant l'opération de convolution du GNN, comme proposé dans RGCNConv [188], afin de distinguer les connexions entre les structures inter-hémisphériques et les connexions entre les structures intra-hémisphériques. De plus, les graphes considérés étant relativement simples, on pourrait envisager de les enrichir avec d'autres informations structurelles et avec d'autres régions cérébrales profondes liées aux fonctions motrices (ex. le faisceau pyramidal [189]).

Ensuite, nous nous contentons dans ce chapitre de présenter la méthode et de montrer sa capacité à aider dans l'étude de la motricité des enfants après AVC néonatal. Cependant, pour la détection des enfants avec paralysie cérébrale (section 5.4) comme pour l'estimation des scores moteurs au BBT (section 5.5), nous n'avons pas comparé nos performances à d'autres techniques existantes, non basées sur des graphes, pour la classification ou la régression. Ainsi, dans de futurs travaux nous prévoyons de comparer nos performances à celles obtenues par des classifieurs classiques d'apprentissage machine (ex. SVM, forêt aléatoire, TabFPN [190]) et par des CNNs (section 1.2.2) opérant directement sur les images.

Enfin, les résultats ont été obtenus sur une cohorte homogène en termes de caractéristiques des patients (enfants nés à terme âgés de 7 ans), de caractéristiques des lésions (lésion sous-corticale apparue en période néonatale) et de conditions d'imagerie (acquisition de données monocentriques avec un scanner IRM 3T). Par conséquent, nos résultats ne sont pas brouillés par les conséquences développementales de la prématurité, la diversité de l'acquisition des compétences motrices avec l'âge ou la variabilité de l'acquisition IRM. Ces résultats doivent être confirmés sur plus d'enfants. Il serait intéressant de voir si notre méthode reste efficace à des stades de développement plus précoces, pour évaluer la motricité des enfants et diagnostiquer la paralysie cérébrale dès le plus jeune âge de l'enfant, en vue de proposer des interventions thérapeutiques adaptées le plus tôt possible.

5.7 Conclusion

Dans ce chapitre, nous avons proposé une stratégie reposant sur des réseaux de neurones sur graphes (GNNs) pour détecter, à partir d'une IRM, la présence de paralysie cérébrale chez des enfants après AVC néonatal ainsi que pour estimer leur fonctions motrices uni-manuelles des 2 mains, évaluées par *Box and Blocks Test* (BBT), en considérant l'organisation macrostructurale inter-hémisphérique des noyaux gris centraux (volumes, allongements, distances). Nous avons mis en évidence que la motricité des enfants, et particulièrement leur fonction motrice manuelle, est corrélée avec l'organisation structurelle générale des noyaux gris centraux, plus qu'avec le volume seul de ces structures comme cela était étudié jusque là. Ainsi, notre méthode parvient à détecter avec précision (89% de précision équilibrée) les enfants avec paralysie cérébrale et à prédire un rapport de score au BBT, pour chaque main, avec une grande précision (distance L1 moyenne < 0.03). Enfin, nous avons constaté que la motricité de chacune des mains (la main affectée et la main "non-affectée") est différemment influencée par les informations structurelles. Le score de la main affectée semble davantage corrélé aux volumes tandis que celui de la main "non-affectée" est lié aux allongements des structures. Ces résultats confirment la pertinence d'étudier les caractéristiques macrostructurelles des noyaux gris centraux en IRM en tant que biomarqueur précoce pour prédire la fonction motrice après une lésion cérébrale précoce. Il est à noter que les travaux relatifs à la détection de la paralysie cérébrale ont fait l'objet d'un article accepté pour communication à la conférence internationale *IEEE ISBI (International Symposium on Biomedical Imaging)* 2023.

Avec ces travaux, nous progressons dans la chaîne de traitement destinée à l'étude de la lésion cérébrale précoce (Figure II.1) puisque nous fournissons une nouvelle méthode pour caractériser le cerveau des enfants après AVC néonatal. L'étude de l'organisation macrostructurale des noyaux gris centraux, combinée à des analyses morphométriques, contribue à proposer et à évaluer l'efficacité de nouvelles interventions thérapeutiques (rééducation, traitements curatifs).

À retenir

- L'analyse des noyaux gris centraux, en IRM, pour étudier la motricité des enfants après AVC néonatal est fréquente. Les travaux existants se focalisent sur le volume des structures et donnent des résultats controversés. Nous proposons d'étudier l'organisation macrostructurale générale des noyaux gris centraux (volumes, allongements, distances).
- Les informations structurelles des noyaux gris centraux sont modélisées sous forme de graphes exploités par des GNNs pour deux applications : une classification de graphes (détection de la paralysie cérébrale) et une régression sur graphes (évaluation de la motricité des 2 mains).
- Notre approche parvient à estimer un rapport de score (évalué par le BBT) entre chaque main des enfants avec précision (distance L1 < 0.03) et à détecter les enfants présentant une paralysie cérébrale dans la plupart des cas (à 89%).
- La motricité de la main affectée semble fortement corrélée aux volumes des structures tandis que celle de la main "non-affectée" est davantage liée aux allongements. Les 3 informations structurelles s'avèrent complémentaires pour une prédiction précise.
- La méthode reste à confirmer sur d'autres populations (enfants plus jeunes, adultes avec AVC) et à être comparée à d'autres techniques dissociées des graphes (CNNs, classifieurs).

CONCLUSION, LIMITES ET PERSPECTIVES

Conclusion

Durant la thèse présentée à travers ces pages, nous avons étudié l'apport des connaissances structurelles, combinées à des techniques d'apprentissage profond, pour diverses tâches d'analyse d'images. Dans la première partie du manuscrit, nous avons apporté des éléments de réponse à notre première interrogation, mentionnée dans l'introduction, liée à l'apport des connaissances structurelles de haut-niveau pour l'analyse d'images en général et à leur intégration sous forme de graphes manipulés par des réseaux de neurones sur graphes (GNNs). Le chapitre 1 a permis de rappeler des notions relatives à l'apprentissage profond, aux connaissances structurelles, à leur modélisation sous forme de graphes et à l'exploitation de ces graphes par des GNNs. Nous avons ainsi rappelé que les réseaux de neurones, notamment convolutifs (CNNs), pouvaient s'avérer efficaces pour différentes tâches d'analyse d'images (ex. classification, segmentation sémantique) même face à un faible nombre de données d'entraînement grâce à diverses stratégies (ex. augmentation de données, apprentissage par transfert). Néanmoins, nous avons vu que la manière d'exploiter l'information spatiale par les CNNs et autres architectures profondes basées sur des mécanismes d'attention restait limitée et pouvait parfois conduire à des erreurs. Nous avons montré que les connaissances structurelles dites de "haut-niveau" pouvaient être modélisées sous forme de graphes destinés à guider l'interprétation de l'image à l'aide de différentes approches, parmi lesquelles on retrouve les réseaux de neurones sur graphes. Nous nous sommes intéressés à l'architecture de ces réseaux sur graphes, en insistant sur le choix de l'opérateur de convolution ainsi que sur la topologie des graphes considérés. Partant de ces acquis, nous avons proposé, dans le chapitre 2, une méthode de post-traitement, basée sur un GNN, pour améliorer la segmentation sémantique d'images fournie par des réseaux de neurones profonds dédiés (CNNs), à travers une tâche de classification de noeuds. L'approche proposée s'appuie sur la carte de segmentation fournie par un CNN et associe chaque région détectée par ce dernier à un noeud d'un graphe.

Les vecteurs d'appartenance issus de la sortie du CNN sont exploités comme attributs de noeuds tandis que les arêtes reliant les noeuds indiquent des informations relatives aux relations spatiales entre régions détectées. Nous avons montré que notre méthode était pertinente et améliorait la qualité des segmentations issues des CNNs (en particulier par une réduction significative de la distance d'Hausdorff). Nous nous sommes intéressés à l'impact de la nature de l'opérateur de convolution ainsi que de la topologie des graphes. Nous avons observé qu'un élagage des arêtes permettait de limiter les informations structurelles considérées lors de l'agrégation de voisinage, conduisant ainsi le plus souvent à de meilleurs résultats. Nous avons également constaté qu'un opérateur comme ECConv, considérant à la fois les attributs multi-dimensionnels sur noeuds et sur arêtes, se montrait plus robuste à la topologie des graphes que d'autres opérateurs moins sophistiqués (ex. GCNConv, GraphConv). La pertinence de notre proposition a été évaluée sur deux bases de données publiques (FASSEG et IBSR), avec plusieurs tailles de base d'entraînement et différentes architectures de CNNs. Nous avons montré que notre approche, combinée à un réseau classique comme U-Net entraîné sur peu de données, pouvait surpasser des réseaux plus complexes (ex. *Visual Transformers*) entraînés sur davantage de données.

Les contributions décrites dans la deuxième partie de ce manuscrit s'emploient à apporter des éléments de réponse à notre seconde interrogation : Dans quelle mesure la combinaison de l'apprentissage profond et des informations structurelles de haut-niveau peut-elle aider dans l'étude de la lésion cérébrale précoce ? Nous avons vu dans cette partie que l'utilisation de modèles animaux était nécessaire pour aider à la compréhension des mécanismes *in vivo* de la lésion cérébrale. Nous avons également souligné que le porcelet était un modèle adapté pour l'étude du développement neurologique précoce de par les similarités de son cerveau avec celui du jeune enfant (morphologie et croissance périnatale, présence de plis, etc.). Néanmoins, nous avons montré que le porcelet restait difficile à étudier à cause de l'absence d'outils automatisant l'entièreté de la chaîne de segmentation du cerveau de cet animal. Ainsi, dans le chapitre 3 de cette thèse, nous avons proposé une méthode basée sur un modèle U-Net 3D par patches pour automatiser l'extraction de cerveau du porcelet sur IRM, étape préliminaire de la chaîne de traitement, cruciale à la segmentation des structures et tissus cérébraux. Nous avons mis en évidence que l'utilisation, durant l'entraînement du modèle, de patches 3D se chevauchant et extraits selon une stratégie adaptée, permettait d'enrichir les informations contextuelles tout en réduisant la mémoire requise par rapport à une approche 3D traditionnelle. Nous avons montré que notre solution surpassait des approches d'apprentissage profond 2D et d'autres outils

standards spécifiques à l’humain (BET). Nous avons aussi proposé une stratégie d’apprentissage par transfert pour utiliser les performances d’un réseau entraîné sur de jeunes porcelets à un stade de développement plus avancé de l’animal afin d’assurer un suivi longitudinal automatisé malgré le peu de données à disposition.

Enfin, nous avons vu que l’AVC néonatal était un bon modèle d’étude de l’impact de la lésion cérébrale précoce du fait des caractéristiques bien connues de la lésion cérébrale découlant de l’AVC néonatal (lésion corticale ou sous-corticale non progressive intervenant toujours en période périnatale chez des enfants nés à terme). Nous avons vu que la lésion cérébrale induite par l’AVC néonatal pouvait, dans environ 30% des cas, entraîner le développement de troubles moteurs permanents regroupés sous le terme de paralysie cérébrale (PC). Ainsi, nous avons, dans nos expériences, considéré une cohorte d’enfants de 7 ans ayant subi ou non un AVC néonatal (cohorte *AVCnn* [120, 140]). Dans une volonté de comprendre la motricité des enfants suite à un AVC néonatal, nous nous sommes intéressés aux noyaux gris centraux du fait du rôle clé de ces structures dans les fonctions sensori-motrices. Nous avons constaté que la segmentation de ces structures profondes, sur IRM, chez le jeune enfant, notamment après lésion cérébrale précoce, était difficile pour diverses raisons (faible contraste tissulaire, présence d’une lésion, peu de données) induisant en erreur les CNNs dédiés à cette tâche. Ainsi, dans le chapitre 4, nous proposons d’appliquer notre post-traitement de segmentation, exploitant les informations structurales par un GNN, compte tenu de l’organisation connue et relativement stable des noyaux gris centraux. Nous avons ajouté une fonction dans l’opérateur de convolution, destinée à apprendre le poids de chaque arête dans la prédiction du GNN, et à permettre l’utilisation de graphes complets. Nous avons vu que notre méthode permettait de corriger, dans une certaine mesure (à l’exception des cerveaux avec une très grosse lésion), les incohérences spatiales produites par les CNNs et qu’elle surpassait des méthodes récentes d’apprentissage profond pour la segmentation des noyaux gris centraux. Une fois les noyaux gris centraux segmentés, nous avons étudié dans le chapitre 5 la relation entre l’organisation macrostructurelle (volumes, allongements, distances) de ces structures profondes et la motricité des enfants ayant eu un AVC néonatal. Nous avons modélisé ces informations structurales sous forme de graphes pour traiter ensuite, grâce à un GNN, deux problèmes médicaux. Le premier consistait à détecter la paralysie cérébrale (PC), par ailleurs diagnostiquée cliniquement par un médecin expert, chez les enfants après AVC néonatal par une classification de graphes. Les expériences ont permis de détecter les enfants atteints de PC dans 89% des cas. Le deuxième problème avait pour objectif d’évaluer, à partir

de l'organisation macrostructurelle des noyaux gris centraux, la motricité uni-manuelle des deux mains des enfants par une régression sur graphes. Dans nos expériences, nous sommes parvenus à estimer un rapport de score entre chaque main (évalué par *Box and Blocks Test*) avec une bonne précision. Nous avons mis en évidence une corrélation forte entre le score prédit par notre modèle et le score réel obtenu par les enfants, traduisant l'existence d'une relation entre la macrostructure des noyaux gris centraux et les fonctions motrices de l'enfant.

Par ces contributions relatives à la segmentation et l'analyse de cerveaux humains comme animaux, nous espérons aider dans l'étude de la lésion cérébrale précoce, notamment dans la proposition et l'évaluation d'interventions thérapeutiques adaptées (rééducation, traitements curatifs).

Les contributions présentées dans ce manuscrit ont fait l'objet d'un article de revue (chapitre 3 [158]) et de trois communications dans des conférences nationales (chapitre 3 [166]) et internationales (chapitres 2 et 4 [108], chapitre 5 (article accepté à *ISBI 2023*)). Nous présentons ci-après les limites et perspectives de ces travaux.

Limites et perspectives

Les travaux présentés dans ce manuscrit ont montré l'intérêt d'associer les connaissances structurelles et l'apprentissage profond, sous forme de graphes, pour l'analyse d'images, en particulier pour la segmentation sémantique et pour l'étude en neuroimagerie de la lésion cérébrale précoce. Mais certaines limites de nos contributions sont à souligner.

Tout d'abord, de nombreux aspects sont à discuter sur notre méthode de post-traitement pour la segmentation sémantique d'images présentée dans les chapitres 2 et 4. Une première faiblesse concerne la définition des régions, associées aux noeuds d'un graphe, à partir de la segmentation initiale fournie par un CNN. En effet, nous définissons ces régions à partir de la classe d'appartenance définie par le CNN et nous donnons comme attributs aux noeuds du graphe ce vecteur de probabilité d'appartenance. En donnant une telle importance à la prédiction du CNN, nous risquons de ne pas parvenir à corriger certaines erreurs de segmentations ou à ne pas distinguer des sous-régions de différentes classes identifiées comme une seule région par le CNN. Cela explique peut-être, en partie, pourquoi notre méthode a principalement un impact sur la distance d'Hausdorff (correction des artefacts et des régions mal classifiées) et limité sur le Dice (les régions mal définies regroupant des sous-régions de classes différentes restent mal définies et associées

à une unique classe). Une première solution à ce problème serait de considérer l’incertitude du CNN lors de la définition des régions [105]. Ainsi, à partir d’un seuil définissant une limite d’ambiguïté, il serait possible de rejeter d’une région les pixels considérés comme incertains et de les associer à une classe temporaire incertaine. En complément de cette solution, pour moins dépendre de la prédiction du CNN, il serait envisageable de fournir aux noeuds du graphe (associés aux régions segmentées par le CNN) d’autres attributs indépendants des probabilités fournies par le CNN (ex. intensité moyenne des pixels de la région). Selon nous, cela supposerait un enrichissement du modèle avec davantage d’informations sur les noeuds (ex. intensités) et sur les arêtes (ex. angles) s’accompagnant, peut-être, d’une intégration de connaissances *a priori* dans le modèle (ex. distance théorique entre les yeux, volume théorique du thalamus), comme mentionné dans [191]. On peut même imaginer que ces relations structurelles pourraient être apprises par le GNN lui-même. Le GNN, manipulant des graphes modélisant un grand ensemble de relations structurelles apprendrait, grâce à une fonction de coût adaptée, à donner plus de poids aux informations structurelles pertinentes.

Une autre faiblesse de la méthode présentée réside sur la topologie des graphes considérés. En effet, dans les chapitres 2 et 4, nous n’avons pas employé la même stratégie. Selon nous, la stratégie basée sur la fonction d’ajustement du poids des arêtes dans l’opérateur ECCConv (chapitre 4) est plus générique car elle ne dépend pas d’un hyperparamètre, appris indépendamment du modèle GNN, tel que le rayon d’élagage ρ utilisé dans le chapitre 2. L’ajout de la fonction, dénommée F_1 dans le manuscrit, s’est fait tardivement durant les travaux et nous avons fait le choix de ne pas refaire les expériences du chapitre 2 avec cette stratégie pour le moment. Néanmoins, cette dernière s’étant avérée efficace pour la segmentation des noyaux gris centraux, nous présumons qu’elle le sera également pour les jeux de données FASSEG et IBSR. Cette vérification sera à réaliser pour homogénéiser la méthode sur toutes les applications et devra être confirmée sur d’autres bases de données usuelles en traitement d’images telles que *Pascal VOC* [106] ou *Cityscapes* [107]. Il serait également intéressant de considérer d’autres méthodes intégrant l’élagage des arêtes dans le GNN (apprentissage des arêtes pertinentes par rétro-propagation) [77, 78, 79].

Les expériences du chapitre 4 ont mis en évidence une autre limitation de notre méthode liée à la présence d’une classe inattendue (lésion cérébrale), de volume considérable, déformant la structure générale du cerveau. Nous avons observé que dans ces situations, notre approche s’avère moins efficace qu’un CNN seul et a tendance à faire disparaître des classes. Dans des travaux futurs, nous envisageons d’utiliser des algorithmes existants

pour segmenter préalablement la lésion du cerveau [172] et d'utiliser ce masque de la lésion comme région supplémentaire dans notre modélisation sous forme de graphes afin que le GNN apprenne mieux à classifier les noyaux gris centraux en présence d'une lésion. Pour entraîner le GNN à faire face à d'importantes déformations structurelles, il sera peut-être nécessaire de complexifier son architecture afin de permettre davantage de degrés de liberté pour supporter les variabilités liées à la lésion.

Enfin, nous avons justifié l'intérêt de notre méthode de post-traitement par le fait qu'elle peut s'appliquer à la sortie de n'importe quel CNN produisant une carte de segmentation, sans nécessiter le développement d'une architecture spécifique basée sur le CNN (comme dans [86, 87, 88, 89]), et sans intégrer d'algorithmes additionnels de sur-segmentation produisant des superpixels (comme dans [83, 84, 85]). De cette manière, nous utilisons les connaissances structurelles dans un objectif de correction et non d'anticipation des erreurs. Il se pose alors la question : est-ce le bon moment pour intégrer ces informations structurelles ? Ne faudrait-il pas les intégrer plus tôt dans la chaîne de traitement ? Pour répondre à cette question, il serait nécessaire de comparer les résultats de notre approche à ceux obtenus par les méthodes intégrant les relations spatiales des images, à l'aide de GNNs, dans l'espace des caractéristiques des réseaux de neurones profonds [86, 87, 88, 89].

Concernant la méthode de prédiction de la motricité des enfants après AVC néonatal basée sur les informations structurelles introduite dans le chapitre 5, beaucoup de perspectives peuvent être évoquées. Il est important de préciser que nous avons, pendant la thèse, uniquement cherché à prouver la pertinence de la méthode. Par conséquent, les résultats obtenus pour la détection de la paralysie cérébrale comme pour l'estimation des scores moteurs manquent de comparaisons avec d'autres méthodes indépendantes des graphes tels que des classifieurs classiques d'apprentissage machine (ex. SVM, forêt aléatoire, TabFPN [190]) ou des CNNs fournissant une prédiction directement à partir de l'IRM (en s'affranchissant de la segmentation des noyaux gris centraux). Ces comparaisons devront être menées prochainement. Notre proposition, bien qu'offrant de bons résultats, reste selon nous améliorable car encore très élémentaire (seulement 8 structures reliées deux à deux et 3 types d'informations structurelles). L'information structurelle modélisée par les graphes pourrait être enrichie avec d'autres types de relations et d'autres régions ou sous-régions cérébrales profondes liées aux fonctions motrices. Ces informations liées aux régions (c.a.d. aux noeuds) et aux relations entre régions (c.a.d. aux arêtes) pourraient être apportées par d'autres modalités d'imagerie telle que l'imagerie fonctionnelle

(ex. imagerie du tenseur de diffusion ou DTI, IRM fonctionnelle ou IRMf), complémentaires à l'IRM structurelle, comme proposé par [177, 178]. A l'aide d'informations plus complètes, nous espérons améliorer les performances de notre modèle et, peut-être, réussir à estimer directement le score moteur brut de chaque main de l'enfant au *Box and Blocks Test* malgré les facteurs externes entrant en jeu.

Durant la thèse, nous avons fait le choix de réaliser ces prédictions en exploitant les informations structurelles sous forme de graphes à l'aide des GNNs. Nous avons déjà au cours du manuscrit (chapitre 1) présenté d'autres alternatives pour exploiter ces informations sous forme de graphes (ex. QAP, GED). Nous pourrions également exploité des méthodes fonctionnant sur des mesures d'entropie [192] cherchant à quantifier le degré de désordre structurel. Cette mesure d'entropie pourrait s'appliquer sur les graphes introduits dans le chapitre 5 [193] ou directement sur l'image [194]. Elle présenterait l'avantage d'être entièrement explicable (pas de paramètres entraînaibles) et robuste au manque de données (pas de données nécessaires pour entraîner un modèle).

Les perspectives ouvertes par notre méthode prometteuse sont très nombreuses sur le plan de la recherche neuroscientifique.

Ainsi, on peut s'interroger si notre méthode, appliquée à des enfants de 7 ans dans nos travaux, pourrait être étendue à des stades de développement plus précoces (pour prédire le plus tôt possible l'apparition d'une paralysie cérébrale) ou au contraire plus avancés afin de mettre en oeuvre une étude longitudinale sur des enfants ayant eu un AVC néonatal. L'application de notre méthode sur des IRMs acquises avant et après rééducation de l'enfant avec lésion cérébrale précoce pourrait permettre de visualiser une possible modification de l'organisation structurelle des noyaux gris centraux liée à la rééducation et surtout d'aider à évaluer l'efficacité des thérapies de rééducation mises en oeuvre. Il est à noter que cette évaluation de la rééducation par notre méthode pourrait être faite conjointement à d'autres outils d'analyse du cerveau telle qu'une analyse morphométrique (ex. VBM, SBM, DBM).

Par ailleurs, il se pose la question de l'efficacité de notre méthode sur des enfants atteints d'autres types de lésions cérébrales précoces (ex. lésion de la substance blanche) unilatérales ou bilatérales. Dans la prolongation de cette perspective, on peut même s'interroger sur la pertinence de notre méthode sur des sujets plus âgés ayant fait un AVC ou présentant une lésion cérébrale d'origine différente. Notre méthode pourrait être testée pour observer s'il existe aussi, dans ce cas distinct de la lésion cérébrale précoce, une corrélation entre organisation macrostructurelle des noyaux gris centraux et motricité.

Les travaux présentés dans cette thèse sont les prémises d'un long travail de recherche en analyse d'images de manière générale. Sur le volet théorique, de nombreuses pistes restent à explorer, notamment compte tenu de l'essor récent des GNNs. Sur le plan médical, nous espérons par ces travaux en neuroimagerie faire avancer la recherche relative à la lésion cérébrale précoce que ce soit chez l'enfant ou sur le modèle animal.

ANNEXES

Dans ces annexes, nous regroupons les Tables et Figures complémentaires aux résultats présentés dans les chapitres 2, 3 et 5.

A.1 Annexes du chapitre 2

L'annexe ci-dessous complète les études du chapitre 2 relatives à l'influence de la topologie des graphes dans les résultats de segmentation de la méthode proposée, basée sur GNN.

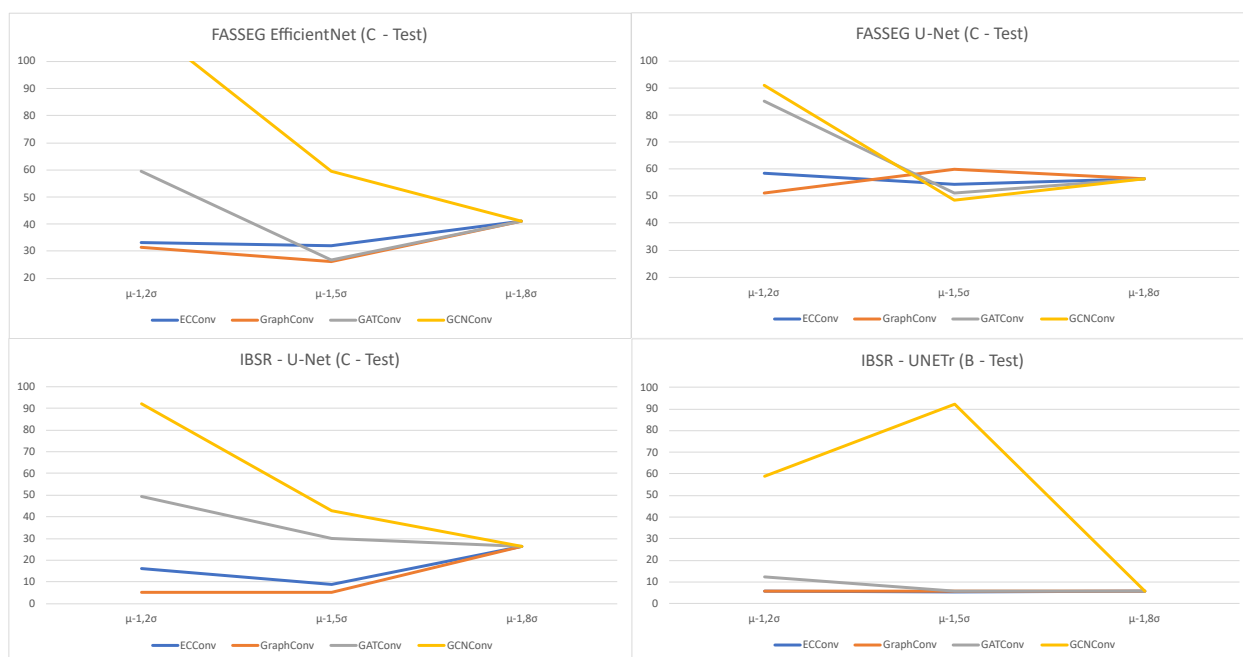


FIGURE A.1.1 – Évolution de la distance d'Hausdorff des segmentations en fonction du rayon d'élagage considéré (comparaison sur les données de test). Deux exemples sur FASSEG (EfficientNet-C et U-Net-C) et deux sur IBSR (U-Net-C et UNETr-B) sont illustrés.

A.2 Annexes du chapitre 3

Les annexes ci-dessous complètent les résultats du chapitre 3 dédié à la segmentation du cerveau de porcelet. Elles illustrent notamment l'impact de la stratégie d'extraction de patches employée, l'influence de la stratégie d'entraînement du modèle U-Net (avec ou sans apprentissage par transfert) ainsi que la comparaison des performances de la méthode proposée sur les porcelets de 2 semaines avec celles de l'outil BET.

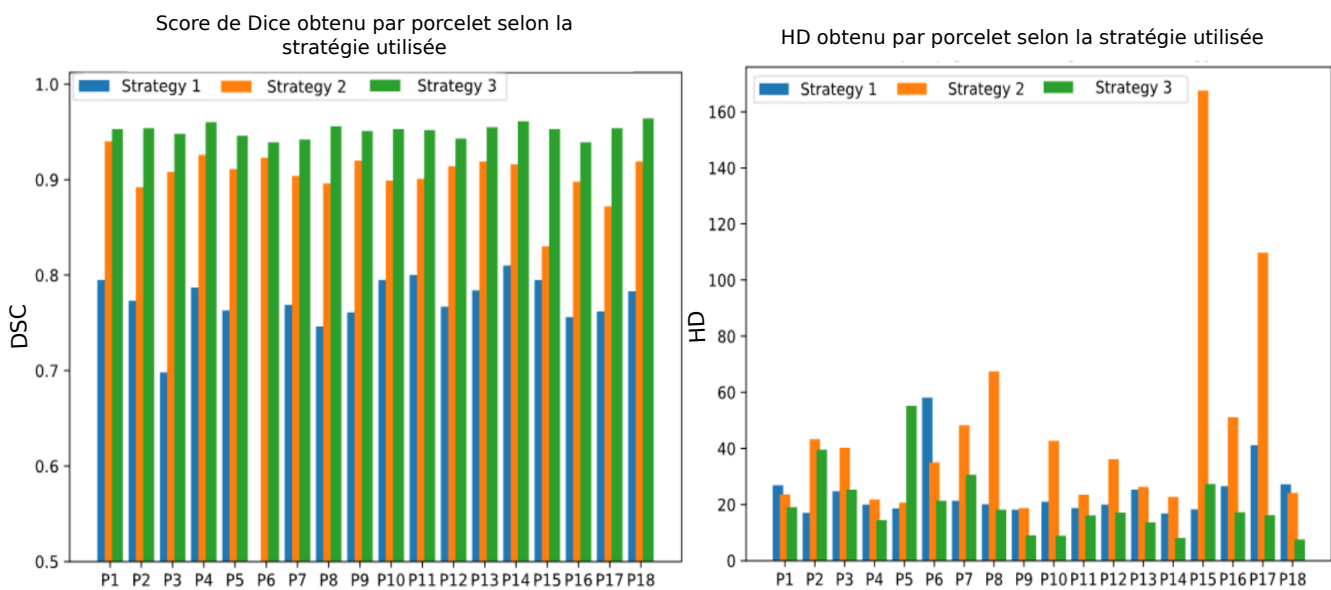


FIGURE A.2.1 – Résultats détaillés des performances (DSC et HD) obtenues sur les 18 porcelets de test âgés de 2 semaines (P1 à P18) avec les trois stratégies de distribution des patches d'entraînement.

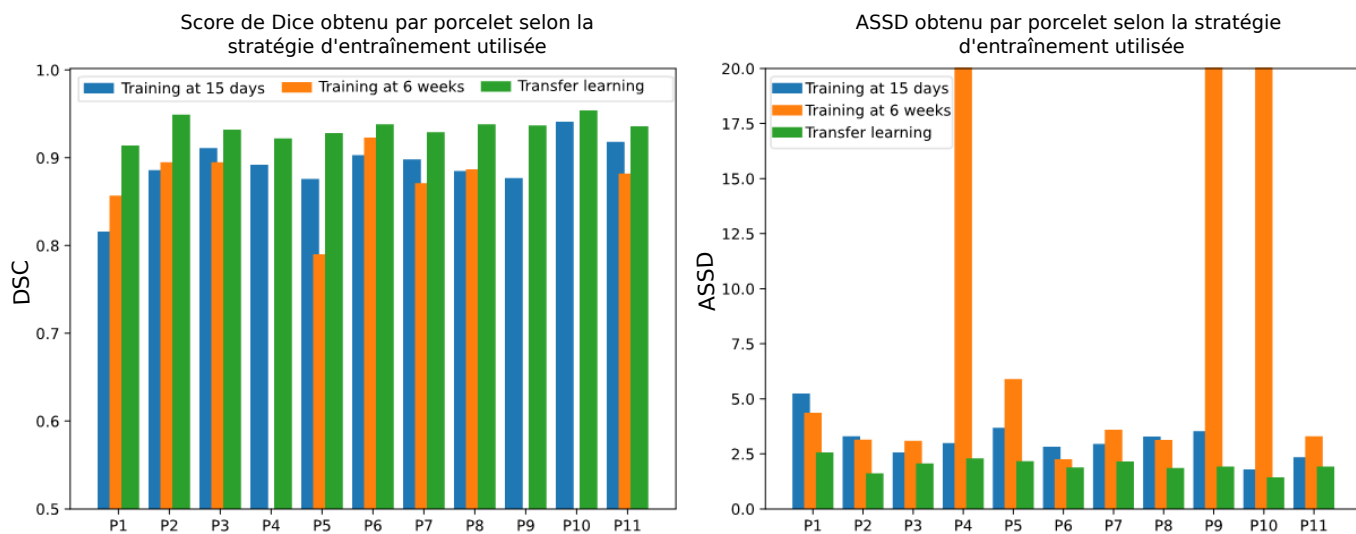


FIGURE A.2.2 – Résultats détaillés des performances (DSC et ASSD) obtenues sur les 11 porcelets de test à 6 semaines avec les 3 stratégies d'entraînement du réseau U-Net.

TABLE A.2.3 – Résultats obtenus avec BET et notre méthode sur les 18 porcelets de test (les âges exacts sont donnés à titre indicatif).

	Âge	Volume cérébral	Volume segmenté		DSC		HD		ASSD	
P1	16 J	$53.45cm^3$	$45.20cm^3$	$55.03cm^3$	0.890	0.959	13.64	7.28	2.91	1.16
P2	13 J	$48.81cm^3$	$46.04cm^3$	$49.57cm^3$	0.870	0.955	14.18	9.43	3.44	1.25
P3	13 J	$46.05cm^3$	$45.83cm^3$	$44.76cm^3$	0.680	0.956	29.07	8.48	8.52	1.21
P4	15 J	$52.40cm^3$	$38.76cm^3$	$54.70cm^3$	0.409	0.959	49.16	5.38	16.49	1.19
P5	15 J	$50.82cm^3$	$32.94cm^3$	$51.99cm^3$	0.675	0.948	37.01	8.60	8.33	1.45
P6	16 J	$54.94cm^3$	$48.54cm^3$	$52.89cm^3$	0.892	0.956	11.04	6.40	2.90	1.25
P7	15 J	$48.72cm^3$	$43.14cm^3$	$49.61cm^3$	0.861	0.946	16.03	8.54	3.55	1.48
P8	16 J	$53.82cm^3$	$26.51cm^3$	$52.45cm^3$	0.639	0.944	48.88	9.43	9.72	1.59
P9	16 J	$52.34cm^3$	$42.76cm^3$	$52.40cm^3$	0.846	0.952	15.26	6.71	4.02	1.33
P10	14 J	$50.68cm^3$	$31.07cm^3$	$49.67cm^3$	0.713	0.949	36.50	10.30	7.14	1.38
P11	14 J	$51.16cm^3$	$42.42cm^3$	$50.04cm^3$	0.868	0.955	15.56	8.06	3.38	1.23
P12	14 J	$50.11cm^3$	$42.87cm^3$	$50.88cm^3$	0.876	0.950	13.00	8.31	3.11	1.33
P13	15 J	$55.67cm^3$	$47.64cm^3$	$53.72cm^3$	0.883	0.936	13.96	14.32	3.09	1.78
P14	15 J	$53.98cm^3$	$33.94cm^3$	$54.66cm^3$	0.632	0.950	44.15	9.85	9.57	1.45
P15	16 J	$53.67cm^3$	$42.78cm^3$	$54.70cm^3$	0.842	0.960	24.06	5.92	4.10	1.15
P16	15 J	$54.23cm^3$	$39.68cm^3$	$51.83cm^3$	0.811	0.945	19.44	11.57	4.86	1.56
P17	15 J	$59.64cm^3$	$30.22cm^3$	$57.07cm^3$	0.581	0.958	39.32	8.66	10.99	1.22
P18	15 J	$52.38cm^3$	$37.11cm^3$	$53.09cm^3$	0.785	0.963	26.09	6.00	5.49	1.06

A.3 Annexes du chapitre 5

Les annexes ci-dessous complètent les résultats du chapitre 5 relatifs à l'étude de la motricité chez les enfants après AVC néonatal. Elles illustrent notamment l'impact des informations structurelles et de la topologie des graphes considérés en entrée de GNN pour le problème de régression (estimation de score moteur au *Box and Blocks Test*).

	L1 moyenne	L1 médiane	L1 max	R (p-value)
Pas d'attribut	0.052	0.044	0.205	-0.056 (0.75)
Distances	0.052	0.045	0.179	0.290 (0.09)
Allongements	0.051	0.045	0.138	0.695 ($3.55e^{-6}$)
Volumes	0.069	0.057	0.283	-0.185 (0.29)
Volumes+Distances	0.059	0.051	0.159	0.425 (0.01)
Allongements+Distances	0.055	0.047	0.136	0.762 ($1.06e^{-7}$)
Volumes+Allongements	0.058	0.061	0.150	0.734 ($5.22e^{-7}$)
Tous les attributs	0.054	0.057	0.162	0.695 ($3.58e^{-6}$)

TABLE A.3.1 – Performance de prédiction du rapport de score y_1 associé à la main affectée en fonction des informations structurelles considérées dans le cas de graphes complets. L1 : distance de Manhattan, R : coefficient de corrélation de Pearson.

	L1 moyenne	L1 médiane	L1 max	R (p-value)
Pas d'attribut	0.048	0.035	0.202	0.240 (0.16)
Distances	0.043	0.032	0.110	0.565 ($0.40e^{-3}$)
Allongements	0.055	0.059	0.123	0.780 ($3.24e^{-8}$)
Volumes	0.061	0.059	0.298	-0.143 (0.41)
Volumes+Distances	0.055	0.047	0.366	-0.141 (0.42)
Allongements+Distances	0.053	0.062	0.133	0.714 ($1.44e^{-6}$)
Volumes+Allongements	0.057	0.062	0.124	0.518 ($1.43e^{-3}$)
Tous les attributs	0.063	0.065	0.188	0.054 (0.76)

TABLE A.3.2 – Performance de prédiction du rapport de score y_2 associé à la main "non-affectée" en fonction des informations structurelles considérées dans le cas de graphes complets. L1 : distance de Manhattan, R : coefficient de corrélation de Pearson.

BIBLIOGRAPHIE

- [1] Jean-Emmanuel Bibault. *2041 : l'odyssée de la médecine - Comment l'intelligence artificielle bouleverse la médecine*. Equateurs, 2023.
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [3] Yan Song, Feiyang Xu, Qi Yao, Jialin Liu, and Shuai Yang. Navigation algorithm based on semantic segmentation in wheat fields using an RGB-D camera. *Information Processing in Agriculture*, 2022.
- [4] Zofie Cimbuřova and Stefan Blumentrath. Viewshed-based modelling of visual exposure to urban greenery—An efficient GIS tool for practical planning applications. *Landscape and Urban Planning*, 222 :104395, 2022.
- [5] Mayank Tiwari, Subir Singh Lamba, and Bhupendra Gupta. An image processing and computer vision framework for efficient robotic sketching. *Procedia Computer Science*, 133 :284–289, 2018.
- [6] Viljami Sairanen, Mario Ocampo-Pineda, Cristina Granziera, Simona Schiavi, and Alessandro Dadducci. Incorporating outlier information into diffusion-weighted MRI modeling for robust microstructural imaging and structural brain connectivity analyses. *NeuroImage*, 247 :118802, 2022.
- [7] Shi Dong, Ping Wang, and Khushnood Abbas. A survey on deep learning and its applications. *Computer Science Review*, 40 :100379, 2021.
- [8] Jean-Baptiste Fasquel and Nicolas Delanoue. A graph based image interpretation method using a priori qualitative inclusion and photometric relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41 :1043–1055, 2019.
- [9] Davide Bacciu, Federico Errica, Alessio Micheli, and Marco Podda. A gentle introduction to deep learning for graphs. *Neural Networks*, 129 :203–221, 2020.
- [10] Jérémy Chopin. *Apprentissage profond et connaissances structurelles pour l'analyse d'images*. PhD thesis, Université d'Angers, 2022.
- [11] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, Pablo Martinez-Gonzalez, and Jose Garcia-Rodriguez. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, 70 :41–65, 2018.
- [12] I. E. Poletaev, Konstantin S. Pervunin, and Mikhail P. Tokarev. Artificial neural network for bubbles pattern recognition on the images. *Journal of Physics : Conference Series*, 754, 2016.
- [13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net : convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351, pages 234–241. Springer, 2015.
- [14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, and Georg Heigold et al. An image is worth

-
- 16x16 words : Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.
- [15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 :640–651, 2017.
- [16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.
- [17] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, page 9650–9660, 2021.
- [18] Tristan Gomez. *Attention spatiale artificielle pour des modèles profonds interprétables de qualité embryonnaire*. PhD thesis, Université de Nantes, 2022.
- [19] Tete Xiao, Mannat Singh, Eric Mintun, Trevor Darrell, Piotr Dollár, and Ross Girshick. Early convolutions help transformers see better. *Advances in Neural Information Processing Systems*, 34 :30392–30400, 2021.
- [20] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. CvT : Introducing Convolutions to Vision Transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 22–31, 2021.
- [21] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger Roth, and Daguang Xu. UNETR : Transformers for 3D medical image segmentation. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1748–1758, 2022.
- [22] Xiang Li, Ying Wei, Lin Wang, Shidi Fu, and Chuyuan Wang. MSGSE-Net : Multi-scale guided squeeze-and-excitation network for subcortical brain structure segmentation. *Neurocomputing*, 461 :228–243, 2021.
- [23] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 2019.
- [24] Karl Weiss, Taghi M. Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3 :1345–1359, 2016.
- [25] Jian Ding, Nan Xue, Gui-Song Xia, and Dengxin Dai. Decoupling zero-shot semantic segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, page 11583–11592, 2022.
- [26] Krishna Chaitanya, Neerav Karani, Christian F. Baumgartner, Ertunc Erdil, Anton Becker, Olivio Donati, and Ender Konukoglu. Semi-supervised task-driven data augmentation for medical image segmentation. *Medical Image Analysis*, 68 :101934, 2021.
- [27] Mickael Komendyak. How to choose the best pre-trained model for your Convolutional Neural Network? <https://data-science-blog.com/blog/2022/04/11/>

-
- how-to-choose-the-best-pre-trained-model-for-your-convolutional-neural-network/, 2022.
- [28] Yuta Tokuoka, Shuji Suzuki, and Yohei Sugawara. An inductive transfer learning approach using cycle-consistent adversarial domain adaptation with application to brain tumor segmentation. In *Proceedings of the 2019 6th international conference on biomedical and bioinformatics engineering*, pages 44–48, 2019.
- [29] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples : A survey on few-shot learning. *ACM computing surveys*, 53(3) :1–34, 2020.
- [30] Vincent Dumoulin, Neil Houlsby, Utku Evci, Xiaohua Zhai, Ross Goroshin, Sylvain Gelly, and Hugo Larochelle. Comparing transfer and meta learning approaches on a unified few-shot classification benchmark. *arXiv*, 2021.
- [31] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet : A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [32] Yongfei Liu, Xiangyi Zhang, Songyang Zhang, and Xuming He. Part-Aware Prototype Network for Few-Shot Semantic Segmentation. *Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science()*, 12354, 2020.
- [33] Adrien Delaye and Eric Anquetil. Fuzzy relative positioning templates for symbol recognition. In *International Conference on Document Analysis and Recognition*, pages 1220–1224, 2011.
- [34] Lars Kunze, Chris Burbridge, Marina Alberti, Akshaya Thippur, John Folkesson, Patric Jensfelt, and Nick Hawes. Combining top-down spatial reasoning and bottom-up object class recognition for scene understanding. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2910–2915, 2014.
- [35] Mateus Riva, Pietro Gori, Florian Yger, and Isabelle Bloch. Is the U-Net directional-relationship aware? In *29th IEEE International Conference on Image Processing (ICIP)*, page 3391–3395, 2022.
- [36] Isabelle Bloch. Fuzzy sets for image processing and understanding. *Fuzzy Sets and Systems*, 281 :280–291, 2015.
- [37] Jean-Baptiste Fasquel and Nicolas Delanoue. Approach for sequential image interpretation using a priori binary perceptual topological and photometric knowledge and k-means-based segmentation. *Journal of the Optical Society of America (JOSA)*, 35 :936–945, 2018.
- [38] RECITAL. <https://recital.univ-nantes.fr/>, 2018.
- [39] ISLES Challenge 2022 : Ischemic Stroke Lesion Segmentation. <https://www.isles-challenge.org/>, 2022.
- [40] CROHME : Competition on Recognition of Online Handwritten Mathematical Expressions. <https://www.isical.ac.in/~crohme/index.html>, 2014.
- [41] Chao-Jung Liu, Vladimir A. Krylov, Paul Kane, Geraldine Kavanagh, and Rozenn Dahyot. IM2ELEVATION : Building Height Estimation from Single-View Aerial Imagery. *Remote sensing*, 12, 2020.

-
- [42] Jean-Baptiste Fasquel, Vincent Agnus, Johan Moreau, Luc Soler, and Jacques Marescaux. An interactive medical image segmentation system based on the optimal management of regions of interest using topological medical knowledge. *Computer Methods and Programs in Biomedicine*, 82 :216–230, 2006.
- [43] Antonio Moreno, Celina M. Takemura, Olivier Colliot, Oscar Camara, and Isabelle Bloch. Using anatomical knowledge expressed as fuzzy constraints to segment the heart in CT images. *Pattern Recognition*, 41 :2525–2540, 2008.
- [44] Aline Deruyver, Yann Hodé, and Luc Brun. Image interpretation with a conceptual graph : Labeling over-segmented images and detection of unexpected objects. *Artificial Intelligence*, 173 :1245–1265, 2009.
- [45] William L. Hamilton. *Graph Representation Learning*, volume 14. Morgan and Claypool, 2020. https://www.cs.mcgill.ca/~wlh/grl_book/.
- [46] Marko A. Rodriguez and Peter Neubauer. A path algebra for multi-relational graphs. In *2011 IEEE 27th International Conference on Data Engineering Workshops*, pages 128–131, 2011.
- [47] Kamel Madi, Hamida Seba, Hamamache Kheddouci, and Olivier Barge. A Graph-based approach for Kite recognition. *Pattern Recognition Letters*, 87 :186–194, 2017.
- [48] Jérémy Chopin, Jean-Baptiste Fasquel, Harold Mouchère, Rozenn Dahyot, and Isabelle Bloch. Improving semantic segmentation with graph-based structural knowledge. In *Pattern Recognition and Artificial Intelligence : Third International Conference, ICPRAI 2022, Paris, France, June 1-3, 2022, Proceedings, Part I*, pages 173–184, 2022.
- [49] Richard Jiang, Anthony T.S. Ho, Ismahane Cheheb, Noor Al-Maadeed, Somaya Al-Maadeed, and Ahmed Bouridane. Emotion recognition from scrambled facial images via many graph embedding. *Pattern Recognition*, 67 :245–251, 2017.
- [50] Frank Julca-Aguilar, Harold Mouchère, Christian Viard-Gaudin, and Nina S. T. Hirata. A general framework for the recognition of online handwritten graphics. *International Journal on Document Analysis and Recognition (IJDAR)*, 23 :143–160, 2020.
- [51] Guixiang Ma, Nesreen K. Ahmed, Theodore L. Willke, and Philip S. Yu. Deep graph similarity learning : A survey. *Data Mining and Knowledge Discovery*, 35 :688–725, 2021.
- [52] Francesc Serratosa and Xavier Cortés. Graph Edit Distance : Moving from global to local structure to solve the graph-matching problem. *Pattern Recognition Letters*, 65 :204–210, 2015.
- [53] Pasquale Foggia, Gennaro Percannella, and Mario Vento. Graph matching and learning in pattern recognition in the last 10 years. *International Journal of Pattern Recognition and Artificial Intelligence*, 28(01) :1450001, 2014.
- [54] Joao Maciel and Joao P. Costeira. A global solution to sparse correspondence problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 :187–199, 2003.
- [55] Andrei Zanfir and Cristian Sminchisescu. Deep learning of graph matching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2684–2693, 2018.
- [56] Kaspar Riesen. Structural pattern recognition with graph edit distance. *Advances in computer vision and pattern recognition*, 2015.

-
- [57] Pep Santacruz and Francesc Serratos. Error-tolerant graph matching in linear computational cost using an initial small partial matching. *Pattern Recognition Letters*, 134 :10–19, 2020.
- [58] Jérémy Chopin, Jean-Baptiste Fasquel, Harold Mouchère, Rozenn Dahyot, and Isabelle Bloch. QAP Optimisation with Reinforcement Learning for Faster Graph Matching in Sequential Semantic Image Analysis. In *Pattern Recognition and Artificial Intelligence : Third International Conference, ICPRAI 2022, Paris, France, June 1-3, 2022, Proceedings, Part I*, pages 47–58, 2022.
- [59] Ziwei Zhang, Peng Cui, and Wenwu Zhu. Deep learning on graphs : A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(1) :249–270, 2020.
- [60] Yujia Li, Chenjie Gu, Thomas Dullien, Oriol Vinyals, and Pushmeet Kohli. Graph matching networks for learning the similarity of graph structured objects. In *International conference on machine learning*, pages 3835–3845, 2019.
- [61] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272, 2017.
- [62] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30 :1025–1035, 2017.
- [63] Thomas Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv*, 2016.
- [64] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13) :457–466, 2018.
- [65] Monica Agrawal, Marinka Zitnik, and Jure Leskovec. Large-scale analysis of disease pathways in the human interactome. In *PACIFIC SYMPOSIUM on BIOCOMPUTING 2018 : Proceedings of the Pacific Symposium*, pages 111–122, 2018.
- [66] Aditya Grover, Aaron Zweig, and Stefano Ermon. Graphite : Iterative generative modeling of graphs. In *International conference on machine learning*, pages 2434–2444, 2019.
- [67] Matthias Fey. Workshop on Geometry and Machine Learning with Applications to Biomedical, University College London, 2020.
- [68] Martin Simonovsky and Nikos Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3693–3702, 2017.
- [69] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1) :4–24, 2020.
- [70] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman Go Neural : Higher-order Graph Neural Networks. In *Conference on Artificial Intelligence (AAAI)*, pages 4602–4609, 2019.
- [71] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. In *International Conference on Learning Representations (ICLR)*, 2018.

-
- [72] Cornelia Caragea, Jian Wu, Alina Ciobanu, Kyle Williams, Juan Fernández-Ramírez, Hung-Hsuan Chen, Zhaohui Wu, and Lee Giles. CiteSeerx : A Scholarly Big Dataset. *Advances in Information Retrieval. ECIR 2014. Lecture Notes in Computer Science*, 8416, 2014.
- [73] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3) :93–93, 2008.
- [74] Abdullah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-STGCNN : A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14412–14420, 2020.
- [75] Alain Trémeau and Philippe Colantoni. Regions adjacency graph applied to color image segmentation. *IEEE Transactions on image processing*, 9(4) :735–744, 2000.
- [76] Guillaume Renton, Muhammet Balçilar, Pierre Héroux, Benoit Gaüzère, Paul Honeine, and Sébastien Adam. Symbols Detection and Classification using Graph Neural Networks. *Pattern Recognition Letters*, 152 :391–397, 2021.
- [77] Han Yang, Xiao Yan, Xinyan Dai, Yongqiang Chen, and James Cheng. Self-enhanced GNN : Improving graph neural networks using model outputs. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021.
- [78] Jiayu Li, Tianyun Zhang, Hao Tian, Shengmin Jin, Makan Fardad, and Reza Zafarani. SGCN : A graph sparsifier based on graph convolutional networks. In *Advances in Knowledge Discovery and Data Mining. PAKDD 2020. Lecture Notes in Computer Science()*, volume 12084, pages 275–287, 2020.
- [79] Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. Dropedge : Towards deep graph convolutional networks on node classification. *arXiv*, 2019.
- [80] Frederik Diehl. Edge contraction pooling for graph neural networks. *arXiv*, 2019.
- [81] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11976–11986, 2022.
- [82] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34 :2274–2282, 2012.
- [83] Song Ouyang and Yansheng Li. Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery. *Remote Sensing*, 13, 2021.
- [84] Qi Diao, Yaping Dai, Ce Zhang, Yan Wu, Xiaoxue Feng, and Feng Pan. Superpixel-based attention graph neural network for semantic segmentation in aerial images. *Remote Sensing*, 14 :305, 2022.
- [85] Jun Zhang, Zhiyuan Hua, Kezhou Yan, Kuan Tian, Jianhua Yao, Eryun Liu, Mingxia Liu, and Xiao Han. Joint fully convolutional and graph convolutional networks for weakly-supervised segmentation of pathology images. *Medical Image Analysis*, 73, 2021.

-
- [86] Rongtao Xu, Ye Li, Changwei Wang, Shibiao Xu, Weiliang Meng, and Xiaopeng Zhang. Instance segmentation of biological images using graph convolutional network. *Engineering Applications of Artificial Intelligence*, 110 :104739, 2022.
- [87] Xia Li, Yibo Yang, Qijie Zhao, Tiancheng Shen, Zhouchen Lin, and Hong Liu. Spatial Pyramid Based Graph Reasoning for Semantic Segmentation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8947–8956, 2020.
- [88] Yunpeng Chen, Marcus Rohrbach, Zhicheng Yan, Shuicheng Yan, Jiashi Feng, and Yannis Kalantidis. Graph-Based Global Reasoning Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 433–442, 2019.
- [89] Qinghui Liu, Michael Kampffmeyer, Robert Jenssen, and Arnt-Børre Salberg. SCG-Net : Self-Constructing Graph Neural Networks for Semantic Segmentation, 2020.
- [90] Seung Yeon Shin, Soochahn Lee, Il Dong Yun, and Kyoung Mu Lee. Deep vessel segmentation by learning graphical connectivity. *Medical Image Analysis*, 58 :101556, 2019.
- [91] Kun Wang, Xiaohong Zhang, Yuting Lu, Xiangbo Zhang, and Wei Zhang. CGRNet : Contour-guided graph reasoning network for ambiguous biomedical image segmentation. *Biomedical Signal Processing and Control*, 75 :103621, 2022.
- [92] Ville Hautamaki, Ismo Karkkainen, and Pasi Franti. Outlier detection using k-Nearest Neighbour graph. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 3, pages 430–433, 2004.
- [93] Shuiwang Ji Hongyang Gao. Graph U-Nets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 :4948–4960, 2022.
- [94] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv*, 2019.
- [95] Alex P. Zijdenbos, Benoit M. Dawant, Ran A. Margolin, and Andrew C. Palmer. Morphometric analysis of white matter lesions in MR images : method and validation. *IEEE Transactions on Medical Imaging*, 13 :716–724, 1994.
- [96] Mario Beauchemin, Keith P. B. Thomson, and Geoffrey Edwards. On the Hausdorff distance used for the evaluation of segmentation results. *Canadian Journal of Remote Sensing*, 24 :3–8, 1998.
- [97] Khalil Khan, Massimo Mauro, Pierangelo Migliorati, and Riccardo Leonardi. Head pose estimation through multi-class face segmentation. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pages 175–180, 2017.
- [98] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip H. S. Torr. Conditional random fields as recurrent neural networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1529–1537, 2015.
- [99] Mingxing Tan and Quoc V. Le. EfficientNet : Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning (ICML)*, pages 6105–6114, 2019.
- [100] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, 2017.

-
- [101] Kaisar Kushibar, Sergi Valverde, Sandra Gonzalez-Villa, Jose Bernal, Mariano Cabezas, Arnau Oliver, and Xavier Llado. Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. *Medical Image Analysis*, 48 :177–186, 2018.
- [102] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net : Learning dense volumetric segmentation from sparse annotation. *Lecture Notes in Computer Science*, pages 424–432, 2016.
- [103] Bumshik Lee, Nagaraj Yamanakkanavar, Muhammad Ammar Malik, and Jae Young Choi. Automatic segmentation of brain MRI using a novel patch-wise U-Net deep architecture. *PLOS ONE*, 15 :1–20, 2020.
- [104] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. Torchio : a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208 :106236, 2021.
- [105] Abdullah A. Abdullah, Masoud M. Hassan, and Yaseen T. Mustafa. A Review on Bayesian Deep Learning in Healthcare : Applications and Challenges. *IEEE Access*, 10 :36538–36562, 2022.
- [106] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) challenge. *International journal of computer vision*, 88 :303–308, 2009.
- [107] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.
- [108] Patty Coupeau, Jean-Baptiste Fasquel, and Mickael Dinomais. On the relevance of edge-conditioned convolution for GNN-based semantic image segmentation using spatial relationships. In *2022 Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, 2022.
- [109] Adam Kirton. Modeling developmental plasticity after perinatal stroke : defining central therapeutic targets in cerebral palsy. *Pediatric neurology*, 48(2) :81–94, 2013.
- [110] Tonse N. K. Raju, Karin B. Nelson, Donna Ferriero, John Kylan Lynch, and NICHD-NINDS Perinatal Stroke Workshop Participants. Ischemic perinatal stroke : summary of a workshop sponsored by the National Institute of Child Health and Human Development and the National Institute of Neurological Disorders and Stroke. *Pediatrics*, 120(3) :609–616, 2007.
- [111] Adam Kirton and Gabrielle deVeber. Paediatric stroke : pressing issues and promising directions. *Lancet Neurology*, 14 :92–102, 2015.
- [112] Stéphane Chabrier, Margaux Pouyfaucou, Alain Chatelin, Yannick Bleyenheuft, Joel Fluss, Vincent Gautheron, Christopher J. Newman, Guillaume Sébire, Patrick Van Bogaert, and Carole Vuillerot et al. From congenial paralysis to post-early brain injury developmental condition : Where does cerebral palsy actually stand? *Annals of physical and rehabilitation medicine*, 63(5) :431–438, 2020.
- [113] Janet Lee, Lisa A. Croen, Camilla Lindan, Kendall B. Nash, Cathleen K. Yoshida, Donna M. Ferriero, A. J. Barkovich, and Yvonne W. Wu. Predictors of outcome in perinatal arterial stroke : a population-based study. *Annals of Neurology*, 58 :303–308, 2005.

-
- [114] Mickaël Dinomais, Lucie Hertz-Pannier, Samuel Groeschel, Stéphane Chabrier, Matthieu Delion, Béatrice Husson, Manoelle Kossorotoff, Cyrille Renaud, and Sylvie Nguyen The Tich. Long term motor function after neonatal stroke : lesion localization above all. *Human Brain Mapping*, 36 :4793–4807, 2015.
- [115] Marie Arsalidou, Emma G. Duerden, and Margot J. Taylor. The centre of the brain : topographical model of motor, cognitive, affective, and somatosensory functions of the basal ganglia. *Human Brain Mapping*, 34(11) :3031–3054, 2013.
- [116] Antonios Makropoulos, Paul Aljabar, Robert Wright, Britta Hüning, Nazakat Merchant, Tomoki Arichi, Nora Tusor, Joseph V. Hajnal, A. David Edwards, Serena J. Counsell, and Daniel Rueckert. Regional growth and atlasing of the developing human brain. *NeuroImage*, 125 :456–478, 2016.
- [117] Brandon T. Craig, Helen L. Carlson, and Adam Kirton. Thalamic diaschisis following perinatal stroke is associated with clinical disability. *Neuroimage : Clinical*, 21 :101660, 2019.
- [118] Jordan Hassett, Helen Carlson, Ali Babwani, and Adam Kirton. Bihemispheric developmental alterations in basal ganglia volumes following unilateral perinatal stroke. *NeuroImage : Clinical*, 35 :103143, 2022.
- [119] Nigul Ilves, Silva Lõo, Norman Ilves, Rael Laugesaar, Dagmar Loorits, Pille Kool, Tiina Talvik, and Pilvi Ilves. Ipsilesional volume loss of basal ganglia and thalamus is associated with poor hand function after ischemic perinatal stroke. *BMC Neurology*, 22 :23, 2022.
- [120] Béatrice Husson, Lucie Hertz-Pannier, Cyrille Renaud, Dominique Allard, Emilie Presles, Pierre Landrieu, and Stéphane Chabrier. Motor outcomes after neonatal arterial ischemic stroke related to early MRI data in a prospective study. *Pediatrics*, 126 :912–918, 2010.
- [121] Bruce Fischl. FreeSurfer. *NeuroImage*, 62 :774–781, 2012.
- [122] José V. Manjón, José E. Romero, Roberto Vivo-Hernando, Gregorio Rubio, Fernando Aparici, Mariam de la Iglesia-Vaya, and Pierrick Coupé. vol2Brain : A New Online Pipeline for Whole Brain MRI Analysis. *Frontiers in neuroinformatics*, 16 :862805, 2022.
- [123] Raghav Mehta, Aabhas Majumdar, and Jayanthi Sivaswamy. BrainSegNet : a convolutional neural network architecture for automated segmentation of human brain structures. *Journal of Medical Imaging*, 4 :024003, 2017.
- [124] Lihao Liu, Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Ψ -Net : Stacking Densely Convolutional LSTMs for Sub-Cortical Brain Structure Segmentation. *IEEE Transactions on Medical Imaging*, 39 :2806–2817, 2020.
- [125] Jacques-Olivier Coq, Marine Kochmann, Diego C. Lacerda, Hanane Khalki, Maxime Delcour, Ana E. Toscano, Florence Cayetanot, Marie-Hélène Canu, Mary F. Barbe, and Masahiro Tsuj. From cerebral palsy to developmental coordination disorder : Development of preclinical rat models corresponding to recent epidemiological changes. *Annals of Physical and Rehabilitation Medicine*, 63 :422–430, 2020.
- [126] Clarissa F. Cavarsan, Monica A. Gorassini, and Katharina A. Quinlan. Animal models of developmental motor disorders : parallels to human motor dysfunction in cerebral palsy. *Journal of Neurophysiology*, 122 :1238–1253, 2019.

-
- [127] Yaël Balbastre, Denis Rivière, Nicolas Souedet, Clara Fischer, Anne-Sophie Hérard, Susannah Williams, Michel E. Vandenberghe, Julien Flament, Romina Aron-Badin, Philippe Hantraye, Jean-François Mangin, and Thierry Delzescaux. A validation dataset for Macaque brain MRI segmentation. *Data in Brief*, 16 :37–42, 2018.
- [128] Arsène Ella and Matthieu Keller. Construction of an MRI 3D high resolution sheep brain template. *Magnetic Resonance Imaging*, 33 :1329–1337, 2015.
- [129] Elizabeth B. Hutchinson, Susan C. Schwerin, Kryslaine Radomski, Neda Sadeghi, Jeffrey Jenkins, Michal E. Komlosh, M. Okan Irfanoglu, Sharon L. Juliano, and Carlo Pierpaoli. Population based MRI and DTI templates of the adult ferret brain and tools for voxelwise analysis. *NeuroImage*, 152 :575–589, 2017.
- [130] Xueru Liu, Rui Tian, Zhentao Zuo, Hui Zhao, Liang Wu, Yan Zhuo, Yong Q. Zhang, and Lin Chen. A high-resolution MRI brain template for adult Beagle. *Magnetic Resonance Imaging*, 68 :148–157, 2020.
- [131] Hexia Gan, Quanzeng Zhang, Han Zhang, Yue Chen, Jianzhong Lin, TaiShan Kang, Jiaying Zhang, Frederic A. Troy 2nd, and Bing Wang. Development of new population-averaged standard templates for spatial normalization and segmentation of MR images for postnatal piglet brains. *Magnetic Resonance Imaging*, 32 :1396–1402, 2014.
- [132] Kirk L. Thibault and Susan S. Margulies. Age-dependant material properties of the porcine cerebrum : effect on pediatric inertial head injury criteria. *Journal of Biomechanics*, 31 :1119–1126, 1998.
- [133] Chelli N. Devi, Anupama Chandrasekharan, V. K. Sundararaman, and Zachariah C. Alex. Automatic segmentation of infant brain MR images : With special reference to myelinated white matter. *Biocybernetics and Biomedical Engineering*, 37 :143–158, 2017.
- [134] Edouard Mazerand, Clémentine Gallet, Luc Le Fournier, Mickaël Dinomais, Claudia Montero-Menei, and Philippe Menei. Development of a porcine model of cerebral palsy : first step before a study on the efficacy of a stereotactic stem cell therapy. In *European Academy of Childhood Disability*, 2019.
- [135] Matthew S. Conrad, Bradley P. Sutton, Ryan N. Dilger, and Rodney W Johnson. An In Vivo Three-Dimensional Magnetic Resonance Imaging-Based Averaged Brain Collection of the Neonatal Piglet (*Sus scrofa*). *PLoS One*, 9 :1396–1402, 2014.
- [136] Stephen M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17 :143–155, 2002.
- [137] David W. Shattuck, Stephanie R. Sandor-Leahy, Kirt A. Schaper, David A. Rottenberg, and Richard M. Leahy. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13 :856–876, 2001.
- [138] Feng Shi, Li Wang, Yakang Dai, John H. Gilmore, Weili Lin, and Dinggang Shen. LABEL : pediatric brain extraction using learning-based meta-algorithm. *Neuroimage*, 62(3) :1975–1986, 2012.
- [139] Snehashis Roy, John A. Butman, Dzung L. Pham, and Alzheimers Disease Neuroimaging Initiative. Robust skull stripping using multiple MR image contrasts insensitive to pathology. *Neuroimage*, 146 :132–147, 2017.

-
- [140] Stéphane Chabrier, Elie Saliba, Sylvie Nguyen The Tich, Aude Charollais, Marie-Noëlle Varlet, Brigitte Tardy, Emilie Presles, Cyrille Renaud, Dominique Allard, Béatrice Husson, and Pierre Landrieu. Obstetrical and neonatal characteristics vary with birthweight in a cohort of 100 term newborns with symptomatic arterial ischemic stroke. *European Journal of Paediatric Neurology*, 14(3) :206–213, 2010.
- [141] Françoise Barré-Sinoussi and Xavier Montagutelli. Animal models are essential to biological research : issues and perspectives. *Future science OA*, 1(4), 2015.
- [142] Riccardo Feo and Federico Giove. Towards an efficient segmentation of small rodents brain : a short critical review. *Journal of neuroscience methods*, 323 :82–89, 2019.
- [143] Hideaki Watanabe, Flemming Andersen, Claus Z. Simonsen, Stephen M. Evans, Albert Gjedde, and Paul Cumming. MR-Based Statistical Atlas of the Göttingen Minipig Brain. *NeuroImage*, 14 :1089–1096, 2001.
- [144] Carly Norris, Jonathan Lisinski, Elizabeth McNeil, John W. VanMeter, Pamela VandeVord, and Stephen M. LaConte. MRI brain templates of the male Yucatan minipig. *NeuroImage*, 235, 2021.
- [145] Stéphan Saikali, Paul Meurice, Paul Sauleau, Pierre-Antoine Eliat, Pascale Bellaud, Gwenaelle Randuineau, Marc Vérin, and Charles-Henri Malbert. A three-dimensional digital segmented and deformable brain atlas of the domestic pig. *Journal of Neuroscience Methods*, 192 :102–109, 2010.
- [146] Paul A. Yushkevich, Yang Gao, and Guido Gerig. ITK-SNAP : An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3342–3345, 2016.
- [147] Ivo Wolf, Marcus Vetter, Ingmar Wegner, Thomas Böttger, Marco Nolden, Max Schöbinger, Mark Hastenteufel, Tobias Kunert, and Hans-Peter Meinzer. The medical imaging interaction toolkit. *Medical image analysis*, 9(6) :594–604, 2005.
- [148] Johannes Lohmeier, Takaaki Kaneko, Bernd Hamm, Marcus R. Makowski, and Hideyuki Okano. atlasBREX : Automated template-derived brain extraction in animal MRI. *Scientific Reports*, 9 :12219, 2019.
- [149] Michael Ebner, Guotai Wang, Wenqi Li, Michael Aertsen, Premal A. Patel, Rosalind Aughwane, Andrew Melbourne, Tom Doel, Steven Dymarkowski, Paolo De Coppi, Anna L. David, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI. *NeuroImage*, 206 :116324, 2020.
- [150] Seyed Sadegh Mohseni Salehi, Seyed Raein Hashemi, Clemente Velasco-Annis, Abdelhakim Ouallah, Judy A. Estroff, Deniz Erdogmus, Simon K. Warfield, and Ali Gholipour. Real-time automatic fetal brain extraction in fetal MRI by deep learning. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 720–724, 2018.
- [151] Nigel Chou, Jiarong Wu, Jordan Bai Bingren, Anqi Qiu, and Kai-Hsiang Chuang. Robust automatic rodent brain extraction using 3D pulse-coupled neural networks (PCNN). *IEEE Transactions on Image Processing*, pages 54–64, 2011.
- [152] Tao Zhong, Fenqiang Zhao, Yuchen Pei, Zhenyuan Ning, Lufan Liao, Zhengwang Wu, Yuyu Niu, Li Wang, and Dinggang Shen et al. DIKA-Nets : Domain-invariant knowledge-guided attention networks for brain skull stripping of early developing macaques. *NeuroImage*, 227, 2021.

-
- [153] Mostefa Ben Naceur, Mohamed Akil, Rachida Saouli, and Rostom Kachouri. Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. *Medical Image Analysis*, 63 :101692, 2020.
- [154] Xindi Wang, Xin-Hui Li, Jae Wook Cho, Brian E. Russ, Nanditha Rajamani, Alisa Omelchenko, Lei Ai, Annachiara Korchmaros, and Stephen Sawiak et al. U-Net model for brain extraction : trained on humans for transfer to non-human primates. *NeuroImage*, 235 :118001, 2021.
- [155] Li-Ming Hsu, Shuai Wang, Paridhi Ranadive, Woomi Ban, Tzu-Hao Harry Chao, Sheng Song, Domenic Hayden Cerri, and Lindsay R. Walton et al. Automatic Skull Stripping of Rat and Mouse Brain MRI Data Using U-Net. *Frontiers in Neuroscience*, 14 :935, 2020.
- [156] Ipek Oguz, Honghai Zhang, Ashley Rumble, and Milan Sonka. RATS : Rapid Automatic Tissue Segmentation in rodent brain MRI. *Journal of Neuroscience Methods*, 221 :175–182, 2014.
- [157] P. Krähenbühl and V. Koltun. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. *Advances in Neural Information Processing Systems*, 24 :109–117, 2011.
- [158] Patty Coupeau, Jean-Baptiste Fasquel, Edouard Mazerand, Claudia N. Montero-Menei, Philippe Menei, and Mickael Dinomais. Patch-based 3D U-Net and transfer learning for longitudinal piglet brain segmentation on MRI. *Computer Methods and Programs in Biomedicine*, 214 :106563, 2022.
- [159] Shie Mannor, Dori Peleg, and Reuven Y. Rubinstein. The cross entropy method for classification. In *In Proceedings of the 22nd international conference on Machine learning*, 2005.
- [160] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, 35 :18–31, 2017.
- [161] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22 :1345–1359, 2010.
- [162] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How Transferable Are Features in Deep Neural Networks? In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, volume 2, pages 3320–3328, 2014.
- [163] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, and Natalia Gimelshein et al. Pytorch : An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [164] Varduhi Yeghiazaryan and Irina Voiculescu. Family of boundary overlap metrics for the evaluation of medical image segmentation. *Journal of Medical Imaging*, 5 :015006, 2019.
- [165] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. DeepLab : Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40 :834–848, 2018.
- [166] Patty Coupeau, Jean-Baptiste Fasquel, Edouard Mazerand, Claudia N. Montero-Menei, Philippe Menei, and Mickael Dinomais. U-Net 3D par patchs et régularisation spatiale avec CRF pour l’extraction de cerveau sur IRM de porcelet. In *ORASIS 2021*, 2021.

-
- [167] Joel Fluss, Mickael Dinomais, and Stéphane Chabrier. Perinatal stroke syndromes : Similarities and diversities in aetiology, outcome and management. *European Journal of Paediatric Neurology*, 23(3) :368–383, 2019.
- [168] Alexander Hammers, Richard Allom, Matthias J. Koepp, Samantha L. Free, Ralph Myers, Louis Lemieux, Tejal N. Mitchell, David J. Brooks, and John S. Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, 19 :224–247, 2003.
- [169] Marko Wilke, Scott K. Holland, Mekibib Altaye, and Christian Gaser. Template-O-Matic : a toolbox for creating customized pediatric templates. *NeuroImage*, 41 :903–913, 2008.
- [170] Mariam Al Harrach, François Rousseau, Samuel Groeschel, Xiaoyu Wang, Lucie Hertz-pannier, Stéphane Chabrier, Amine Bohi, Julien Lefevre, Mickael Dinomais, and AVCnn group. Alterations in cortical morphology after neonatal stroke : compensation in the contralesional hemisphere? *Developmental Neurobiology*, 79(4) :303–316, 2019.
- [171] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing : Overview, challenges and the future. *Classification in BioApps*, pages 323–350, 2018.
- [172] Alex M. Pagnozzi, Yaniv Gal, Roslyn N. Boyd, Simona Fiori, Jurgen Fripp, Stephen Rose, and Nicholas Dowson. The need for improved brain lesion segmentation techniques for children with cerebral palsy : A review. *International Journal of Developmental Neuroscience*, 47 :229–246, 2015.
- [173] Stéphane Chabrier, Béatrice Husson, Mickaël Dinomais, Pierre Landrieu, and Sylvie Nguyen The Tich. New insights (and new interrogations) in perinatal arterial ischemic stroke. *Thrombosis Research*, 127 :13–22, 2011.
- [174] Mickael Dinomais, Lucie Hertz-Pannier, Samuel Groeschel, Matthieu Delion, Béatrice Husson, Manoelle Kossorotoff, Cyrille Renaud, Stéphane Chabrier, and Sylvie Nguyen The Tich. Does contralesional hand function after neonatal stroke only depend on lesion characteristics? *Stroke*, 47(6) :1647–1650, 2016.
- [175] Andrea M. Kuczynski, Adam Kirton, Jennifer A. Semrau, and Sean P. Dukelow. Bilateral reaching deficits after unilateral perinatal ischemic stroke : a population-based case-control study. *Journal of neuroengineering and rehabilitation*, 15 :1–13, 2018.
- [176] Curt A. Sandman, Kevin Head, L. Tugan Muftuler, Lydia Su, Claudia Buss, and Elysia Poggi Davis. Shape of the basal ganglia in preadolescent children is associated with cognitive performance. *NeuroImage*, 99 :93–102, 2014.
- [177] Liangliang Liu, Yu-Ping Wang, Yi Wang, Pei Zhang, and Shufeng Xiong. An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Medical Image Analysis*, 81, 2022.
- [178] Wenju Yang, Guangqi Wen, Peng Cao, Jinzhu Yang, and Osmar R. Zaiane. Collaborative learning of graph generation, clustering and classification for brain networks diagnosis. *Computer Methods and Programs in Biomedicine*, 219, 2022.
- [179] Martin Hanik, Mehmet Arif Demirtaş, Mohammed Amine Gharsallaoui, and Islem Rekik. Predicting cognitive scores with graph neural networks through sample selection learning. *Brain imaging and behavior*, pages 1–16, 2021.

-
- [180] Surveillance of Cerebral Palsy in Europe. Surveillance of cerebral palsy in Europe : a collaboration of cerebral palsy surveys and registers. *Developmental Medicine and Child Neurology*, 42 :816–824, 2000.
- [181] Marjolein Jongbloed-Pereboom, Maria W.G. Nijhuis van der Sanden, and Bert Steenbergen. Norm scores of the Box and Block Test for children ages 3-10 years. *The American Journal of Occupational Therapy*, 67(3) :312–318, 2013.
- [182] Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M. Buhmann. The Balanced Accuracy and Its Posterior Distribution. In *20th International Conference on Pattern Recognition*, pages 3121–3124, 2010.
- [183] Ricco Rakotomalala. Analyse de corrélation. Étude des dépendances–Variables quantitatives (Version 1.1.). Tiré de http://www.eric.univ-lyon2.fr/~ricco/cours/Analyse_de, 2015.
- [184] Ronald A Thisted. What is a P-value. *Departments of Statistics and Health Studies*, 1998.
- [185] Wai Yen Loh, Peter J. Anderson, Jeanie L.Y. Cheong, Alicia J. Spittle, Jian Chen, Katherine J. Lee, Charlotte Molesworth, Terrie E. Inder, and A. Connelly et al. Neonatal basal ganglia and thalamic volumes : very preterm birth and 7-year neurodevelopmental outcomes. *Pediatric research*, 82(6) :970–978, 2017.
- [186] Mariam Al Harrach, Pablo Pretzel, Samuel Groeschel, François Rousseau, Thijs Dhollander, Lucie Hertz-Pannier, Julien Lefevre, Stéphane Chabrier, Mickael Dinomais, and AVCnn Study Group. A connectome-based approach to assess motor outcome after neonatal arterial ischemic stroke. *Annals of clinical and translational neurology*, 8(5) :1024–1037, 2021.
- [187] Erin D. Bigler. Structural image analysis of the brain in neuropsychology using magnetic resonance imaging (MRI) techniques. *Neuropsychology Review*, 25 :224–249, 2015.
- [188] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web : 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings 15*, pages 593–607, 2018.
- [189] Christina R. Marmarou. *Pyramidal System*. Springer, New York, NY., 2011.
- [190] Noah Hollmann, Samuel Müller, Katharina Eggenberger, and Frank Hutter. TabPFN : A Transformer That Solves Small Tabular Classification Problems in a Second, 2022.
- [191] Mateus Riva. *Spatial Relational Reasoning in Machine Learning : Deep Learning and Graph Clustering*. PhD thesis, Institut Polytechnique de Paris, Télécom Paris, 2022.
- [192] Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3) :379–423, 1948.
- [193] John Stewart Fabila-Carrasco, Chao Tan, and Javier Escudero. Permutation Entropy for Graph Signals. *IEEE Transactions on Signal and Information Processing over Networks*, 8 :288–300, 2022.
- [194] Andreia S. Gaudêncio, Mirvana Hilal, João M. Cardoso, Anne Humeau-Heurtier, and Pedro G. Vaz. Texture analysis using two-dimensional permutation entropy and amplitude-aware permutation entropy. *Pattern Recognition Letters*, 159 :150–156, 2022.

Titre : Apprentissage profond et relations structurelles pour l'analyse d'images : application à l'étude de la lésion cérébrale précoce chez l'enfant après AVC néonatal

Mots clés : Relations structurelles, réseaux de neurones sur graphes, segmentation sémantique, imagerie par résonance magnétique cérébrale, paralysie cérébrale, motricité

Résumé : L'apprentissage profond, notamment en analyse d'images, a connu un essor considérable ces dernières décennies sans pour autant exploiter la totalité de l'information mise à disposition par les images. Dans ce contexte, nous nous sommes intéressés à l'apport des connaissances structurelles de « haut-niveau » correspondant à des relations observables entre les objets présents dans l'image (ex. relations spatiales ou photométriques). Nous cherchons à intégrer ces informations, en s'appuyant sur des réseaux de neurones sur graphes (GNNs), pour améliorer les performances de l'apprentissage profond en analyse d'images. Nous présentons une contribution à la segmentation sémantique ainsi qu'une application en neuroimagerie dans laquelle les relations structurelles apparaissent

centrales: l'étude de la lésion cérébrale précoce. Cette lésion, intervenue sur un cerveau en cours de développement, peut provoquer des troubles moteurs permanents (paralysie cérébrale) et requiert des études sur l'humain (ex. enfants après AVC néonatal) ainsi que des modèles animaux pour mieux comprendre son impact. Dans ce cadre, nous présentons plusieurs contributions sur l'enfant et l'animal où les informations structurelles aident à la segmentation sémantique en IRM, en complément de l'apprentissage profond, mais également où elles permettent d'établir un lien entre la possible perte de motricité provoquée par la lésion et l'organisation macrostructurelle altérée d'une région spécifique du cerveau liée aux fonctions motrices: les noyaux gris centraux.

Title: Deep learning and structural relationships for image analysis: application to the study of early brain lesion in children following neonatal stroke

Keywords : Structural relationships, graph neural networks, semantic segmentation, brain magnetic resonance imaging, cerebral palsy, motor skills

Abstract: Deep learning, especially in image analysis, has grown considerably in the last decades without exploiting all the information provided by images. In this context, we are interested in the contribution of "high-level" structural relationships corresponding to observable relations between the objects present in the image (e.g. spatial or photometric relationships). We seek to integrate this information, using graph neural networks (GNNs), to improve the performance of deep learning in image analysis. We present a contribution to semantic segmentation as well as an application in neuroimaging in which structural relationships seem

central: the study of early brain lesion. This lesion, occurring on a developing brain, can cause permanent motor disorders (cerebral palsy) and requires human studies (e.g. children after neonatal stroke) as well as animal models to better understand its impact. Thus, we present several contributions on children and animals where structural information helps for semantic segmentation on MRI, as a complement to deep learning, but also where it allows to establish a link between the possible loss of motor skills caused by the lesion and the altered macrostructural organization of a specific brain region related to motor functions: the basal ganglia.