Design of a top-down computer vision algorithm dedicated to the recognition of cortical sulci

Léonie Borne

université PARIS-SACLAY

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

NNT : 2019SACLS322

Thèse de doctorat

Conception d'un algorithme de vision par ordinateur "top-down" dédié à la reconnaissance des sillons corticaux

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

Ecole doctorale n°575 Electrical, Optical, Bio: Physics and Engineering (EOBE)
Spécialité de doctorat : Imagerie et physique médicale

Thèse présentée et soutenue à Saclay, le 01/10/2019, par

LÉONIE BORNE

Composition du jury :

Isabelle Bloch Professeur, Télécom Paris-Tech, Univ. Paris-Saclay (LTCI)	Présidente
Pierrick Coupé Chargé de recherche, CNRS, Univ. Bordeaux (LABRI)	Rapporteur
François Rousseau Professeur, INSERM, IMT Atlantique (LaTIM)	Rapporteur
Arnaud Cachia Professeur, CNRS, Univ. Paris-Descartes (LaPsyDÉ)	Examinateur
Olivier Coulon Directeur de Recherche, CNRS, Univ. Aix-Marseille (INT)	Examinateur
Mark Jenkinson Professeur, Univ. Oxford (FMRIB)	Examinateur
Jean-François Mangin Directeur de recherche, CEA, Neurospin (UNATI)	Directeur de thèse
Denis Rivière Chargé de recherche, CEA, Neurospin (UNATI)	Co-encadrant de thèse

université PARIS-SACLAY

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

NNT : 2019SACLS322

Doctoral thesis

Design of a top-down computer vision algorithm dedicated to the recognition of cortical sulci

PhD thesis of University Paris-Saclay
prepared at University Paris-Sud

Doctoral school n°575 Electrical, Optical, Bio: Physics and Engineering (EOBE)
PhD speciality : Imaging and Medical Physics

Thesis presented and defended in Saclay, on October, 1^{st}, 2019, by

LÉONIE BORNE

Composition of the jury :

Isabelle Bloch Professor, Télécom Paris-Tech, Univ. Paris-Saclay (LTCI)	President of the jury
Pierrick Coupé Professor, CNRS, Univ. Bordeaux (LABRI)	Reviewer
François Rousseau Professor, INSERM, IMT Atlantique (LaTIM)	Reviewer
Arnaud Cachia Professor, CNRS, Univ. Paris-Descartes (LaPsyDÉ)	Examiner
Olivier Coulon Research director, CNRS, Univ. Aix-Marseille (INT)	Examiner
Mark Jenkinson Professor, Univ. Oxford (FMRIB)	Examiner
Jean-François Mangin Research director, CEA, Neurospin (UNATI)	Supervisor
Denis Rivière Professor, CEA, Neurospin (UNATI)	Co-supervisor

Acknowledgements

To Nicolas, who had not chosen to do a thesis, but who still experienced it by proxy. Thank you for supporting me, feeding me and taking me out even when it was raining.

To Denis, for his availability, his attentive ear and for showing me the pleasures of research and of Parisian subsoils. Thank you for these cata-ptivating experiences!

To Jean-François, for his support, encouragement and eternal optimism throughout this thesis. Thank you for being so for both of us.

To my colleagues for their help, whatever the solicitation, the discussions in the canteen, the cakes in the office that never stayed very long and the Chilean parties. Thank you for being there, simply.

To my running colleagues, for our discussions, our shared sweat, our cheesy evenings and our wild weekend with the plastic chicken. Thank you for your good humour at all times.

To my whole family, for their support and constant interest in my research, although I doubt I have been able to explain it properly. Thank you for your encouragement, from the very beginning.

Thank you to all of you without whom this thesis would not have been such an enriching experience, in discovery, knowledge and encounter.

Contents

Acknowledgements **i**

Introduction **1**
Context . . . 1
Challenges . . . 2
Thesis organization . . . 3
Thesis contribution . . . 3

I Background **5**

1 Cortical sulci **7**
1.1 Brain anatomy . . . 7
1.2 Sulci: anatomy, formation and link with functional organization 8
1.2.1 Sulci and gyri anatomy . . . 8
1.2.2 Mechanism at the origin of cortical folds . . . 9
1.2.3 Relations with cortical functional organization . . . 10
1.2.3.1 Abnormal fold patterns and psychiatric pathologies . . . 10
1.2.3.2 Sulci shape and brain function . . . 11
1.3 Sulci representation . . . 12
1.3.1 Overview . . . 12
1.3.1.1 Sulcus regions . . . 13
1.3.1.2 Sulcus roots/pits . . . 13
1.3.1.3 Sulcus bottom lines . . . 13
1.3.1.4 Sulcus segments . . . 14
1.3.2 Morphologist/BrainVISA representation . . . 14

2 Multi-Atlas Segmentation (MAS) **19**
2.1 General principles . . . 19
2.2 Patch-based MAS approach . . . 20
2.2.1 Nonlocal means label fusion . . . 21
2.2.2 Voxel-wise vs. Patch-wise label fusion . . . 21
2.2.3 Optimized PAtch-match Label fusion (OPAL) . . . 22

2.2.3.1 PatchMatch (PM) algorithm 22
2.2.3.2 Optimized PatchMatch (OPM) algorithm . . 23
2.2.4 Multi-scale and multi-feature label fusion 24

3 Convolutionnal Neural Networks (CNNs) 27
3.1 A model inspired by the brain's organization 27
3.2 Building Blocks of CNNs 29
3.2.1 Convolutional Layers 29
3.2.2 Pooling Layers 30
3.2.3 Fully-Connected Layers 30
3.3 CNNs architectures 31
3.3.1 Image Classification 31
3.3.2 Image Segmentation 34
3.4 How to train a neural network? 35
3.4.1 General approach 35
3.4.2 Tricks to improve training 36
3.4.2.1 How to avoid the vanishing/exploding gradient problem? 36
3.4.2.2 How to deal with overfitting? 36

II Applications 39

4 Automatic recognition of local patterns of cortical sulci 41
4.1 Introduction 42
4.2 Databases 44
4.2.1 Anterior Cingulate Cortex (ACC) pattern 44
4.2.2 Power Button Sign (PBS) 45
4.3 Method 45
4.3.1 Fold representation 45
4.3.2 Classification approaches 47
4.3.2.1 Support Vector Machine (SVM) 47
4.3.2.2 Scoring by Non-local Image Patch Estimator (SNIPE) 49
4.3.2.3 Convolutional Neural Network (CNN) 51
4.3.3 Performance evaluation 53
4.4 Results 54
4.4.1 Which is the best model? 54
4.4.2 Looking for PBSs 55
4.5 Conclusion 56

Appendix 61

5 Automatic labeling of cortical sulci 69
5.1 Introduction . . . 70
5.1.1 Overview of automatic sulci recognition methods . . . 71
5.1.2 New approaches: MAS and CNN . . . 71
5.1.3 Bottom-up geometric constraints . . . 72
5.2 Database . . . 73
5.3 Method . . . 74
5.3.1 Folds representation . . . 75
5.3.2 Labeling methods . . . 78
5.3.2.1 Statistical Probabilistic Anatomy Map (SPAM) models . . . 78
5.3.2.2 MAS approaches . . . 79
5.3.2.3 CNNs based approaches . . . 86
5.3.3 Bottom-up geometric constraints . . . 88
5.3.4 Performance evaluation of labeling models . . . 89
5.3.4.1 Mean/max error rates . . . 89
5.3.4.2 Error at the sulcus scale: E_{local} . . . 90
5.3.4.3 Error at the subject scale: E_{SI} . . . 91
5.3.4.4 Error rate comparison . . . 91
5.4 Results . . . 92
5.4.1 Which is the best model? . . . 92
5.4.2 Which sulci are better recognized? . . . 94
5.4.3 Does this improvement have an impact in practice? . . . 94
5.5 Discussion . . . 96
5.5.1 Patch-based MAS (PMAS) . . . 96
5.5.2 Patch-based MAS with High level representation of the data (HPMAS) . . . 97
5.5.3 Patch-based CNN (PCNN) and UNET . . . 98
5.5.4 Unknown label . . . 99
5.6 Conclusion . . . 99

Appendix 101

Conclusion 113
Contributions . . . 113
Limitations . . . 113
Perspectives . . . 114
Closing remarks . . . 115

Summary in French 117
Introduction . . . 117
Classification automatique des motifs locaux de plissements corticaux 118
Étiquetage automatique des sillons corticaux . . . 121
Conclusion . . . 123

Bibliography **125**

Introduction

Context

Considered as the seat of our intelligence, the brain is currently one of the most challenging enigmas that remains to be solved. Its function and in particular its dysfunctions fascinate. The link between the anatomy of the brain and its function is one of the unknowns of this enigma. For a while, it was common to think that our ability to calculate was related to the size of a bump in our cranium box: the "math bump". This expression comes from a pseudoscience, the phrenology, which was very popular in the 19th century. This pseudoscience attributed each aspect of an individual's personality to the extent of a bump in the cranium. Of course, the shape of the skull is not linked to that of the brain and this pseudoscience no longer has any credit. However, it has been shown that calculation actually involves specific areas of our brain, whose location is relatively stable from one individual to another. In fact, the appearance of non-invasive imaging, such as Magnetic Resonance Imaging (MRI), about 40 years ago allowed us to visualize the brain, first anatomically and then functionally. So is the functioning of these areas related to the anatomical organization of the brain? Regarding the anatomy of the brain's surface, the most noticeable feature is its convolutions, which are so variable that each individual has unique fold patterns. However, this striking feature is still poorly understood today.

From a Darwinian point of view, these folds are considered as a trick of evolution to increase the cortical surface without modifying the volume of the cranial cavity. Thus, the degree of folding, the complexity of their arrangement and their variability seem to be related to the degree of intelligence of each species. Indeed, most animals have a smooth brain, only some mammals, such as great apes (but also ferrets, cats, dogs, horses, cows, elephants, etc.), and some cetaceans, such as dolphins, have a folded brain. Note that some cetaceans have a brain with a volume and degree of folding greater than ours, which suggests a much higher degree of intelligence! This observation nevertheless shows the importance of folds in the uniqueness of the human species.

Regarding their formation, the cortex folds during the last trimester of

pregnancy and the sulci do not change their overall configuration afterwards. This period also corresponds to the emergence of cortical architecture: the migration of neurons into different predefined areas occurs just before the formation of folds. This observation suggests a link between cortical folds and the functional organization of the cortex. In fact, unusual fold patterns are generally linked to abnormal developments that can lead to psychiatric syndromes such as epilepsy or schizophrenia. Moreover, a close link has been demonstrated between the shape of certain sulci and cognitive functions such as manual laterality or reading. Therefore, the fold patterns appear to be signatures of each individual's functional organization. This hypothesis is now relatively accepted for the most important sulci but is more controversial for secondary sulci.

Anyway, the study of cortical folds requires a high level of expertise that few neuroanatomists have at this time. In order to facilitate their study, artificial intelligence algorithms are used in this thesis to automate tasks requiring advanced learning. But how do you teach a task to an algorithm? This question is addressed by a field of study of artificial intelligence: the machine learning. This approach is inspired by our own way of learning: for exemple, learning to read involves teaching based on decoding many words and texts. Similarly, machine learning algorithms "learn" from data. Today, artificial intelligence has made significant progress, especially in the field of computer vision, which is revolutionizing our daily lives. This is due to the emergence of a new method inspired by the brain function: the deep learning. This technique has allowed smartphones to talk, computers to beat chess champions, cars to drive on their own, etc. Will they also allow us to decipher our sulcal lines?

Challenges

The first challenge addressed during this thesis is the automation of local fold pattern classification. Based on a pattern identified by experts, the objective is to automate its recognition, which raises several issues. First, the distinction of patterns is sometimes difficult to the naked eye and requires ad-hoc criteria. Second, due to the extreme variability of cortical folds, the choice of classifier is a crucial step. Third, as we are interested in local patterns, the classifier must be able to handle particularly noisy data. Fourth, since manual labeling is time-consuming, the available learning databases are limited in size and can be largely unbalanced if the pattern is rare.

The second challenge concerns the automation of cortical sulci labeling. In order to study the geometry of cortical folds, neuroanatomists have attempted to define a sulci dictionary to differentiate these structures. Among the several proposed nomenclatures, the main sulci generally have a common label, but the fold variability is so high that there is no consensual

nomenclature for all sulci. However, whatever the nomenclature used, sulci labeling is generally tedious and requires a long expertise in the field. Thus, the automation of sulci recognition is essential to enable large-scale studies. This challenge has already been addressed several times, including in the laboratory hosting this thesis, but the results obtained are not yet satisfactory, on the one hand because the recognition rates are too low to reproduce the observations obtained with manual labeling and on the other hand because the recognition pipeline is dependent on an error-prone sub-segmentation process.

Thesis organization

This thesis is organized in two parts. The first part gives a state of the art on the topic under study (i.e. cortical sulci) and on the methods of medical image analysis on which the work of this thesis is based (i.e. multi- atlas segmentation techniques and convolutional neural networks). The second part presents the methods implemented for the automatic classification of local fold patterns and the automatic labeling of the sulci themselves. For these two problems, several methods are compared and the best method obtained is tested in practice.

Thesis contribution

The work carried out during this thesis has led to the emergence of two types of tools dedicated to the study of cortical folds, which will soon be available in the BrainVISA/Morphologist toolbox (http://brainvisa.info). The first tool automates the classification of local cortical fold patterns. This issue had never been addressed before. The second tool automates sulci labeling by eliminating the main defects of the model previously proposed by the BrainVISA/Morphologist toolbox. Thus, in addition to improving performance and speed, the proposed new model is robust to sub-segmentation errors, which is one of the greatest weaknesses of the old system.

This thesis continues the work of J.-F. Mangin (Mangin, 1995), D. Rivière (Riviere, 2000) and M. Perrot (Perrot, 2009) during their respective theses. It took place within the NAO (*Neuroimagerie Assistée par Ordinateur*) team of the Neurospin laboratory at CEA (*Commissariat à l'énergie atomique et aux énergies alternatives*) in Saclay. The work has benefit from an outstanding visualization environment provided by the Simulation house in order to manipulate and visualize hundreds of brains simultaneously (http://maisondelasimulation.fr).

Part I

Background

Chapter 1

Cortical sulci

Abstract

The complex and diverse arrangement of cortical folds is the most striking, interesting and yet poorly understood coarse morphological feature of the human brain. First, what is a sulcus? To answer this question, I will briefly review the anatomy of the brain in order to understand the place of sulci in its organization. Second, why study sulci? To understand the interest aroused by cortical folds, I will focus on their anatomy and its main components before presenting the mechanism of fold formation and their link with the functional organization of the brain. Third, how to study sulci? The choice of the fold representation is a key step for studying these structures. I will make a quick overview of the different representations used so far before developing the one used in this thesis.

1.1 Brain anatomy

Let's do a quick reminder about the anatomy of the human brain.

On a cellular scale, the brain is composed of nerve cells, called neurons, and glial cells. Neurons transmit information in the form of nerve impulses, while glial cells nourish, support and protect neurons. Glial cells also seem to have a significant, but still poorly understood, role in brain function. In order to communicate with each other, neurons have two types of extensions: dendrites that collect nerve impulses and carry them to the cell body, and the axon, a long fiber that carries information to other neurons.

On an anatomical scale, the cerebrum represents the largest part of the human brain. It is separated into two hemispheres communicating through the corpus callosum. Several brain nuclei are located in the centre of the brain: the central grey nuclei, the thalamus and the hypothalamus. Also in the center of the brain, there are cavities filled with cerebrospinal fluid: the ventricles. Finally, the cerebellum is below the rest of the brain.

The cellular bodies of cerebrum neurons are located either on the periphery where they form a grey layer, the cortex, or in nuclei buried in the depths of the brain. The axons form clusters of fibers, which constitute the white matter. In this thesis, we focus on a particular region of the brain, the cortex, which is supposed to be the seat of high-level functions.

1.2 Sulci: anatomy, formation and link with functional organization

1.2.1 Sulci and gyri anatomy

The cortex is divided into several convolutions, called gyri, delimited by folds, called sulci. On a larger scale, these structures are grouped into several large cortical lobes on each hemisphere, generally delimited by large deep sulci (Figure 1.1). Indeed, the central sulcus separates the frontal lobe from the parietal lobe; the parieto-occipital fissure separates the occipital lobe from the parietal lobe; the cingulate sulcus separates the frontal lobe from the limbic lobe on the inner side, etc. There is also a buried lobe at the bottom of the sylvian valley, between the temporal and frontal lobes: the insula.

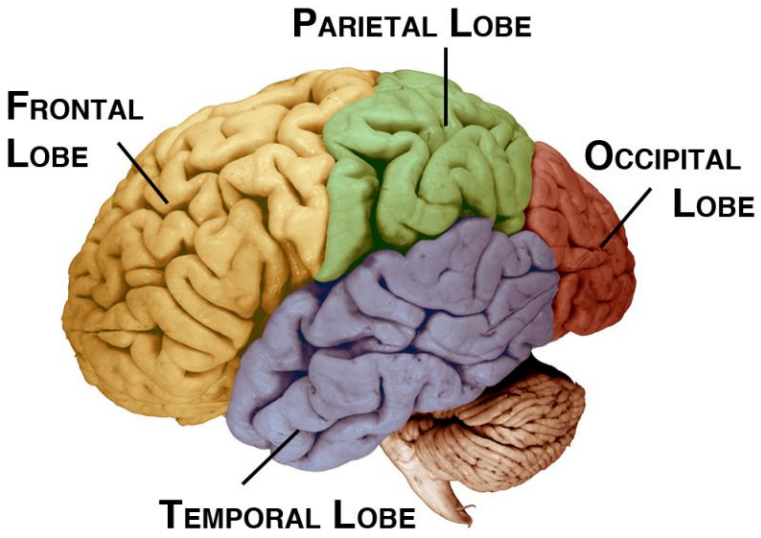


Figure 1.1 – Four main cortical lobes. (Miller and Cummings, 1999)

In order to study cortical folds, neuroanatomists have sought to associate a name to each sulcal or gyral structure, which has led to the emergence of several nomenclatures such as those described by Ono et al. (1990), Rademacher et al. (1992), Destrieux et al. (2010), etc. However, applying one of these nomenclatures, designed on a limited number of brains, to the

thousands of brains now available is much more complicated than expected. In fact, the folds are so variable that the labeling of a sulcus, even a primary one, can raise many questions. Note that if among the nomenclatures proposed so far, none of them is a consensus in the community, it is also because of the extreme variability of these structures.

1.2.2 Mechanism at the origin of cortical folds

Cortical folds appear during the third trimester of pregnancy. Once formed, their shape may vary slightly depending on the environment and the age of the subject. However, they maintain roughly the same configuration, which makes them excellent markers of early brain development. Nevertheless, the mechanism behind the formation of folds is not yet fully understood. Recently, a multicultural community has begun to focus on modeling the mechanisms that drive the folding process.

From a physical point of view, it has been demonstrated that the mechanical instability induced by the tangential expansion of the grey matter of the cortex is sufficient to simulate the folding process (Tallinen et al., 2014, 2016) (Figure 1.2). This result is a big step forward, but this modeling does not seem sufficient to obtain realistic folding patterns.

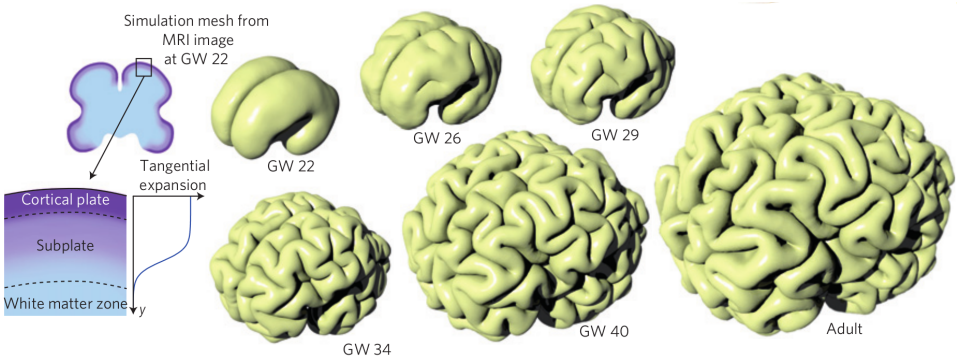


Figure 1.2 – Simulation of uniform tangential expansion of the cortical layer from a smooth fetal brain. This expansion is sufficient to reproduce the cortex gyrification process. (Tallinen et al., 2016)

Two other major hypotheses exist to explain the mechanism of sulci formation.

The first hypothesis is based on the relationship between folding and the existence of primary molecular maps, called protomaps (Fernández et al., 2016). These maps, supposed to be the result of heterogeneous gene expression, are present before the neuronal migration accompanying the construction of the cortex, followed by its folding. The existence of a "protomap" of the primary folding scheme is now relatively accepted (de Juan Romero et al., 2015). Thus, the areas of the protomap intended to become gyri

would present an increased multiplication of neurons, explaining the expansion maxima responsible for the formation of "bumps". Therefore, the mechanical model mentioned above could be improved by taking into account this protomap model.

The second hypothesis, compatible with the first, is that slight fibre tensions could play a role (Van Essen, 1997). After the appearance of primary sulci, these tensions could explain the following dynamic steps forming the secondary and tertiary folds. However, this hypothesis is far from being accepted.

The appearance of cortical folds is concomitant with the distribution of the cortex into different functional areas (Kostovic and Rakic, 1990; Sur and Rubenstein, 2005; Reillo et al., 2010). Is that a coincidence? I will discuss the links between folding patterns and functional organization in the next section.

1.2.3 Relations with cortical functional organization

The cortex is often divided into several zones, each supposed to be specialized in a different cognitive function (Amunts and Zilles, 2015; Glasser et al., 2016). The existence of close links between primary sulci and functional architecture is generally accepted (Welker, 1990; Fischl et al., 2007). These links are more controversial for other sulci, where the number of studies reporting similar results is very low (Watson et al., 1993; Amiez et al., 2006; Weiner et al., 2014). However, if we consider anatomical studies that are not based on a functional investigation, several of them suggest a much closer link than what has been demonstrated today.

1.2.3.1 Abnormal fold patterns and psychiatric pathologies

Abnormal folds patterns are usually related to anomalies in early brain development. For example, in (Bénézit et al., 2015), the abnormal development of interhemispheric fibres strongly disrupts the folding process.

Less obvious links have been identified with some neurodevelopmental diseases. For example, epilepsy is often due to incomplete migration of a group of neurons trapped in the white matter before reaching the cortical mantle: this group is called focal cortical dysplasia. This dysplasia is sometimes linked to abnormal fold patterns (Besson et al., 2008; Régis et al., 2011; Mellerio et al., 2014; Roca et al., 2015), even when the dysplasia is invisible on MRI (Figure 1.3).

Similarly, a link has been established between the patterns of the cingulate sulcus and schizophrenia. The shape of this sulcus can be categorized into two patterns: the single type and the double parallel type (Figure 1.4). It has been shown that patients with schizophrenia tend to have

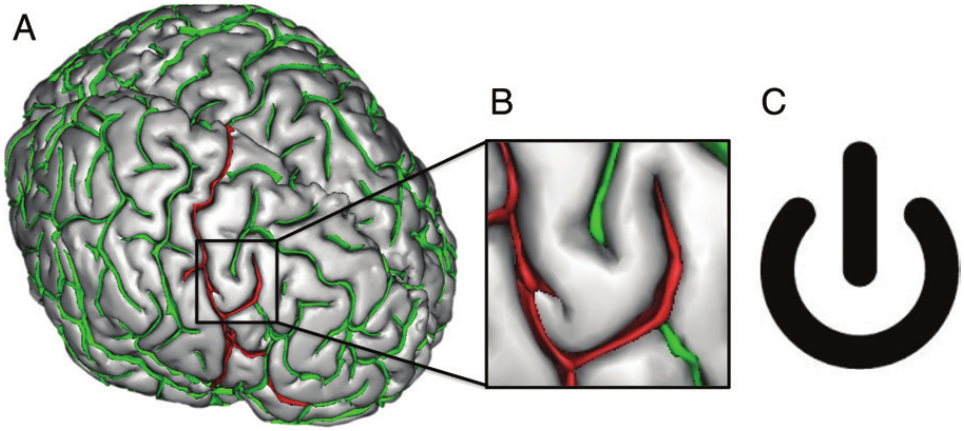


Figure 1.3 – The Power Button Sign: a fold pattern related to epilepsy. (Mellerio et al., 2014).

more symmetric patterns (same pattern on each hemisphere) than controls (Le Provost et al., 2003; Yücel et al., 2003).

It is interesting to note that children with the same lack of asymmetry have poor inhibitory control (Borst et al., 2014; Cachia et al., 2014). Thus, the study of folding patterns is not limited to psychiatric diseases but extends to the functioning of normal brains.

1.2.3.2 Sulci shape and brain function

In addition to studies based on fold patterns, other anatomical studies are based on a comparison of sulci length and depth between two groups of subjects (Mangin et al., 2010; Alemán-Gómez et al., 2013; De Guio et al., 2014; Janssen et al., 2014; Hamelin et al., 2015; Leroy et al., 2015; Muellner et al., 2015). These studies show that the links between function and sulci shape can be much more subtle than an abnormal pattern.

Consider the example of the central sulcus. It is commonly accepted that this sulcus delimits the primary motor cortex of the primary somatosensory cortex. This special position makes it a particularly interesting structure to study. Indeed, it has been shown that the shape of this sulcus is related to manual laterality (Sun et al., 2012) and recovery after a subcortical stroke (Jouvent et al., 2016). The link between its form and functional organization was also studied through functional MRI activations: Sun et al. (2016) shows that the position of the "hand knob" is related to the hand's motor activation, as well as to the premotor linguistic zone.

Therefore, studies on the geometry of cortical folds reveal links with the functional organization of the brain. However, these studies would not have

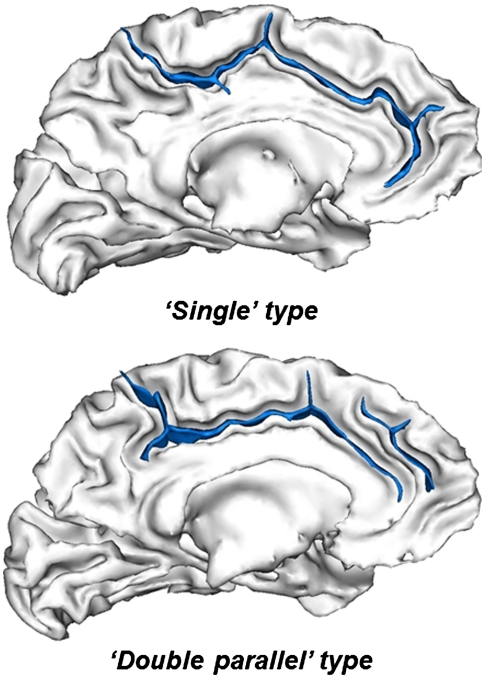


Figure 1.4 – Two folding patterns identified in the cingulate region. (Cachia et al., 2014)

been possible without representing in 3D the folds from the MRI acquisition, allowing for example to facilitate their visualization when defining patterns or to evaluate the characteristics of a sulcus (length, depth, opening, etc).

1.3 Sulci representation

In the literature, there are many different approaches for representing cortical folds, each with its advantages and disadvantages. In this thesis focusing on the classification of local fold patterns and on the sulci labeling, the representation chosen must be sufficiently exhaustive to allow the distinction of these complex structures while remaining standardized enough to be generalizable.

Therefore, in this section, I will give a quick overview of the representations of sulci presented in the literature, before developing the representation used in this thesis.

1.3.1 Overview

When considering cortical folds, two types of representations are possible: either the representations of bumps (gyri) or those of hollows (sulci).

The first type of representation, used for example by Fischl et al. (2004) and Chen et al. (2017), is based on a labelisation of the cortical surface into different gyri. These representations are probably more appropriate when considering functional MRI or diffusion MRI. In fact, white matter fibers appear to preferentially fan out in gyri (Van Essen, 1997), so there is a high probability that this is where the functional activity is also concentrated. However, gyri do not have a well delineated boundary and are generally defined by the sulci surrounding them.

Therefore, the second type of representation, based on sulci, is more obvious from an anatomical point of view because they are better defined structures. As the representation used in this thesis is based on sulci, I will develop in particular this type of representation below.

1.3.1.1 Sulcus regions

As for gyri representation, this representation of sulci is based on the cortical surface labelisation. Each sulcus is represented by a sulcus region, i. e. the surface of the cortex that surrounds the sulci (Behnke et al., 2003; Rettmann et al., 2005; Vivodtzev et al., 2006; Yang and Kruggel, 2009).

This representation is particularly effective in capturing all variations of cortical surface depth, including dimples. However, it is not ideal for representing the 3D layout of these structures. Moreover, it takes into account the sulci's opening, which is highly dependent on the subject's age. Other representations allow better normalization of the data.

1.3.1.2 Sulcus roots/pits

In order to overcome the high variability of folds, it was proposed to focus on sulcus roots, also called sulcus pits. These structures, corresponding to local maximum depths, represent the position of the first folds during their formation in utero. Their selection is supposed to provide more information on the architectural organization of the cortex (Régis et al., 2005; Lohmann et al., 2007; Im et al., 2009; Mangin et al., 2015a; Im and Grant, 2019).

However, these approaches represent the sulci by a set of points where only critical parts of the structure shape are preserved, which is not appropriate for our applications.

1.3.1.3 Sulcus bottom lines

In the late 1990s, the use of pipelines representing the cortical surface, no longer in 3D, but as a 2D spherical surface (Dale et al., 1999), promoted surface representations of sulci. These surface representations are based on the sulci bottom line as it is more accessible once the cortical surface is represented by a 2D mesh (Kao et al., 2007; Shi et al., 2008; Li et al., 2010; Shattuck et al., 2009; Seong et al., 2010; Lyu et al., 2018).

Although this representation is more appropriate than sulcal roots for sulci labeling, it does not consider information related to the deep extension of sulci. Yet, knowing the depth or inclination helps to recognize some sulci: for example, primary sulci are deeper than others, collateral sulcus has a characteristic inclination, etc.

1.3.1.4 Sulcus segments

When sulci are considered as hollows filled with cerebrospinal fluid, they can be extracted by making a negative mold of the cortical surface. To avoid representing the sulci opening, the extracted structure can be displayed by median surfaces between the two walls surrounding each sulcus. These median surfaces are then divided into several sulcus segments, also called elementary folds.

Compared to sulcus bottom lines, the sulcus segments are particularly interesting for representing information related to the deep extension of sulci. Two representations are based on this notion of sulcus segments. The first representation corresponds to smooth structures, called sulcus ribbons, which do not represent the branches of sulci (Le Goualher et al., 1997; Vaillant and Davatzikos, 1997; Zeng et al., 1999; Zhou et al., 1999). The second representaion corresponds to the interface between the two walls bordering the sulci (Mangin, 1995; Le Goualher et al., 1998; Riviere et al., 2002; Klein et al., 2005; Perrot et al., 2011). By displaying branches of sulci, this representation is therefore much more suitable for studying local patterns of cortical folds. Thus, this more exhaustive representation has been used in this thesis.

1.3.2 Morphologist/BrainVISA representation

The Morphologist pipeline of the BrainVISA toolbox provides a representation of sulcus segments taking into account the branches of sulci (Rivière et al., 2009). This pipeline, first described in (Mangin et al., 1995), then more recently in (Riviere et al., 2002) and (Mangin et al., 2004), constitutes the preprocessing, allowing the extraction of sulci, used in this thesis. It consists of four major steps: first, the segmentation of white and grey matter from MRI, then the extraction of the skeleton of cortical folds, followed by its division into elementary folds and finally the representation of these folds as a graph (Figure 1.5).

- **White/grey matter segmentation:** First, MRIs are analyzed to segment white matter and grey matter. To do this, a bias correction is made to eliminate signal intensity inhomogeneities depending on spatial position due to weaknesses in the acquisition process. Then, the intensity distribution is analyzed to detect grey and white matter, whose intensities vary according to the MR sequence and the subject.

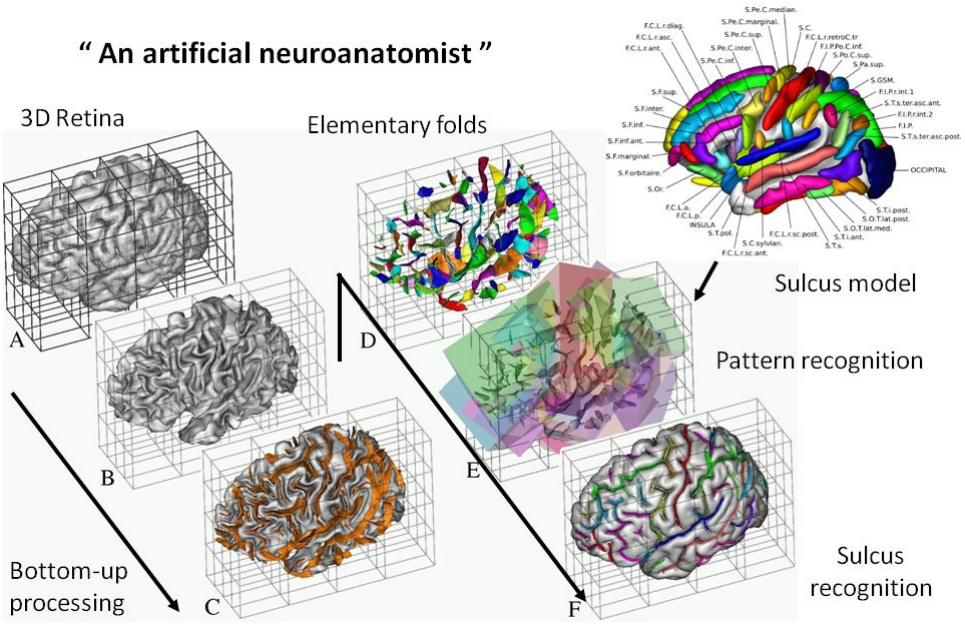


Figure 1.5 – A computer vision pipeline mimicking a human anatomist (Mangin et al., 2015b). A: interface between the cerebral envelope and the cortex. B: interface between white matter and grey matter. C: extraction of the fold skeleton. D: cutting of the skeleton into elementary folds. E: Folds labeling using the model of Perrot et al. (2011)

Finally, by using the histogram produced, the image is binarized by keeping the range of intensities that can belong to the brain. The binary image is eroded iteratively until a seed is obtained which is then dilated to form the shape of the brain.

- **Sulci skeleton:** Second, the resulting segmentation is used to extract the sulci skeleton, located in the cerebrospinal fluid filling the folds. After separation of the two hemispheres and the cerebellum, each hemisphere is represented by a spherical object whose outer surface corresponds to the interface between the cerebral envelope and the cortex and the inner surface to that between white and grey matter. The object obtained, corresponding to a 3D negative mold of the white matter, is then skeletonized. The skeleton points connected to the outside of the brain are used to form the brain envelope. The rest of the points represent cortical folds.

- **Elementary folds:** Third, this part of the skeleton is divided into several elementary folds. Each elementary fold is supposed to correspond to a single sulcus label. Several factors are considered to cut the fold skeleton: first, topological characterization is used to isolate sur-

face pieces that do not include any junctions (Malandain et al., 1993; Mangin et al., 1995) and second, the skeleton is cut to the minimum depths, which correspond to buried gyri (Figure 1.6).

- **Graph representation:** Finally, the pipeline also proposes to aggregate the elementary folds in the form of graphs designed for the sulci recognition method described in (Riviere et al., 2002). Each elementary fold represents a node of the graph and the relationship with the other nodes corresponds either to a topological separation (branches), or to the presence of a buried gyrus (pli de passage), or to that of a convolution (gyrus) between two unconnected elementary folds.

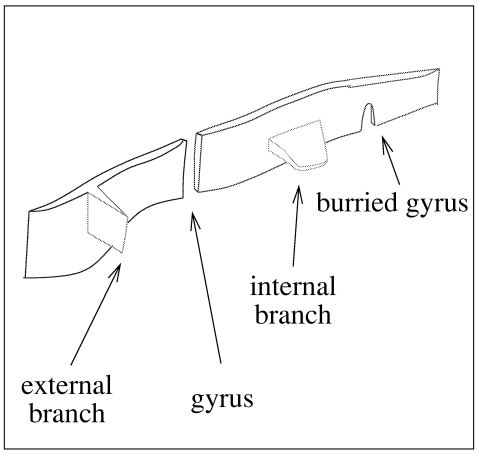


Figure 1.6 – Schematic representation of the fold skeleton. The fragmentation into elementary folds isolates the internal and external branches and cuts the skeleton at the level of the buried gyri. (Riviere et al., 2002).

This pipeline extracts the sulci skeleton reliably enough to allow the study of local fold patterns and the labeling of cortical sulci. However, the cutting of the skeleton into elementary folds, which is particularly useful for regularizing sulci labeling, is considerably less robust. Indeed, when labeling sulci, each elementary fold is supposed to have a unique label. However, it happens that the fragmentation of the skeleton is insufficient and that some elementary folds actually contain several labels. Moreover, vastly different fragmentations can be observed from the same MRI. In fact, several stochastic optimizations are included in the segmentation pipeline (e. g. for bias correction, brain masking, skeletonization, etc.). These optimizations only have a slight impact on the shape of the resulting fold skeleton, but for the topological fragmentation into elementary folds, a single voxel can then make the difference. Thus, these stochastic optimizations can have important consequences on the fragmentation of large simple surfaces.

The Morphologist/BrainVISA pipeline offers two methods for automatic labeling of elementary folds: the first is a graph matching approach (Riviere et al., 2002) and the second is based on Statistical Probabilistic Anatomy Map (SPAM) (Perrot et al., 2011). Each of these two methods assigns a label by elementary fold and does not allow to reconsider the bottom-up segmentation of the elementary folds. To eliminate the shortcomings associated with the fragmentation into elementary folds, this thesis propose to complement the regularization scheme using a top-down perspective which triggers an additional cleavage of the elementary folds when required.

Conclusion

This chapter has highlighted why cortical folds are still poorly understood and why new tools are required to study them. In this thesis, the tools developed aim on the one hand to automate the classification of local patterns of cortical folds and on the other hand to automatically label sulci. These tools will allow the study of sulcal geometry characteristics on the very large databases currently available, which is not possible with manual recognition. For each of these tools, the Morphologist/BrainVISA pipeline is used to extract thes folds. This pipeline allows the fold skeleton to be segmented in a robust and reliable way, but the fragmentation of the skeleton into elementary folds is sometimes problematic. Thus, one of the major objectives of this thesis is to overcome these fragmentation errors when labeling sulci.

Chapter 2

Multi-Atlas Segmentation (MAS)

Abstract

The automatic recognition of cortical sulci shares the same characteristics as segmentation problems. Indeed, as explained in the previous chapter, sulci are represented in this thesis in the form of a skeleton of folds from which we want to label each voxel. Similarly, segmentation aims to label each voxel in an image. Among the most common segmentation methods nowadays, Multi Atlas Segmentation (MAS) is now widely used in medical imaging because it allows a better representation of the variability of anatomical structures than a model based on an average template. Note that the SPAM model proposed today by the BrainVISA/Morphologist toolbox is indeed based on an average template while cortical folds are known for their extreme variability. Thus, a MAS approach seems to have the potential to significantly improve the performance of the current model.

In this section, I will describe the general principle of MAS techniques and then detail the patch approaches used in this thesis.

2.1 General principles

MAS approaches, introduced about ten years ago (Rohlfing et al., 2004; Klein et al., 2005; Heckemann et al., 2006), are among the most effective methods for medical image segmentation today. Indeed, these approaches manipulate each manually labeled image in the database as an "atlas" used to label a new image. This allows them to better represent anatomical variability than a model based on an average representation, and therefore to have a better segmentation accuracy. This advantage, however, generally comes at a high calculation cost.

As proposed in (Iglesias and Sabuncu, 2015), MAS approaches can be described in four main steps:

Generation of atlases The first step consists in creating atlases, i.e. labeled training images. Each labeling generally requires a costly effort from domain-specific experts, using interactive visualization tools.

Registration In order to establish spatial correspondences between the atlases and the image to be segmented, they must be registered to the image. This step is essential for successful segmentation: it is necessary to optimize spatial alignment while maintaining an anatomically plausible deformation. To do this, it is important to carefully choose the deformation model (i.e. rigid or non-rigid registration), the distance to be minimized and the optimization method. However, this crucial step can be extremely costly depending on the modalities used.

Label Propagation Once the spatial correspondence of the atlases has been established with the image to be segmented, the atlas labels are propagated to the coordinates of the new image.

Label Fusion The last step consists in combining the propagated labels of the atlases in order to obtain the final segmentation. This step, ranging from simple selection of the best atlas to probabilistic fusion of atlases, is one of the main components of MAS.

One of the critical steps in MAS approaches is the registration of atlases to the image to be segmented. Indeed, a poor registration can significantly penalize segmentation accuracy. However, in our opinion, there is currently no sufficiently precise registration of sulci for this task. Some approaches recently proposed, however, do not require a precise registration: these are the patch approaches.

2.2 Patch-based MAS approach

Introduced by Coupé et al. (2011) and Rousseau et al. (2011), the nonlocal patch-based label fusion have become more popular in recent years (Wang et al., 2011; Fonov et al., 2012; Zhang et al., 2012; Asman and Landman, 2013; Bai et al., 2013; Konukoglu et al., 2013; Wolz et al., 2013; Wang et al., 2014; Ta et al., 2014; Romero et al., 2017a,b). These approaches use a patch search strategy to identify matches with the atlases and use the non-local means method of Buades et al. (2005) to fuse the labels. Moreover, since they no longer require costly non-linear registration, these approaches can

be computationally efficient with an appropriate implementation (Ta et al., 2014; Giraud et al., 2016), which is a significant advantage in practice.

In this section, I will first briefly present the foundations of patch label fusion approaches, and then discuss the various proposed improvements.

2.2.1 Nonlocal means label fusion

In the label fusion strategy described in (Coupé et al., 2011), each voxel of the image to be segmented is labeled based on a cubic patch surrounding that voxel. This patch is compared to several patches in the atlas library: the distance between two patches is used to make a robust weighted average of all patch matches. This weighted average is based on the non-local mean estimator proposed by Buades et al. (2005). The value of this estimator is finally used to determine the label of the voxel.

Since comparing each patch of the image to be segmented to all the patches of all the atlases in the library would be extremely computationally expensive, different techniques are used to limit the number of patch matches. First, the region of interest of the image to be segmented is extracted in order to reduce the number of voxels to be labeled. Second, not all atlases in the library are used: only some atlases are pre-selected based on their similarity to the image to be segmented. Finally, a selection is made among the patches of the selected atlases: only the atlas patches located in the neighbourhood of the position of the voxel to be labeled are taken into account.

This label fusion strategy has produced results similar or even superior to the contemporary segmentation algorithms (Coupé et al., 2011; Eskildsen et al., 2012). In addition, this segmentation approach has also been adapted for the classification of patients with Alzheimer's disease (AD) and cognitively normal subjects (CN) (Coupé et al., 2012). This classification is based on the non-local mean estimator to asses the proximity of each voxel to both populations in the training database (Figure 2.1). Thus, this approach provides an estimate of the grade (i.e., the degree of proximity of one group or another) for each voxel. The average of these scores is then calculated to obtain an estimate of the subject's grade.

2.2.2 Voxel-wise vs. Patch-wise label fusion

In the method presented above, only the central voxel of the patches is propagated to the image to be segmented. However, if the patch strongly matches the image to be segmented, it is likely that it is not the only voxel that is interesting to propagate. Thus, Rousseau et al. (2011) successfully propose to propagate all the voxels of the compared patches. This approach also provides better regularization of the results and has proven its efficiency in several applications (Wu et al., 2014; Manjón et al., 2014a).

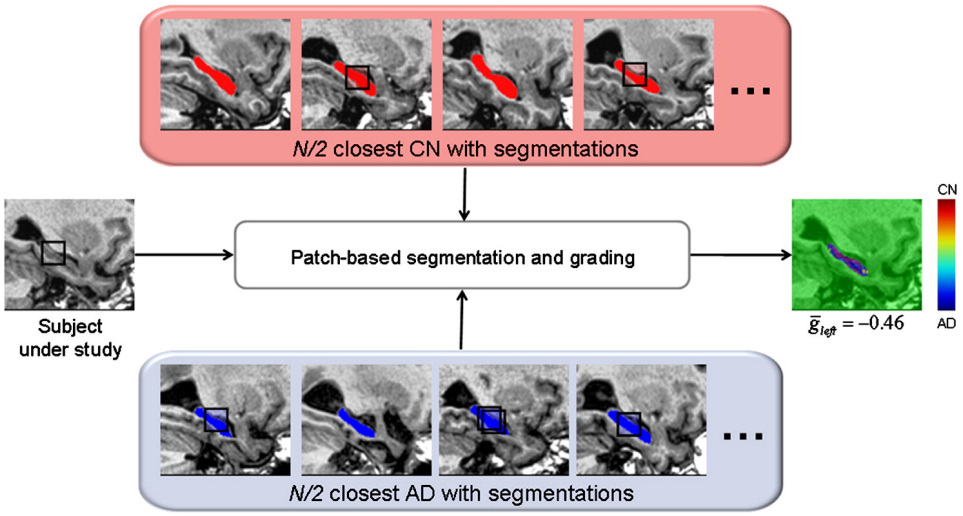


Figure 2.1 – Global overview of the grading and segmentation method proposed in (Coupé et al., 2012), known as SNIPE (Scoring by Non-local Image Patch Estimator). The nearest N/2 subjects in the atlas library are selected from both populations (AD and CN). Each patch of the studied subject is compared with the patches of the selected atlases in order to segment the studied structure and note its similarity with the two populations.

2.2.3 Optimized PAtch-match Label fusion (OPAL)

Despite the different techniques used in (Coupé et al., 2011) to reduce the number of patches to be compared, the computation time is still important. To remedy this, a search strategy to find the Approximate Nearest Neighbours (ANNs) of each patch to be labeled, inspired by the PatchMatch (PM) algorithm (Barnes et al., 2011), was proposed: the Optimized Patch-Match (OPM) algorithm. This algorithm gave its name to the new label fusion strategy that results from it: the Optimized PAtch-Match Label fusion (OPAL) strategy.

2.2.3.1 PatchMatch (PM) algorithm

The PM algorithm is an efficient strategy to find good patch matches between two images (initially in 2D). It is based on the assumption that, given a patch $p_A(i,j)$, belonging to image A and whose central pixel is located at coordinates (i,j), that matches with the patch $p_B(i',j')$ belonging to image B, the adjacent patches of $p_A(i,j)$ will probably match the adjacent patches of $p_B(i',j')$.

This algorithm is implemented in three steps: initialization, propagation and random search. During initialization, a random patch of the image B is assigned to each patch of the image A. The adjacent patches are then

tested during the propagation step, in order to improve the current match by following the principle mentioned above. Finally, during the random search step, randomly selected patches in concentric neighbourhoods of the current match are tested in order to avoid local minima. These last two steps are performed iteratively to improve correspondences.

2.2.3.2 Optimized PatchMatch (OPM) algorithm

The OPM algorithm proposes to adapt the PM algorithm to biomedical applications by exploiting the fact that two anatomical images can generally be roughly aligned (Ta et al., 2014; Giraud et al., 2016). Thus, the search for a similar patch no longer needs to be in the entire image but only in a part of it.

Compared to the PM algorithm, the OPM algorithm manipulates 3D images and does not look for patch matches between two images but between an image and a library of images (i.e. the atlases). One of the major advantages of this algorithm is that its duration depends only on the size of the image to be segmented and not on the size of the library.

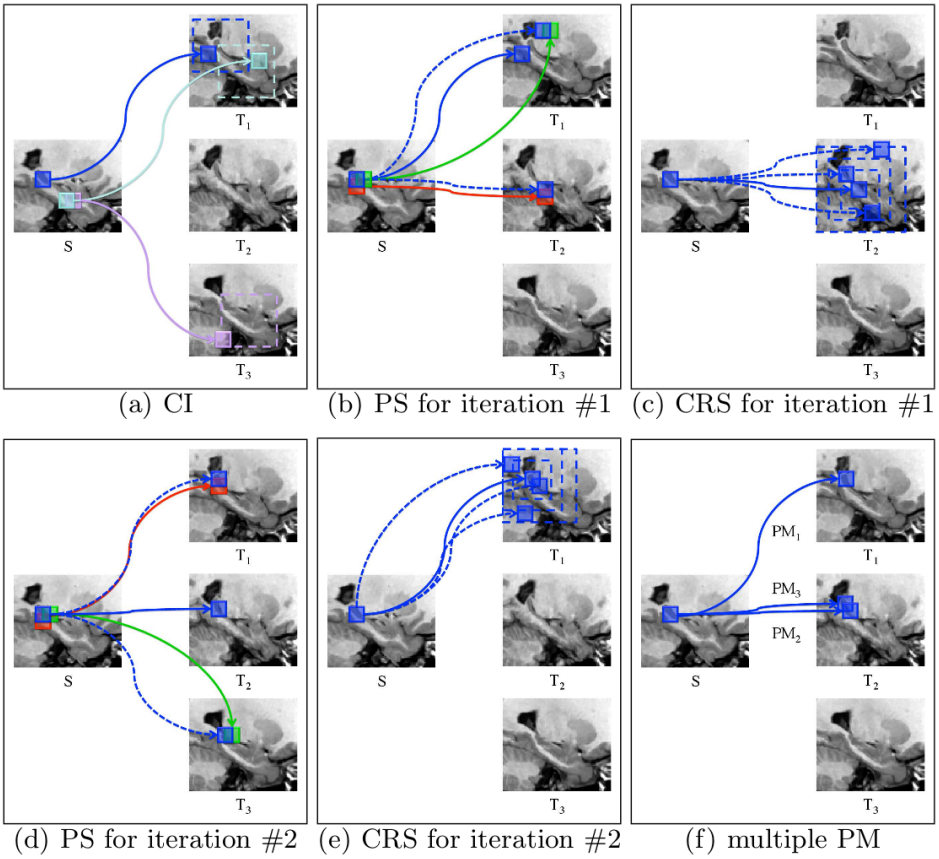
As with PM, the OPM algorithm is implemented in three steps: constrained initialization, propagation step and constrained random search. These last two steps are performed iteratively a predetermined number of times (Figure 2.2).

- **Constrained Initialization** The constrained initialization aims to assign to each voxel of the image to be segmented, a patch of the atlas library being located approximately in the same region. To do this, an atlas is randomly selected in the library. Then, a patch is randomly selected in the neighbourhood of the voxel to be labeled. The width of this neighbourhood is determined a priori according to the variability of the problem to be managed.

- **Propagation Step** Once all the voxels in the image have been matched to a patch in the atlas library, the next step is to propagate the best matches. Thus, as for the PM algorithm, for each voxel of the image to be labeled, it is checked whether the propagation of the patches of one of its adjacent voxels does not allow a better match.

- **Constrained Random Search** Once the propagation step is completed, a random constrained search is performed. As with the PM algorithm, for each match, it looks for a better match in the same image by testing several randomly selected patches. For the OPM algorithm, this search is constrained by a search window that closes a little more each time around the position of the matched patch.

Whereas before, all the tested patches were involved in the label fusion, this time only those returned by the OPM algorithm are propagated. In



(a) CI (b) PS for iteration #1 (c) CRS for iteration #1

(d) PS for iteration #2 (e) CRS for iteration #2 (f) multiple PM

Figure 2.2 – OPAL main steps. (a) Constrained initialization (CI), (b) and (d) propagation step (PS) for iteration #1 and #2, respectively (c) and (e) constrained random search (CRS) for iteration #1 and #2, respectively and (f) multiple Patch Match (PM). (Ta et al., 2014)

order to obtain several patches per voxel to label, the algorithm is run several times independently.

2.2.4 Multi-scale and multi-feature label fusion

Some approaches propose to use different patch sizes and/or features, which allows them to significantly improve the results in some cases (Eskildsen et al., 2012; Manjón et al., 2014b; Giraud et al., 2016; Romero et al., 2017b,a). For example, Giraud et al. (2016) proposes to merge the score maps obtained for each modality by following the late fusion principle of Snoek et al. (2005) (Figure 2.3).

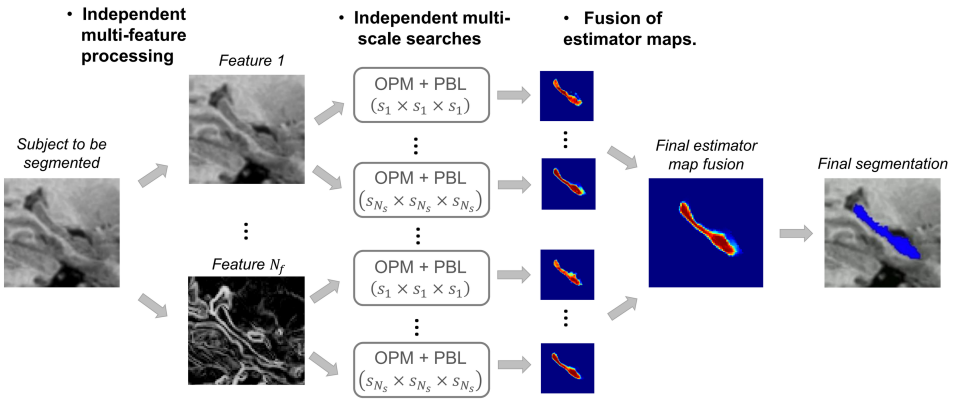


Figure 2.3 – Fusion of multi-feature and multi-scale label estimator maps. The algorithm is applied with N_s different patch sizes, on N_f different features, so $N = N_s * N_f$ estimator maps are computed and merged to provide the final segmentation. (Giraud et al., 2016)

Conclusion

In conclusion, MAS patch approaches seem particularly appropriate for cortical sulci labeling for two main reasons. First, they effectively represent the variability of the structures to be segmented. Second, they do not require any precise registration between the atlases and the image to be segmented, which is difficult to obtain for cortical sulci without first labeling them.. For the same reasons, these approaches will also be tested for the automatic classification of cortical folds, based on the approach proposed in (Coupé et al., 2012).

Chapter 3

Convolutionnal Neural Networks (CNNs)

Abstract

Neural networks, and in particular convolutional neural networks (CNNs), are famous for their remarkable effectiveness in dealing with many computer vision problems. These algorithms, inspired by the brain architecture, have indeed revolutionized image recognition over the past ten years. In this thesis, this trendy approach is used to automatically classify fold patterns and automatically recognize sulci. Therefore, this chapter will detail the history and principles of this method in order to understand why it was selected for these two problems.

3.1 A model inspired by the brain's organization

In the 1950s, attempts were made to model the brain function by representing its operating units: the neurons. The Perceptron of Rosenblatt (1958) is a simplified mathematical model of how neurons work (Figure 3.1): it receives a set of binarized inputs (from several other neurons), it processes this information by weighting the inputs (representing the strength of the synapse of the neuron transmitting the information) and it thresholds the sum of the weighted inputs to have a binarized output according to the sum magnitude (representing a neuron activation when the signal is high enough). By adjusting the input weights, the Perceptron allows to model the basic OR/AND/NOT functions, which was a challenge at that time.

In the 1980s, the Perceptron was integrated into a model inspired by the organization of the visual cortex: the Neocognitron (Fukushima, 1980). In this model, neurons are organized into several hierarchical layers where each layer is responsible for detecting a pattern from the outputs of the previous layer using a sliding filter. The first CNN was created!

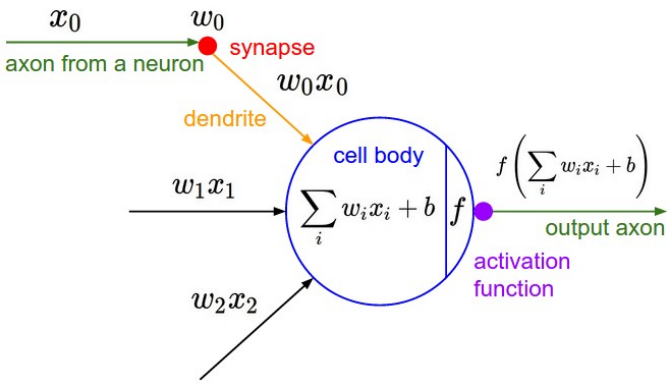


Figure 3.1 – A diagram showing how the Perceptron works (Fei-Fei et al., 2015). The Perceptron receives a set of binarized inputs x_0, x_1 and x_2. It processes this information by weighting the inputs by w_0, w_1 and w_2 respectively. It thresholds the sum of the weighted inputs $\sum_i w_i x_i$ at bias b thanks to an activation function f, in order to have a binarized output according to the sum magnitude. Note that in most current models, output and inputs are no longer necessarily binary because other activation functions, imitating the thresholding process, are now used, such as the sigmoid function.

Although the Neocognitron is effective for some pattern recognition tasks, these convolutional filters were then manually configured. It was only in the 1990s that automatic learning of neural networks was proposed using backpropagation (Werbos et al., 1990). The first convolutional neural network trained in this way is LeNet-5 (Figure 3.2), which allows automatic recognition of handwritten numbers (LeCun et al., 1998). The architecture of this network is composed of several types of neural layers, which are still widely used in more recent architectures. The following section will develop the principles of these different layers.

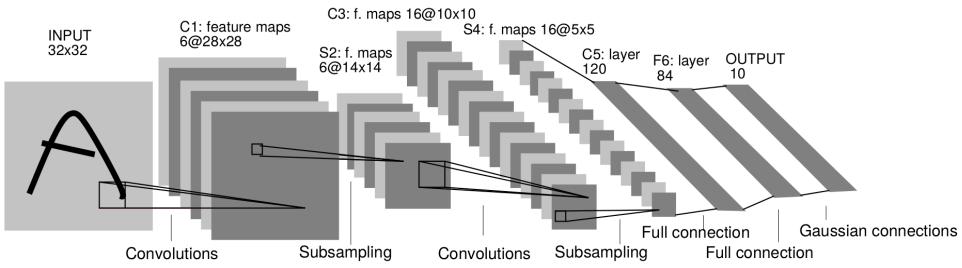


Figure 3.2 – LeNet-5 architecture, a CNN for digits recognition (LeCun et al., 1998).

3.2 Building Blocks of CNNs

The major advantage of CNNs is that they process images efficiently by taking into account their spatial organization. For 2D image processing, each layer of these networks is organized in three dimensions: width, height and depth. The width and height define the size of the feature maps associated with the layer, and its depth determines their number. Three types of layers are mainly used by CNNs: convolutional layers, pooling layers and fully connected layers.

3.2.1 Convolutional Layers

In order to avoid connecting each neuron in a layer to all the neurons in the previous layer, which would be extremely computationally expensive, convolution layers limit the inputs of a neuron to a fixed size window. For a given depth of the output layer, the weights are shared through the different view windows on the input image. Therefore, the neurons along the depth "watch" the same window but analyze it differently (Figure 3.3).

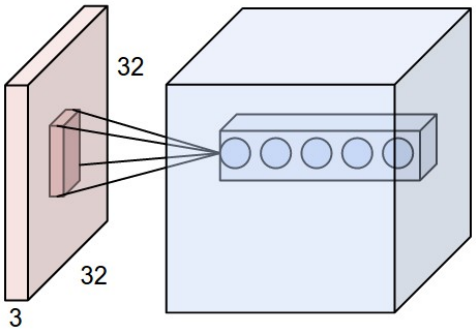


Figure 3.3 – Illustration of a convolution layer. The input volume is shown in red: it corresponds to a 2D image, of size 32*32, whose 3 RGB color channels are represented along the depth of the input. The layer of neurons is represented in blue. Each neuron in the layer is connected only to a local region of the input volume but to the entire depth. Several neurons (5 in this example) along the depth are connected to the same region of the input volume. (Fei-Fei et al., 2015)

The first hyperparameter, called the receptive field, corresponds to the size of the input region observed by a column of neurons. Three other hyperparameters allow to control the output volume: the depth, i.e. the number of neurons in a column of neurons, the stride, i.e. the step between each observation window of two successive columns of neurons, and the size of the zero-padding, i.e. the quantity of zeros added at the edges of the input volume.

Therefore, the convolutional layers allow the image to be analyzed using different filters, whose nature depends on the weights assigned to each area of the input window. Each filter sees all the input features and produces a new output feature. Usually, the number of output features is greater than the number of input features. In order to control the size of these features, convolutional layers go hand in hand with pooling layers.

3.2.2 Pooling Layers

Pooling layers are generally placed between two successive convolution layers. They gradually reduce the size of the representation, and therefore the number of parameters and calculations in the network. Two hyperparameters are used to define these layers: the stride, i. e. the filter application step, and the spatial extent, i. e. the filter size. The most commonly used filter is the maximum function (Figure 3.4).

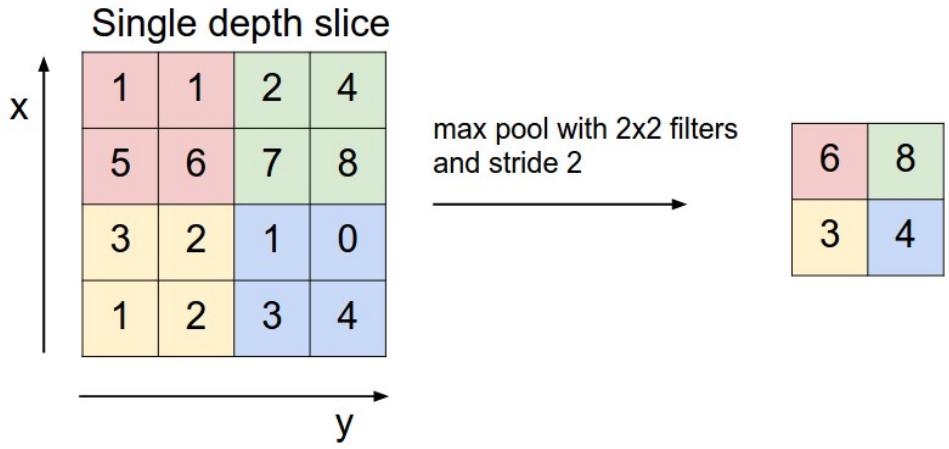


Figure 3.4 – Illustration of a Max-Pooling Layer. Each color represents a different input window for filters with a spatial range of 2*2 and stride 2. For each input window, only the highest value is kept. (Fei-Fei et al., 2015)

Convolutional layers and pooling are repeated iteratively in order to reduce the size of the features while increasing their number. This allows the CNN to represent simplified features of the images. Once these features are extracted, they must be processed to perform the classification: for this purpose, the fully connected layers are used.

3.2.3 Fully-Connected Layers

Fully connected layers include neurons connected to all the neurons of the previous layer. This type of layer is used to learn how to classify images, once the images have been "analyzed" by the convolutional and pooling layers.

After LeNet-5 for handwritten number recognition, other neural network architectures have been proposed to address different problems. The following section describes the historical evolution of architectures developed for issues related to those of this thesis.

3.3 CNNs architectures

This section discusses the CNNs architectures used for image classification and segmentation. In fact, this thesis focuses on two applications: the automatic classification of local folding patterns and the automatic labeling of the sulci themselves. The first task is similar to image classification: we want to classify the subjects, or their hemispheres, according to the pattern they present. The second task is more like image segmentation: for each voxel of the fold skeleton, we want to assign a label to it.

3.3.1 Image Classification

The classification of 2D images has made spectacular progress thanks to the use of neural networks. This is particularly well illustrated by the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015), held every year from 2010 to 2017.

In 2012, the first CNN wins this challenge (Krizhevsky et al., 2012) by reducing the top-5 error from 26.2% to 15.3%. This five-layer convolutional network, known as AlexNet, is trained using Graphics Processing Units (GPUs) and has an architecture similar to LeNet-5 (LeCun et al., 1998). This article uses several techniques that are now widespread: data augmentation and dropout to avoid overfitting and the use of GPUs to accelerate learning.

In 2014, new neural network architectures significantly reduced the top-5 error from 15.3% to about 7%. GoogleNet, also known as Inception V1 (Szegedy et al., 2015), wins the competition by offering a deeper architecture (22 layers) thanks to an optimized implementation. The second winner is the VGG network (Simonyan and Zisserman, 2014) which uses small convolutional filters (3*3*3) to deepen its network architecture which goes up to 16 or 19 layers. It is now one of the most popular architectures because of its simplicity to train.

In 2015, the ResNet network (He et al., 2016) wins the competition with a top-5 error of 3.5%. The architecture of this network has been designed to allow a much deeper network than the VGG (i.e. 152 layers instead of 19) while maintaining a lower complexity (Figure 3.6). Moreover, in order to facilitate the training of this particularly deep network, this model uses "skip connections", i.e. shortcuts performing identity mapping (Figure 3.5). Thanks to these connections, the layers only have to learn the residual functions with reference to the layer inputs, instead of learning unreferenced

functions. These skip connections have also inspired the architecture of the DenseNet (Huang et al., 2017) where all layers are connected to each other instead of just connecting the input and output of a residual block. This enabled, among other things, to obtain similar results to the ResNet by using a less deep architecture, containing fewer parameters to learn, and therefore being less computationally expensive.

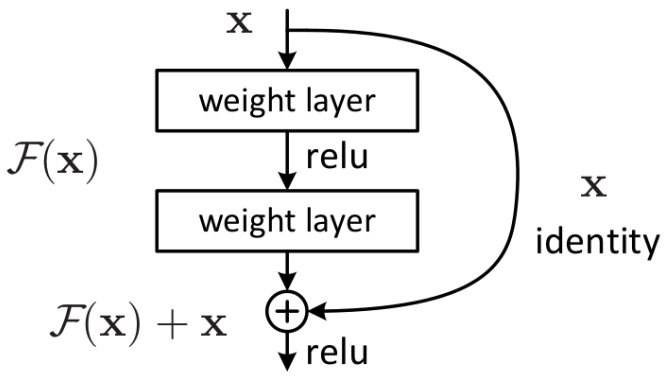


Figure 3.5 – Illustration of a "skip connection" for residual learning. (He et al., 2016)

In 2017, the Squeeze-and-Excitation (SE) network won the competition with a top 5 error of only 2.3% (Hu et al., 2018). This network is based on a new architectural unit, the SE block, which adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels.

All these advances show the importance of network architecture in the search for the best error rate. Today, a community of researchers is interested in automating the design of the network architecture to adapt it to the given problem (Zoph et al., 2018).

Note that all the spectacular advances described above concern the classification of 2D images, available in very large quantities for the ILSVRC. The classification of 3D images raises new challenges: on the one hand because of the large size of the inputs to be processed and on the other hand because of the difficulty of obtaining images labeled in large quantities. To date, there is no challenge as renowned as the ILSVRC for 3D image classification, which prevents comparing the different neural network architectures for this problem. However, it is planned that the ILSVRC will be replaced by a new competition that aims to help robots see the world in all its depth.

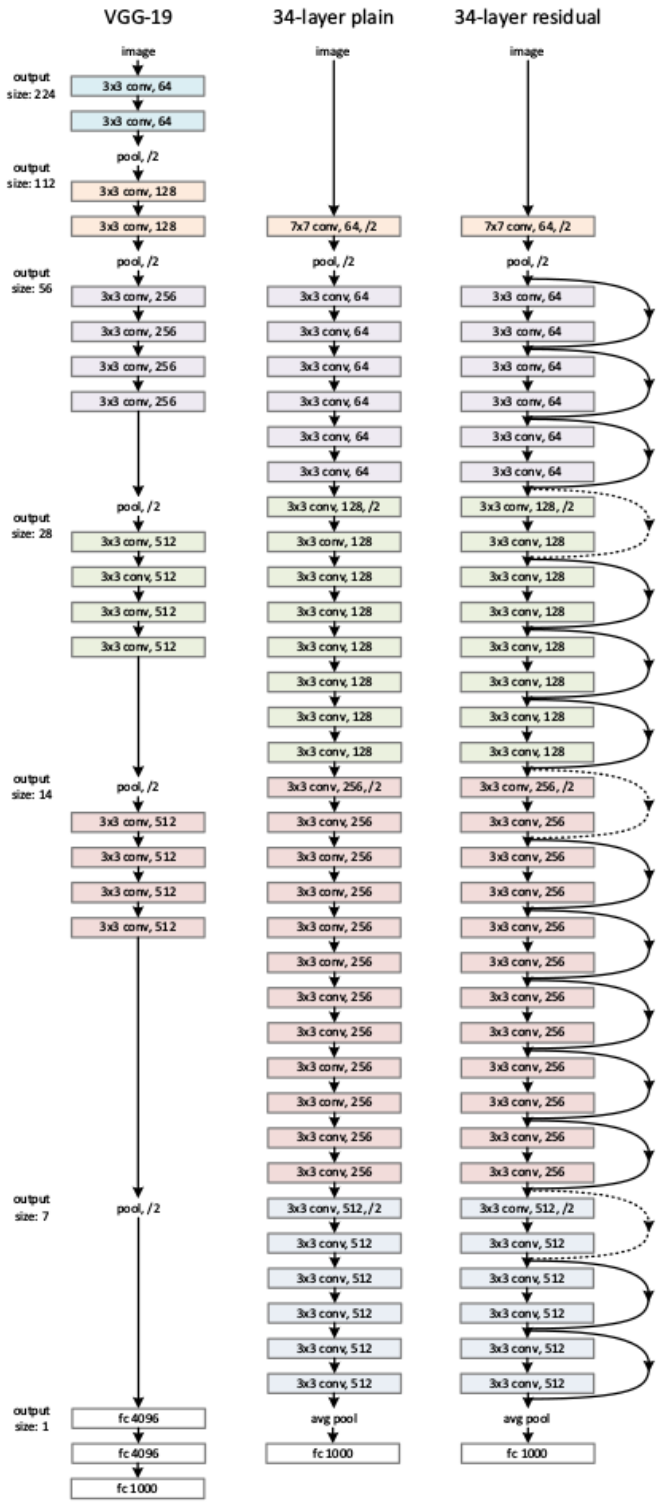


Figure 3.6 – Example network architectures for ILSVRC. **Left:** the VGG-19 model (19.6 billion FLOPs, i.e. floating-point operations per second) as a reference. **Middle:** a plain network with 34 parameter layers (3.6 billion FLOPs). **Right:** a residual network with 34 parameter layers (3.6 billion FLOPs). (He et al., 2016)

3.3.2 Image Segmentation

Regarding image segmentation, the use of a patch approach was first proposed by Ciresan et al. (2012). The method described in this article aims to train a neural network to classify each pixel according to its neighbourhood, i.e. a square patch of which it is the central pixel. This approach, although particularly effective, is nevertheless costly to apply.

More recently, fully connected networks have been used for semantic segmentation (Long et al., 2015). This approach is based on the network architecture designed for image classification mentioned above and adds a layer to reproduce the size of the input image. Note that the image size no longer needs to be fixed: the output image has the same size as the input image. Based on this approach, the use of deconvolutions was then proposed (Noh et al., 2015). Unlike the previous network, this approach does not brutally reproduce the segmented image after several convolutions but allows to progressively enlarge the feature maps obtained to reproduce the image. This idea inspired Ronneberger et al. (2015) to propose a specific architecture for medical imaging: the U-Net (Figure 3.7). This architecture is now widely used and has been adapted for 3D medical image segmentation by Çiçek et al. (2016) and Milletari et al. (2016).

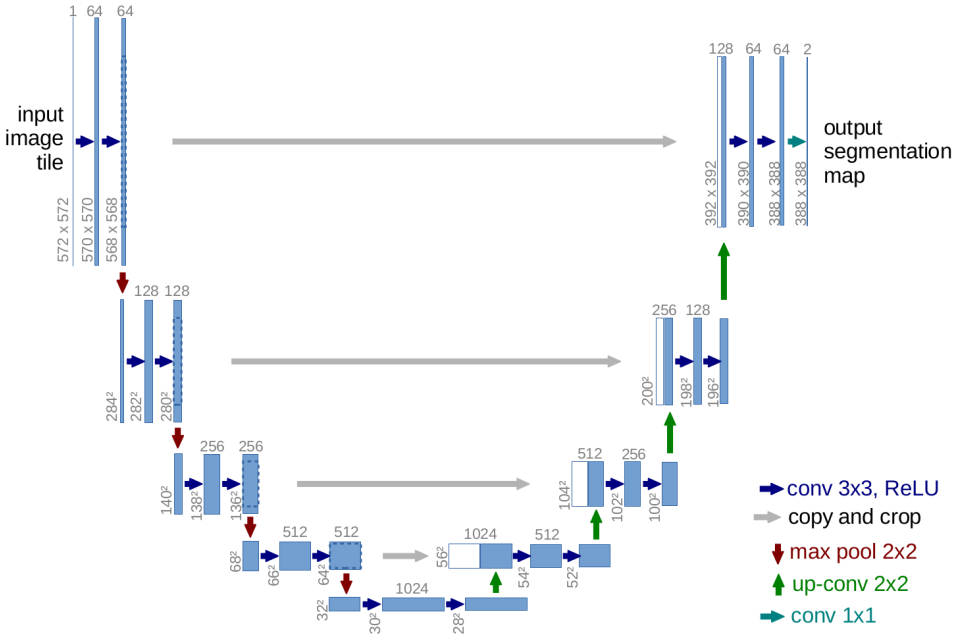


Figure 3.7 – U-net architecture, a fully convolutional network for medical image segmentation. (Ronneberger et al., 2015)

Recently, Google proposed a new neural network for semantic segmen-

tation (Chen et al., 2018) that introduces atrous convolutions (also called dilated convolutions). These convolutions introduce a new parameter called the dilation rate that defines a spacing between the input values considered. This allows a more global vision, supposed to be more adapted to segmentation than the convolutions used for image classification. The interesting results obtained by this approach show that, even today, image segmentation is a hot topic.

Although it is important to understand recent neural network architectures, automating image classification or segmentation does not only depend on the network organization. Training the network is also a critical step that I will discuss in the next section.

3.4 How to train a neural network?

The design evolution of the CNN architecture shows a tendency to create deeper models. However, the deeper the network, the more parameters it contains to learn and the harder the training is to converge. For a long time, training neural networks was almost impossible to converge in a reasonable time. The use of GPUs has considerably accelerated the process, especially as they become more and more powerful and accessible. However, today, it is mainly many tricks that help them to converge. In this section, I will briefly explain how a neural network is trained and then detail some of these tricks.

3.4.1 General approach

When a new data to be modeled is presented to the neural network, the first step is to calculate the output of the network: this is the **forward propagation**.

During training, we want to modify the architecture of the neural network, i. e. its parameters, so that the output obtained is closer to the desired one. For this purpose, a **loss function** is calculated. This function must be derivable and get higher values when the output is far from the expected learning value.

In order to minimize the loss function, several optimization algorithms have been proposed, the first and most common being the **Stochastic Gradient Descent** (SGD). SGD is a gradient descent method used to minimize an objective function, here the loss function. The stochastic, i. e. random, side of this algorithm allows it to avoid small local minima and therefore to be more robust. When training a neural network, this algorithm adapts the parameters of each neuron by calculating the gradient of the loss function according to the parameter to be optimized. Depending on the gradient value, the parameter value is then adjusted to minimize the loss function.

In order to calculate the gradients for each parameter of the model, SGD uses **backpropagation**. This tool first calculates the gradients of the neuron parameters of the last layer, then iteratively goes up the network to calculate the gradients of the neuron parameters of the previous layers.

3.4.2 Tricks to improve training

In this section, I will discuss two of the main problems encountered when training a neural network: the vanishing/exploding gradient problem that prevents the network from learning and overfitting that makes the network learn incorrectly.

3.4.2.1 How to avoid the vanishing/exploding gradient problem?

First of all, what is the vanishing gradient problem? During learning, gradients are used to update neuron parameters. If these gradients are too low, updating the parameters will have only a small impact. However, during backpropagation, the gradients of the model parameters with respect to the loss function are calculated by chain rule: each gradient is a multiplication of partial derivatives that depend on the previous layers. If some of these derivatives are close to zero, multiplying them will make the gradients even smaller, i.e. " vanishing ", which will cause the mentioned problem. This is all the more complicated to manage if the network is deep because the parameter gradients of the first layers will depend on an even higher number of partial derivatives. Conversely, if these gradients are greater than one, the gradient values may explode, causing the opposite problem: the parameters are unstable.

Now, how can we avoid it? First, by carefully initializing the parameters of the neural network. For example, Glorot and Bengio (2010) initializes the connection weights so that the input of the activation functions is normalized. Indeed, if the activation function is a sigmoid, input values too far from zeros have almost zero derivatives. Other activation functions have been proposed to avoid this problem: for example, the ReLU function has a zero derivative only for inputs below zero, large inputs will have the same derivative, i.e. one.

Then, the input data are generally normalized. Indeed, this makes it possible on the one hand to ensure that they are in the same range of values and on the other hand to avoid saturation of the activation functions. In the same vein, batch normalization aims to normalize the input of each layer during training.

3.4.2.2 How to deal with overfitting?

Another recurrent problem when training a neural network, especially when it is deep, is overfitting. This phenomenon is manifested by an increase

of the neural network's performance on the learning set, accompanied by a decrease on the validation set. In fact, neural networks are particularly data-intensive and easily overfit on small learning bases.

In order to limit overfitting, the first solution is to expand the size of the learning base. To do this, the data can be increased by various transformations of the images: rotation, translation, cropping, noise effect, etc.

Another important step to avoid overfitting is the training design. In particular, the choice of the learning rate, the number of epochs, etc. For example, a too short learning time may lead to underfitting the model on the learning and test set, while a too long learning time may lead to overfitting on the learning set while having poor performance on the test set. In order to find the right balance, early stopping is generally used: this technique aims to stop learning when the model's performance starts to deteriorate on a validation set, i.e. a part of the learning set reserved to test the model at each epoch.

Finally, one of the most popular tricks is the use of dropout during network training (Srivastava et al., 2014). This technique aims to "drop out" some neurons during learning: a neuron is kept active according to a probability determined a priori, otherwise it is "extinguished" (Figure 3.8). Thus, only a small network is trained on the data at this stage. The deleted neurons are then reintegrated into the network with their original weight.

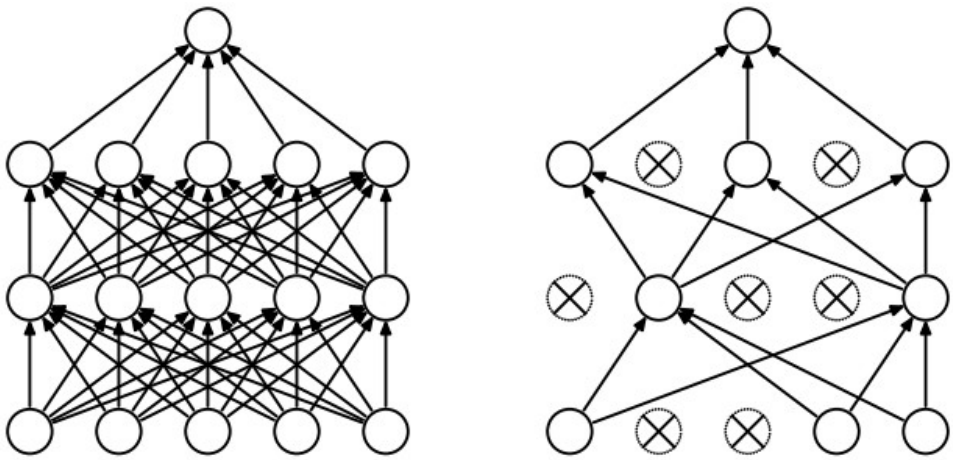


Figure 3.8 – Illustration of the Dropout. (Srivastava et al., 2014)

Conclusion

To conclude, this chapter highlights the importance of choosing the architecture of the neural network on the one hand and the training design on

the other. In this thesis, these CNN models inspired by the brain function will be tested to study one of its most striking anatomical characteristics: the cortical folds.

For the automatic classification of local folding patterns, a 3D ResNet will be used. Based on the study conducted by Hara et al. (2018) comparing the spatiotemporal 3D CNN adaptations of the 2D CNNs used for ILSVRC, a 3D ResNet has been choosen for the automatic classification of local folding patterns, as it is the architecture obtaining the best performance in (Hara et al., 2018).

For the automatic recognition of cortical sulci, two types of approaches will be tested: the first inspired by the patch segmentation approach proposed by Ciresan et al. (2012) and the second using a fully convolutional network, i.e. a 3D U-Net (Çiçek et al., 2016). These two approaches were initially proposed for the segmentation of medical images and have proven their effectiveness in this task.

Part II

Applications

Chapter 4

Automatic recognition of local patterns of cortical sulci

Abstract

Despite the tools now available to visualize cortical folds in 3D, manually classifying local sulcal patterns is a time-consuming and tedious task. In fact, 3D visualization of folds helps experts to identify different sulcal patterns but fold variability is so high that the distinction between these patterns sometimes requires the definition of complex criteria, making manual classification difficult. However, the assessment of the impact of these patterns on the functional organization of the cortex could benefit from the study of large databases. In this chapter, the automation of the classification of two types of patterns is proposed to allow morphological studies to be extended and confirmed on such large databases. To do so, three methods are proposed, the first based on a Support Vector Machine (SVM) classifier, the second on the Scoring by Non-local Image Patch Estimator (SNIPE) approach proposed by (Coupé et al., 2012) and the third based on a 3D Convolution Neural Network (CNN). Two types of patterns are studied: the Anterior Cingulate Cortex (ACC) patterns and the Power Button Sign (PBS). The two ACC patterns are almost equally present whereas PBS is a particularly rare pattern in the general population. The three models proposed have interesting performances for these two classification problems. The CNN-based model is more interesting for the classification of ACC patterns thanks to its rapid execution. However, SVM and SNIPE-based models are more effective in managing unbalanced problems such as PBS recognition.

4.1 Introduction

During the third trimester of pregnancy, the surface of the human fetus' brain forms convolutions, called gyri, delimited by folds, called sulci. The mechanism of fold formation and the impact of this anatomical characteristic on brain function are still poorly understood today. However, it has been demonstrated that the spatial pattern of some cortical sulci, markers of early brain development (Cachia et al., 2016), is associated with local brain function. Indeed, several studies have reported a correlation between visually recognized sulcal patterns and some cognitive characteristics – e.g. cognitive control efficiency (Fornito et al., 2004; Cachia et al., 2014) or handedness (Sun et al., 2012) – and neuropsychiatric illnesses – e.g. epilepsy (Mellerio et al., 2014) or schizophrenia (Plaze et al., 2015). By allowing the automatic recognition of these sulcal patterns, it will be possible to extend and confirm such studies on larger databases and better understand the subtle links between the sulcal shapes and the functional architecture.

In this chapter, I will focus on automating the identification of two types of sulcal patterns: first, the classification of the Anterior Cingulate Cortex (ACC) patterns, and second, the recognition of the Power Button Sign (PBS).

The sulcal pattern of the ACC can be classified with structural MRI (Cachia et al., 2014, 2016) in two different types (Figure 4.1): a "single" type, with only the cingulate sulcus and the "double parallel" type, with an additional paracingulate sulcus. Initially, a link has been demonstrated between these patterns and schizophrenia: schizophrenic subjects tend to have symmetric patterns between their right and left hemispheres, unlike controls (Le Provost et al., 2003; Yücel et al., 2003). Later, it was also shown that children with the same lack of asymmetry have poor inhibitory control (Borst et al., 2014; Cachia et al., 2014). Thus this pattern is not only related to a psychiatric disease but also to the healthy brain function. Automating the recognition of this pattern is particularly difficult because of its high variability (Figure 4.2). For this purpose, we have a database of 348 hemispheres, including 207 with the double parallel pattern, of which the cingulate and paracingulate sulci have been manually labeled.

The second pattern we are interested in is related to epilepsy. This disease is often due to incomplete migration of a group of neurons trapped in the white matter before reaching the cortical mantle: this group is called focal cortical dysplasia. This dysplasia is sometimes linked to abnormal fold patterns (Besson et al., 2008; Régis et al., 2011; Mellerio et al., 2014; Roca et al., 2015), even when the dysplasia is invisible on MRI. Here, we are interested in the automatic detection of one of these patterns: the PBS (Figure 4.3), present in about 60% of patients with dysplasia of the motor area, whereas in the general population, this pattern is so rare that we do

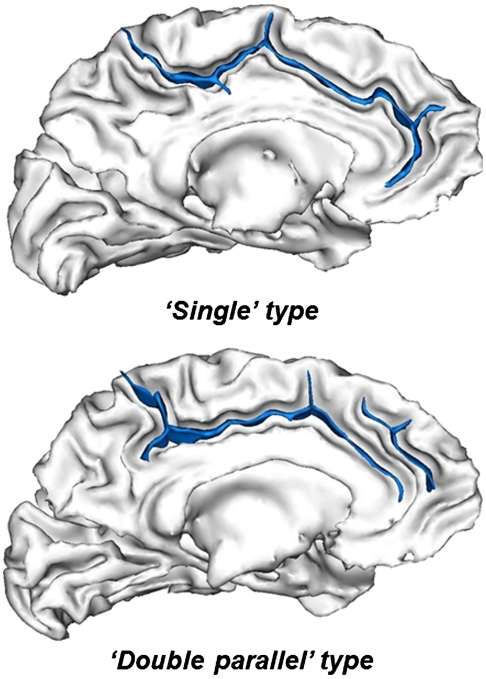


Figure 4.1 – Two folding patterns identified in the cingulate region. The "single" pattern is characterized by the presence of the cingulate sulcus alone. The "double parallel" pattern has an additional paracingulate sulcus, which must be parallel to the cingulate sulcus and long enough to be identified. (Cachia et al., 2014)

not know its proportion (Mellerio et al., 2014). This pattern is characterized by the interposition of a precentral sulcal segment between the central sulcus and one of its hook-shaped anterior ascending branches. Automating PBS recognition is particularly challenging because, in addition to the extreme variability of this pattern (Figure 4.4), the available labeled database is limited (19 controls and 38 patients with dysplasia, i. e. 114 hemispheres) and the proportion of hemispheres with a PBS is particularly unbalanced (22 subjects, including 1 control and 21 patients with dysplasia, have at least one of the two hemispheres with PBS, i. e. 28 hemispheres with PBS in the database).

In this chapter, I will first present the three proposed methods for the automatic recognition of these patterns. Two of them use manually labeled hemispheres to evaluate the similarity of a new hemisphere to the patterns. The third method uses a 3D neural network to automatically classify patterns.

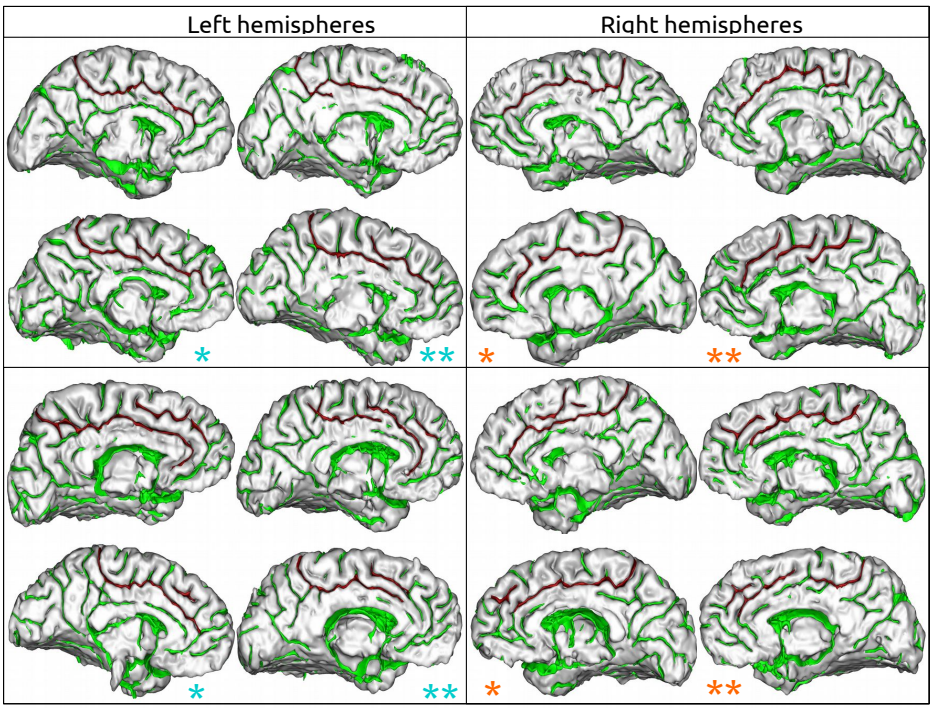


Figure 4.2 – Illustration of ACC patterns variability. The upper part of the table shows the single patterns and the lower part shows the double parallel patterns. We observe that the position, size and number of components of the paracingulate sulcus are variable. It is sometimes difficult to distinguish with the naked eye between a large paracingulate sulcus and a double cingulate sulcus (see **). Additionally, small paracingulate sulci can sometimes be confused with other small sulci (see *).

4.2 Databases

The BrainVISA/Morphologist pipeline, freely available at http://brainvisa.info/, was used to segment and provide 3D graphical representation from raw MRI data. This 3D mesh-based reconstruction of cortical folds was used to manually visualize and label brains.

4.2.1 Anterior Cingulate Cortex (ACC) pattern

The MRI of 190 subjects (including 74 normal controls, 67 schizophrenia patients and 49 healthy siblings) were selected from a large prospective study on brain development at the National Institute of Mental Health (NIMH). The ACC sulcal pattern was visually labeled for each individual hemisphere (Cachia et al., 2014, 2016). The cingulate and paracingulate sulci were manually labeled. Sixteen subjects with at least one hemisphere with am-

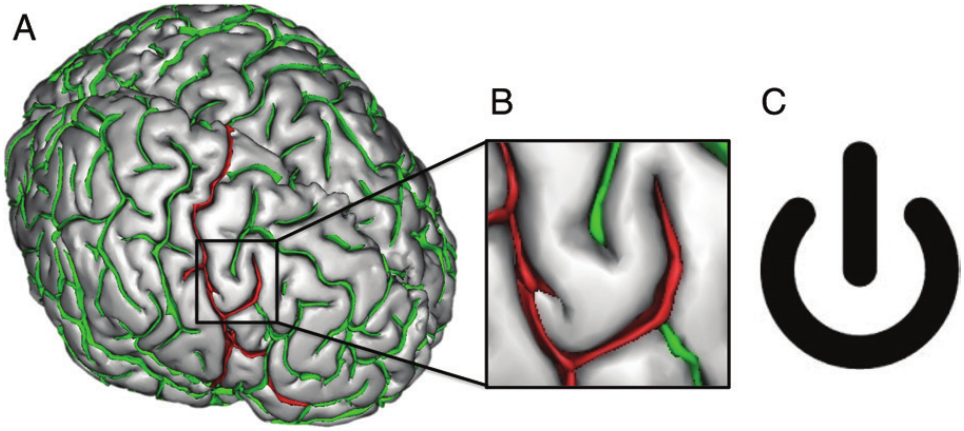


Figure 4.3 – The Power Button Sign: a fold pattern related to epilepsy. This pattern is characterized by the interposition of a precentral sulcal segment (in green) between the central sulcus (in red) and one of its hook-shaped anterior ascending branches (in red). (Mellerio et al., 2014)

biguous pattern types were not included in the analysis. The learning base was therefore composed of 348 labeled ACC sulcal patterns, including 141 "single" types and 207 "double parallel" types.

4.2.2 Power Button Sign (PBS)

The MRIs of 57 subjects (including 19 normal controls, 38 epileptic patients) were selected from the study detailed in (Mellerio et al., 2014). The PBS was visually labeled for each individual hemisphere according to the criteria described in the same study. In addition to the identified PBSs, some hemispheres have been labeled as having an intermediate pattern. Intermediate patterns are considered as PBSs in this study. Among the 114 hemispheres in the database, 28 have the PBS (including 13 intermediate patterns). The central sulcus, its ascending branch and the pre-central sulcus forming the PBS were manually labeled.

4.3 Method

4.3.1 Fold representation

The BrainVISA/Morphologist pipeline, used to visualize the sulci in 3D and manually label the patterns, is also used for data preprocessing. This pipeline represents the folds by a set of voxels corresponding to a skeleton of the cerebrospinal fluid filling the fold. The representation of the folds therefore corresponds to a negative mold of the brain. In addition to facilitating manual labeling of sulcal patterns, the skeleton representation enables to

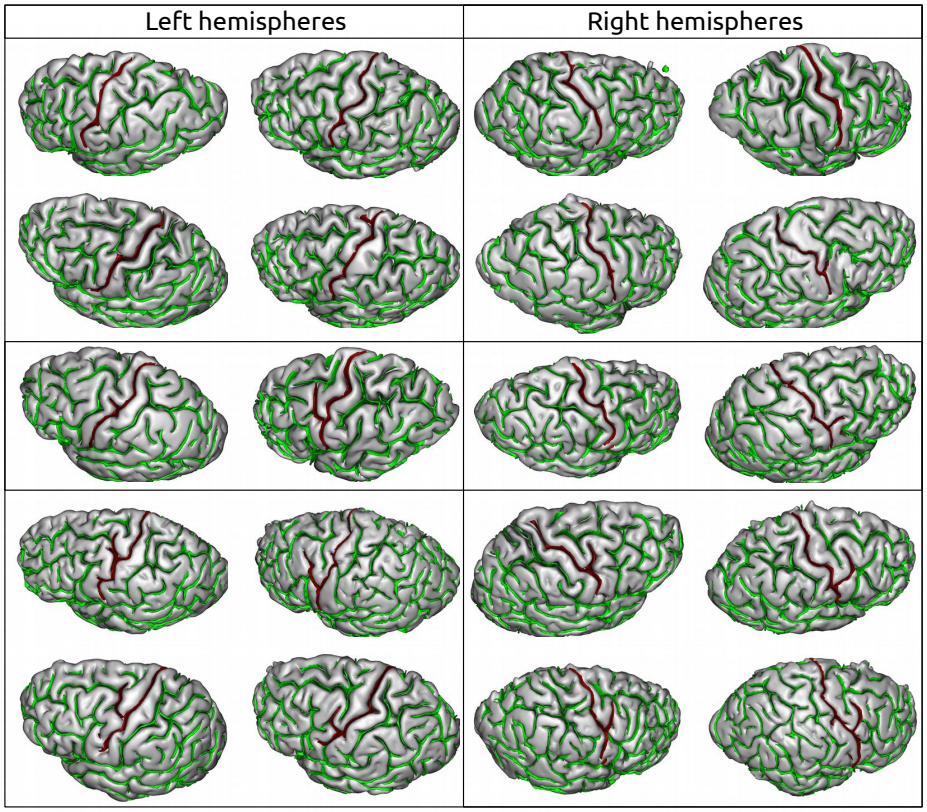


Figure 4.4 – Illustration of PBS variability. The first two lines represent hemispheres without PBS and the last two lines with PBS. The middle line represents hemispheres with an intermediate pattern. In this study these intermediate patterns are considered as PBS.

normalize the data optimally. Indeed, the data are particularly influenced by the type of MRI sequence, the age of the subject (which has a significant impact on the opening of the sulci) or even pathologies. The fold skeletons are affinely registered in the MNI space (Collins et al., 1994). In order to superimpose the right and left hemispheres, the right hemispheres are flipped.

With regard to ACC patterns, the single and double parallel types are respectively characterized by the absence and presence of the paracingulate sulcus. Thus, the objective is to detect the presence of the paracingulate to perform the classification. Similarly, for PBS, the algorithm must detect its presence. Paracingulate sulcus and PBS will be considered as the "searched patterns". In both databases, the searched patterns were extracted from manual labeling, as described in Figure 4.5.

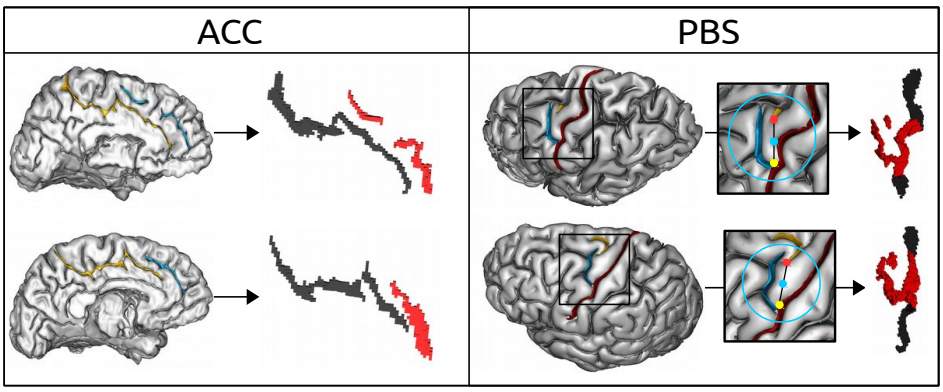


Figure 4.5 – Extraction of searched patterns from manual labeling. The voxels extracted from the manual labels represent the searched pattern in red and the sulcus to which it is attached in black. **Left:** Paracingulate sulcus, labeled in blue, corresponds to red voxels while cingulate sulcus, labeled in yellow, corresponds to black voxels. **Right:** The central sulcus is labeled in red, the ascending branch of the central sulcus in blue and the pre-central sulcus, wedged between the blue and red sulci, in yellow. As these manual labels are insufficient to properly delimit the PBS, the extraction of red voxels requires additional processing. 1. The yellow point corresponds to the intersection between the blue and red sulci. 2. The red point is the point of the yellow sulcus closest to the yellow point. 3. The blue point is in the centre of the red and yellow points. 4. The radius of the blue sphere, with the blue point as its centre, is calculated to be at least 20mm and to ensure that all the whole blue sulcus is contained in the sphere. 5. All the voxels contained in the blue sphere and belonging to the red/yellow/blue sulci define the red voxels. The black voxels correspond to the central sulcus voxels that do not belong to the red voxels.

4.3.2 Classification approaches

4.3.2.1 Support Vector Machine (SVM)

The idea of this first approach is to use the distances from the searched patterns of the labeled database to the image to be labeled to train a SVM classifier (Figure 4.6).

Registration: In order to realign the searched patterns to the hemisphere to be classified, the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992), with the robust implementation of Holz et al. (2015), is used. As in our previous study investigating the link between handedness and the shape of the central sulcus (Sun et al., 2012), the voxels of the fold skeleton, extracted by the BrainVISA/Morphologist pipeline, are used as

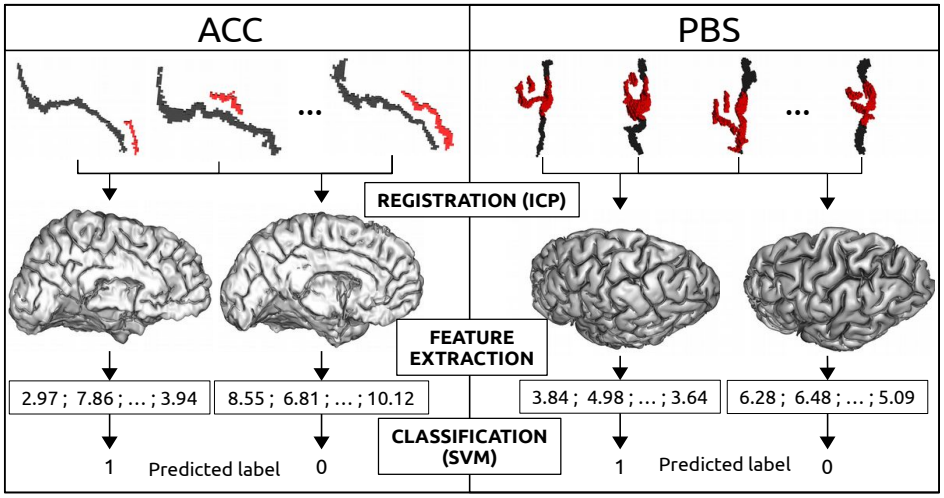


Figure 4.6 – Illustration of the SVM-based method. For each hemisphere of the training database with a paracingulate sulcus/PBS, the searched patterns (in red) and their attached sulcus (in black) are registered on the hemispheres to be classified using the Iterative Closest Point (ICP) algorithm. Once they have been registered, the average distance (from the points of the searched pattern in red to their nearest points in the hemisphere to be classified) is used as a feature for the SVM classifier. Each hemisphere of the training database with the searched pattern is therefore a feature of the SVM classifier.

point clouds to perform the registration.

In order to constrain the registration of the searched patterns, they are registered with their attached sulcus (i. e. cingulate sulcus for paracingulate sulcus, and central sulcus for PBS). This constraint with large sulci increase the specificity of the pattern and prevents the registration of the searched patterns to be trapped in a spurious local minima.

Since the ICP algorithm is particularly sensitive to the initial positions of the point clouds to be registered, several initializations are performed in order to avoid local minima. From the position of the point clouds in the MNI space (Collins et al., 1994), the pattern to be registered and its attached sulcus are translated before applying the ICP algorithm. The amplitude of the translations tested is determined by inner cross-validation. The registration minimizing the distance from the pattern and its attached sulcus is selected.

Feature extraction: Once registered, the distances between the searched patterns and the hemisphere to be classified are calculated. This distance corresponds to the average Euclidean distance from the voxels of the searched

pattern to their nearest voxel in the new hemisphere. The distance is low if the hemisphere to be classified also has the searched pattern.

Classifier: The calculated distances are used to train a Support Vector Machine (SVM) algorithm to classify hemispheres with or without the searched pattern. An ordered vector containing the distances to the hemisphere to be classified is used to train the classifier: each feature corresponds to a searched pattern in the training database. The Radial Basis Function is used as a kernel. Penalty parameter C of the error term and kernel coefficient γ are determined by inner cross validation.

4.3.2.2 Scoring by Non-local Image Patch Estimator (SNIPE)

The second approach is based on the Scoring by Non-local Image Patch Estimator (SNIPE) proposed in (Coupé et al., 2012). In order to classify images, SNIPE uses a grading measure based on a nonlocal patch-based framework. This measure estimates the similarity of the patches surrounding the voxels in the image under study with all the patches present in different training populations. In this study, the training database contains two "populations": hemispheres with the searched pattern and those without it (Figure 4.7).

Region Of Interest (ROI) delimitation: For this approach, the hemispheres to be classified are represented by binary volumes of 2*2*2mm resolution, containing the fold skeleton registered in the MNI space. A masking is performed on these volumes in order to extract the ROI. The mask is calculated from the location in the MNI space of the searched patterns in the training database. All voxels within 1cm of one of the searched patterns in the training database belong to the ROI. Note that the calculation of the mask is based on the location of the searched patterns in the training database, so it is recalculated for each cross-validation fold.

Optimized PatchMatch (OPM) algorithm: Once the ROI has been extracted, patches with central voxels belonging to the fold skeleton are associated with similar patches in the training database. To do this, the Optimized Patch Match (OPM) algorithm (Ta et al., 2014; Giraud et al., 2016) is used. This algorithm searches for similar patches in an atlas library using a cooperative and random strategy resulting in a very low computational burden. Compared to the PatchMatch algorithm (Barnes et al., 2009) from which it is inspired, OPM is adapted to 3D anatomical atlases by taking into account the rough alignment of images.

Here, as only patches with the central voxel belonging to the fold skeleton are considered, an adapted version of the OPM algorithm has been implemented (see appendix for more details). In addition, to manage the unbal-

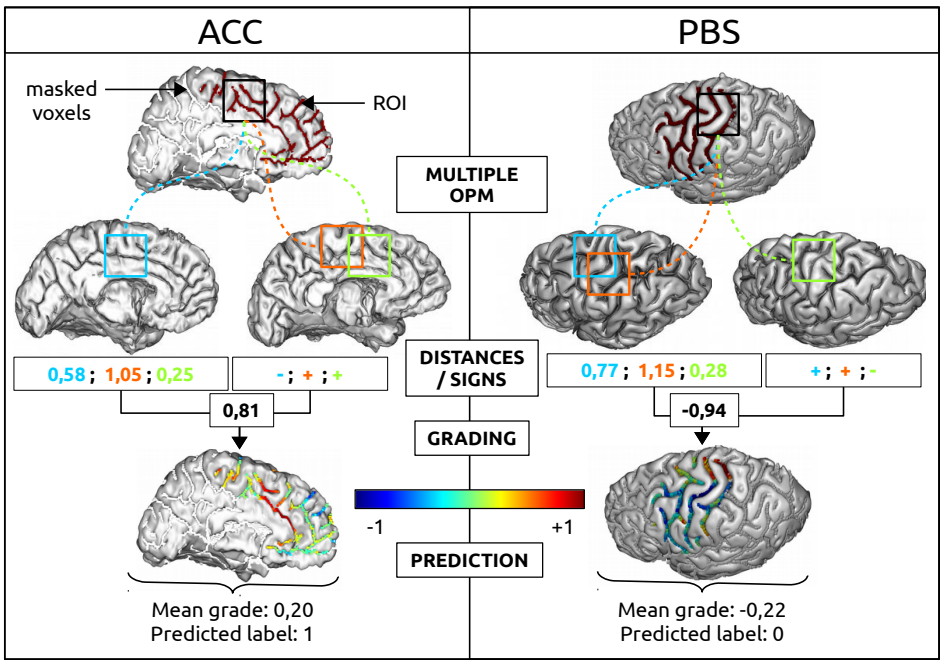


Figure 4.7 – Illustration of the SNIPE-based method. For each voxel in the Region Of Interest (ROI) of the hemisphere to be classified, its patch is associated with several similar patches (here 3) in the training database by using the OptimizedPatchMatch (OPM) algorithm. The distance of the similar patches to the patch and their location in a hemisphere with the searched pattern (+) or without it (-), are used to calculate the grade. All grades associated with ROI voxels are averaged to obtain the grade of the hemisphere. A positive grade predicts the presence of the searched pattern and vice versa for a negative grade.

anced presence of populations in the training database, under-represented hemispheres are proportionately more probable to be selected during the OPM initialization. Finally, since the compared patches are binarized volumes and not grayscale images, the distance to be minimized is different from that used in (Ta et al., 2014). This distance is the average Euclidean distance between the skeleton voxels of one patch and their nearest neighbour in the skeleton of the other patch. Please refer to the appendix for more details.

To obtain several similar patches, the OPM algorithm is run several times. The patch sizes and the number of similar patches selected are determined by inner cross-validation.

Grading calculation: As in (Coupé et al., 2012), in order to estimate the proximity of each voxel in the hemisphere to be classified to both populations, the selected similar patches are used to calculate the degree of proximity to one of the populations. To do so, for each voxel in the hemisphere to be classified, a robust weighted average of the distances of the patches selected to the patch surrounding the voxel is calculated based on the non-local average estimator proposed by Buades et al. (2005). In our case, a negative (respectively positive) classification value indicates that the neighbourhood surrounding the voxel is more specific to hemispheres without the searched pattern than with it. The ROI grades are then averaged to obtain the hemisphere grade.

In summary, several adaptations have been made compared to (Coupé et al., 2012):

- The search for similar patches is performed using the OPM algorithm and not using an exhaustive search.
- In order to manage the unbalanced databases, the initialization of the OPM algorithm is weighted to compensate.
- The average of the grades is done on the ROI instead of on a simultaneously segmented region. In fact, since the searched patterns are often missing, it is impossible to define an area to be segmented for the hemispheres without the searched pattern.
- Only patches with a central voxel belonging to the fold skeleton are considered. By doing so, it limits the number of patch matches that require optimization as the voxels belonging to the skeleton represent only a small part of the image's voxels. Moreover, since the patches are extracted from binarized images, the calculation of the distance between two patches can be successful only if the patches contain a minimum number of skeleton voxels.

4.3.2.3 Convolutional Neural Network (CNN)

The last proposed approach uses a 3D Convolutional Neural Network (CNN) (Figure 5.12).

Preprocessing: The input data of the neural network corresponds to a 3D volume of resolution 2*2*2mm containing the fold skeleton registered in the MNI space. Voxels that belong to the fold skeleton are 1; the others are 0. The volume size is calculated to contain all the searched patterns in the training database. Note that unlike the SVM-based method where a mask was calculated from the position of the searched patterns to determine the

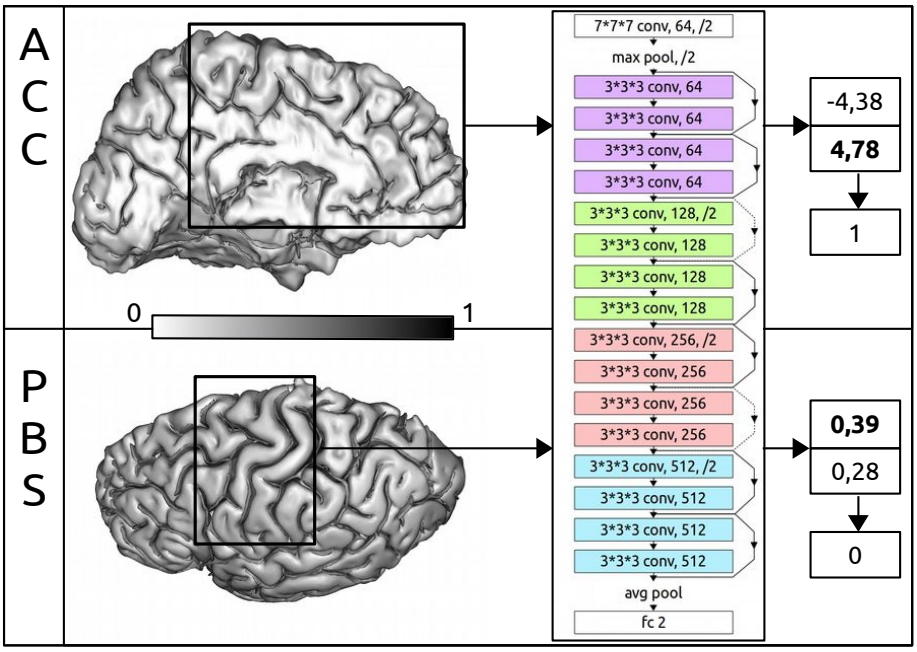


Figure 4.8 – Illustration of the CNN-based method. The binary volumes of resolution 2*2*2mm containing the fold skeleton, registered in the MNI space, are given as inputs to the neural network. The volume sizes are calculated to contain all the searched patterns of the training database. The neural network used is an 18-layer 3D ResNet. Two values are obtained at the output, one for each class. The highest value determines the predicted label.

ROI, here it is a parallelepipedic bounding box that is calculated for each fold of the cross validation.

In order to augment the training database, a rotation in a random direction with a random angle (following a Gaussian distribution $\mathcal{N}(0, \frac{\pi}{40}^2)$) and a random translation (each coordinate following a Gaussian distribution $\mathcal{N}(0, 2^2)$) are applied successively to the volumes at each epoch.

Network architecture: Concerning the classification of 2D images, spectacular progress has been made in the last decade. This is particularly well illustrated by the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which saw the emergence of several network architectures that are now widely used for image classification. However, for the classification of 3D images, there is currently no such challenge to assess the advantages and disadvantages of several architectures. In the study (Hara et al., 2018), the authors evaluate the performance of adaptations to 3D spatiotemporal images of neural networks used for ILSVRC 2D image classification. Based

on this study, the architecture of the 3D ResNet was chosen for our problem. The parameters of the architecture used correspond to those of the initially proposed 2D architecture (He et al., 2016), with 3D convolutional layers, a single input channel and only 18 layers. The use of deeper networks did not appear to improve the results for this study and was therefore not tested. In order to be able to manage variable input volume size (since it depends on the folds of the cross-validation), the adaptive average pooling proposed by the Pytorch library (Paszke et al., 2017) is used.

Training: The weight initialization of the convolutional layers follows the method described in (He et al., 2015), using a normal distribution. The stochastic gradient descent with Nesterov momentum (Sutskever et al., 2013) is used for training, with a batch size of 10 samples and a learning rate and momentum determined by an 3-folds inner cross validation. The loss function corresponds to the cross entropy loss, weighted by the inverse of the number of samples of each class in the training database. The learning rate is halved when the loss function has not improved for five consecutive epochs. In accordance with the early stopping strategy, the training is stopped after ten consecutive epochs without improvement. The selected neural network corresponds to the epoch at which the balanced accuracy is the highest.

Ideally, in order to perform early stopping, part of the hemispheres in the training set should be reserved to evaluate the performance of the neural network at each epoch, and train on the others. In this study, in order to maintain a training set of sufficient size during 3-folds inner cross-validation, the validation fold is used to test the model at each epoch and the 2 others folds are used for training.

4.3.3 Performance evaluation

The evaluation of the performance of the trained model was estimated using a double loop of Cross Validation (CV): one inner 3-folds CV loop to select the best hyper-parameters and one outer 10-folds CV loop to evaluate classification performances. Note that the strategies used to set the method hyperparameters are detailed in the appendix. In order to define the folds of the two cross-validations, special attention was paid to ensure that two hemispheres of the same subject were not separated, one in the training set and the other in the test set. In addition, the hemispheres were distributed so that each fold contained approximately the same proportion of searched patterns.

Balanced accuracy is used to evaluate the performance of each model (Brodersen et al., 2010). This score is defined as the average of recall R_c obtained on each class c:

$$R_c = \frac{TP_c}{TP_c + FN_c} \tag{4.1}$$

with TP_c and FN_c, respectively the number of true positive and false negative samples.

In order to compare the models in pairs, a paired T-test was performed between the balanced accuracy lists of each CV fold. If the p_{value} is less than 0.05, the models are considered significantly different.

In order to determine whether the average classification rate obtained is significant or not, it is compared with chance, i.e. the classification obtained with a Binomial distribution of success probability 0.5 and with a number of samples corresponding to the number of hemispheres to be classified in the database. Since this is an exact, two-sided test of the null hypothesis, the p_{value} can be divided by 2 as we test that the accuracy is superior to the chance level.

Finally, a S score can be calculated for each model to assess its confidence in the classification of a given hemisphere. For the SVM-based model, the probabilities of presence/absence of a searched pattern is calculated according to the procedure proposed in (Platt et al., 1999). The proposed S score is the difference between these two probabilities. For the SNIPE-based model, the value of the estimator is directly used as the S score. For the CNN-based model, two scores are obtained at the output of the network, one for each pattern. The S score used corresponds to the difference between these two outputs. In order to determine whether the S score significantly distinguishes fold patterns, the Mann-Whitney test (Mann and Whitney, 1947) is used. This test is complementary with the paired T-test comparing the balanced accuracies obtained during the cross validation because the latter does not take into account the confidence of the model in its classification.

4.4 Results

4.4.1 Which is the best model?

When analysing the balanced accuracy by model (Figure 4.9), we observe that whatever the fold pattern studied, all models are better than chance and that none of them is significantly better than the other ones. Thus, the balanced accuracy of the ACC patterns classification is about 80% and that of the PBS about 60%. Models therefore have more difficulty managing rare patterns where the training database is limited and unbalanced.

Regarding the distribution of the S score (Figure 4.10) for PBS recognition, the Mann-Whitney test indicates that, except for the CNN-based model, the S scores significantly separate the hemispheres with and without a searched pattern. Thus, the CNN-based model does not provide sufficiently reliable scores to assist in the search for PBSs, while the other two

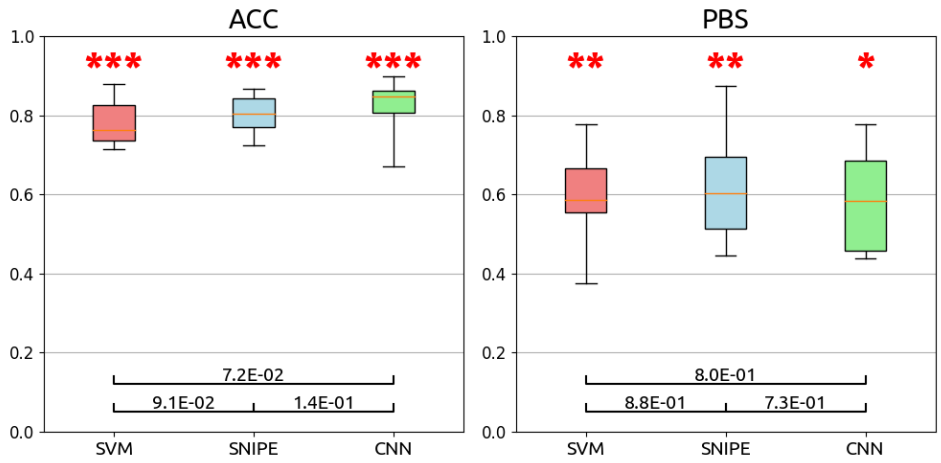


Figure 4.9 – Balanced accuracies for the 10 folds of the cross-validation by model. The SVM-based model is in red, the SNIPE-based model in blue and the CNN-based model in green. The values of the balanced accuracies obtained for each fold of the cross-validation were compared by matched T-test. The p_{values} obtained are indicated below the compared models. The red stars above the models indicate that the model is better than chance (one star for $p_{value} < 0.05$, two for $p_{value} < 0.02$, three for $p_{value} < 0.001$). The box extends from the lower to upper quartile balanced accuracy values, with a line at the median. The whiskers extend from the box to show the minimum and maximum values.

models can be used to find PBSs on a new database and enrich the current training database. This application has been tested on the Human Connectome Project (HCP) database.

4.4.2 Looking for PBSs

The three models proposed were trained on the whole training database by fixing the hyperparameters thanks to a 3-folds cross validation. They were then used to label the HCP database, containing 1023 subjects, i. e. 2046 hemispheres. In order to see if these methods can identify some PBSs on this new database, the 9 hemispheres with the highest S score are examined (Figure 4.11, 4.12 and 4.13). Note that we do not know the proportion of PBSs in the healthy population, so it is difficult to estimate how many PBSs are actually present in the HCP database. However, the three models proposed make it possible to find some of them, which is already promising, in particular for SVM and SNIPE-based models that allow to find more PBSs than the CNN-based model. Indeed, the first two models identify three PBSs out of the nine hemispheres examined, while the CNN-based

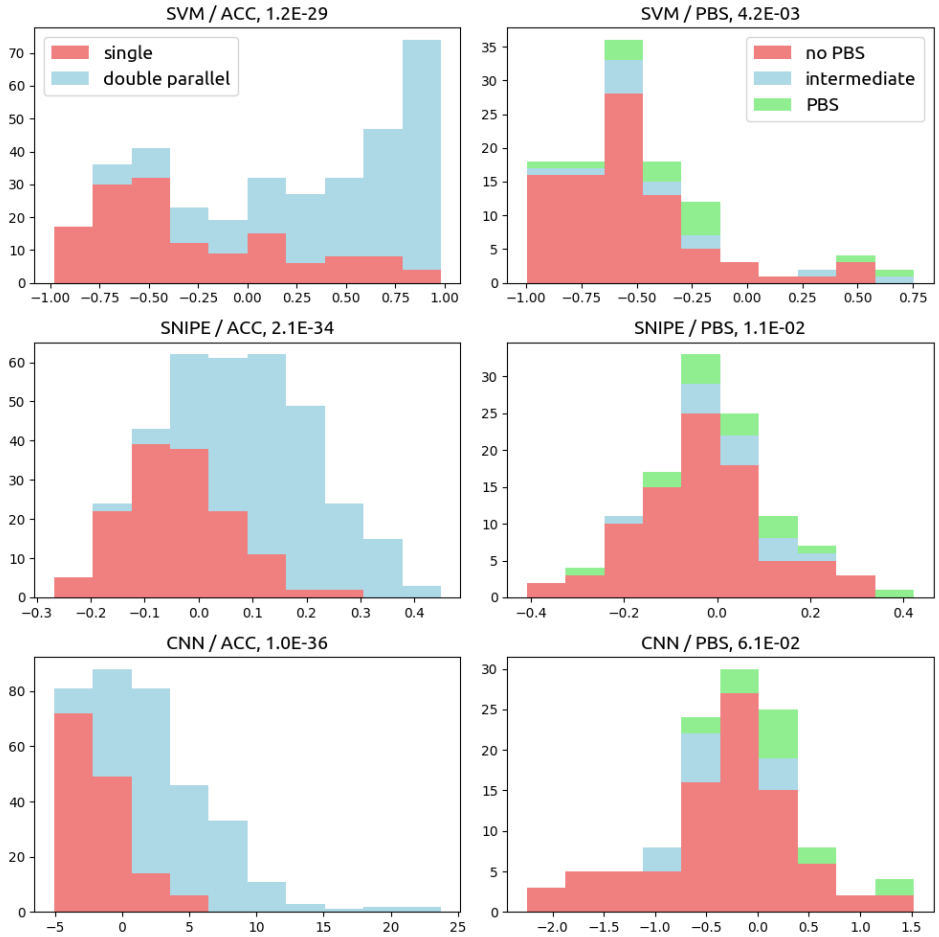


Figure 4.10 – S score per model. The proportions of hemispheres with or without the searched pattern are represented as a function of the S score value. The proportions are stacked. The p_{value} of the Mann-Whitney test are indicated in the title of each graph. Note that only the p_{value} of the CNN-based model for PBS recognition is less than 0.05.

model identifies only one. Note that the three PBSs identified by the SVM-based and SNIPE-based models are different. These methods therefore seem to be complementary for PBSs research.

4.5 Conclusion

In this study, three methods were proposed for the automatic classification of sulcal patterns. The first method uses the distances of the patterns of the training database to the hemisphere to be classified to train a SVM. The second method is based on the SNIPE patch classification approach

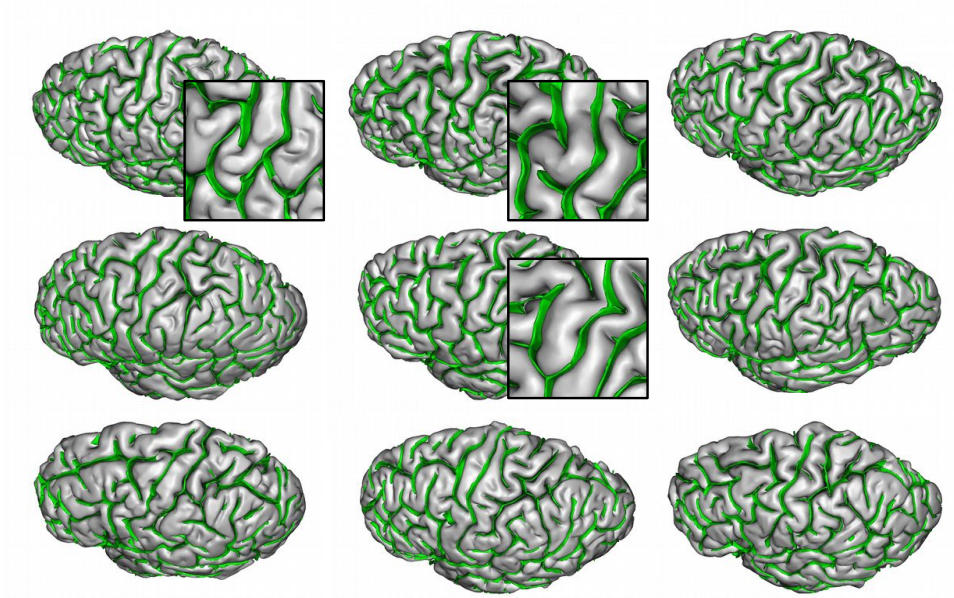


Figure 4.11 – Looking for PBSs with the SVM-based method. The hemispheres presented are those with the highest S scores on the HCP database (score decreasing in the reading direction). Hemispheres with a PBS, or an intermediate pattern, are zoomed. The right hemispheres are flipped for this image.

proposed by Coupé et al. (2012). The third method uses an 18-layer 3D recurrent neural network.

For the recognition of ACC patterns, the three proposed models have equivalent labeling performance, with a balanced accuracy of about 80%. However, the fastest model to apply is by far the one based on CNN. For PBS recognition, all three models also have equivalent labeling performance, with a balanced accuracy of about 60%. This score is still insufficient to carry out large-scale morphological studies. However, the models provide scores to assess their confidence in the proposed labeling. Except for the CNN-based model, these scores significantly distinguish between hemispheres with and without PBS. Thus, the use of these scores facilitates the enrichment of the training database by searching for PBSs on a new database, which should lead to an improvement in the classification scores of current models.

In conclusion, the proposed models are promising and have the potential to extend and confirm morphological studies on larger databases in order to better understand the links between sulcal shapes and functional organization of the cortex. In fact, even if they have difficulty managing unbalanced training databases, they can still be used to enrich the training database at a lower cost.

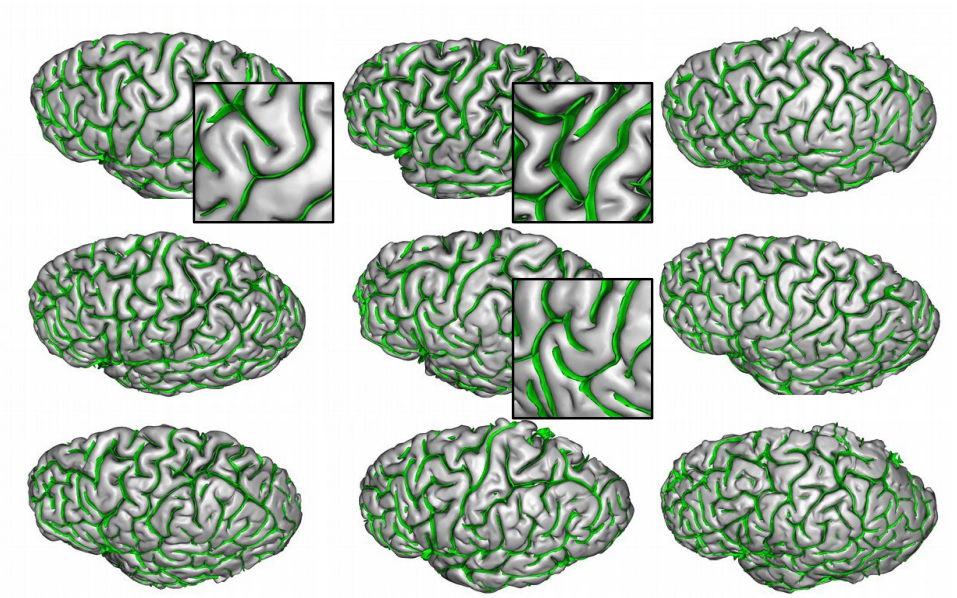


Figure 4.12 – Looking for PBSs with the SNIPE-based method.

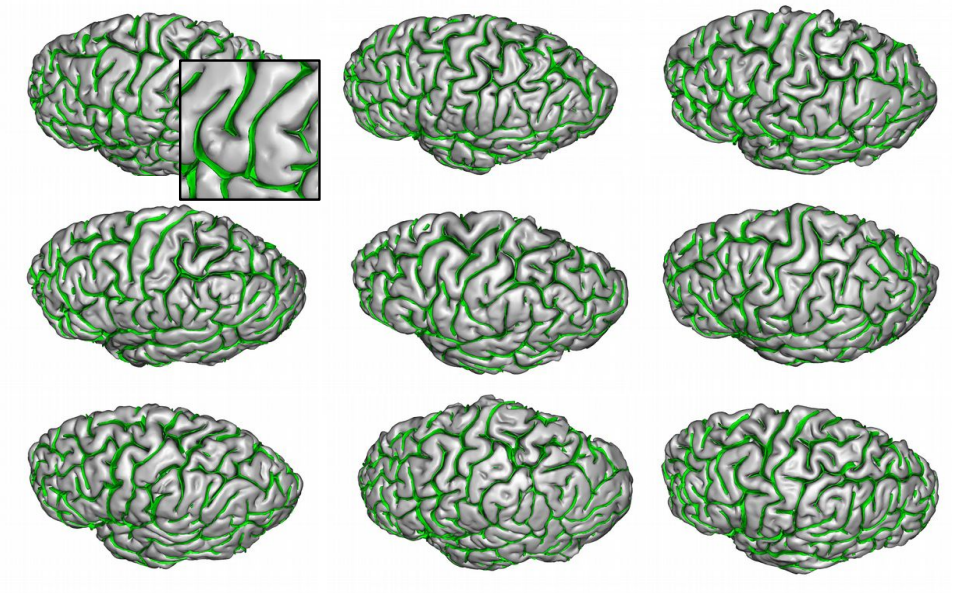


Figure 4.13 – Looking for PBSs with the CNN-based method.

In order to improve the current performance of the proposed models, one possible approach would be to train a 2D neural network from the surface projections of the folds on a 2D image representation. Indeed, 2D networks are easier to train than 3D networks, thanks to their reduced number of parameters to learn. The difficulty of such an approach would be to determine an optimal way to project the folds on a surface, without excessively distorting the patterns. Such projections have recently been proposed for sulci recognition using a 2D CNN (Parvathaneni et al., 2019). Once this step is completed, it opens the way to many promising neural network architectures that have already proven their worth in 2D image classification, such as for the ImageNet challenge (Russakovsky et al., 2015).

However, classification of local patterns is sometimes difficult to carry out because of the continuity between the different configurations. For this reason, in the study of ACC patterns, some subjects were excluded from the study due to a lack of consensus on their labeling. Similarly, different levels of PBS presence were used when labeling them. Also, some approaches prefer to use manifold based techniques to better represent the continuity between two extreme configurations (Sun et al., 2012, 2016). Nevertheless, these approaches are based on the prior labeling of the structures of interest, which is difficult to obtain for patterns absent in at least half of the population and as difficult to delimit as the paracingular sulcus and the PBS. Similarly, however, it would be interesting to predict a score for the presence of these patterns rather than attempt to distinguish them into separate groups.

Appendix

OPM algorithm adaptation

In this section, I focus only on the adaptations made in comparison to the Optimized Patch-Match (OPM) algorithm presented in (Ta et al., 2014). These adaptations are due to the fact that only patches whose central voxel belongs to the fold skeleton are considered. The information presented here is insufficient to understand how OPM works. Please refer to (Ta et al., 2014; Giraud et al., 2016) for more information.

The OPM algorithm itself is an adaptation of the PatchMatch (PM) algorithm. The PM algorithm aims to assign to each patch of an image A, a patch similar to it in an image B. For this, it is based on the fact that given a patch $p_A(x, y)$ in image A that matches well with a patch $p_B(x', y')$ in image B, then there is a good chance that the patches adjacent to $p_A(x, y)$ match well with the patches adjacent to $p_B(x', y')$. The OPM algorithm allows to adapt this algorithm to the segmentation of 3D anatomical images by no longer searching for the correspondences of the patches of image A with an image B but with an atlas library and by constraining the search for patches, thanks to a rough alignment between the anatomical images.

OPM first includes a constrained initialization, followed by a propagation and a random search step performed iteratively to improve patch matching. In our case, propagation and random search steps are done three times.

- **Constrained initialization:** During constrained initialization, for each voxel of the skeleton to be labeled, we want to randomly associate a patch of the database whose central voxel is located around its position in MNI space. To do so, an atlas is randomly selected in the atlas library. Then, a voxel is randomly selected from the skeleton voxels of the selected atlas included in a window of 14 mm side drawn around the position of the voxel to be labeled. However, a problem arises if no sulcus is present in this area. To remedy this, the search window is increased by one millimeter until at least one voxel belonging to a sulcus is found.

- **Propagation step:** During the propagation step, we try to improve the patch matches of each skeleton voxel by testing whether the at-

lases associated with its adjacent voxels provide better matches (Figure 4.14). Let s_A be a voxel to be labeled of the S_A skeleton and s'_A one of its adjacent voxels, associated with the s'_B voxel of the S_B skeleton. We then look for the voxel s_B closest to the position adjacent to s'_B where the voxel s_A would be located relatively to s'_A. The patch around s_B is then compared to the patch around s_A to see if it allows a better match.

- **Constrained random search:** During the constrained random search, we randomly select a skeleton voxel in the matched image within a decreasing window (42mm, 22mm, 10mm, 6mm) to see if it provides a better match.

Since the compared patches are binarized volumes and not grayscale images, the distance to be minimized is different from that used in (Ta et al., 2014). The following distance d between two patches $P(S_A)$ and $P(S_B)$, respectively belonging to the fold skeletons S_A and S_B (superimposed by a simple translation) is used:

$$d(P(S_A), P(S_B)) = \frac{d(P(S_A) \to S_B) + d(P(S_B) \to S_A)}{2} \tag{4.2}$$

The measurement from a patch $P(S_A)$ to a fold skeleton S_B corresponds to the average of quadratic Euclidean distances d_E of the skeleton voxels $p_A \in P(S_A)$ and their nearest neighbor in the fold skeleton S_B (Figure 4.15):

$$d(P(S_A) \to S_B) = \frac{1}{|P(S_A)|} \sum_{p_A \in P(S_A)} \min_{p_B \in S_B} [d_E^2(p_A, p_B)] \tag{4.3}$$

Note that, in order to avoid border effects, the closest neighbor of p_A is searched in the entire skeleton S_B and not only among the skeleton voxels contained in the patch $P(S_B)$.

Hyperparameters

In order to set the hyperparameters, a 3-folds inner cross validation is performed. The procedures for selecting hyperparameter values for each method are described below.

SVM-based model Three hyperparameters are fixed by grid search: the translations used to optimize the registration and the γ/C parameters of the SVM classifier corresponding respectively to the kernel coefficient and the penalty parameter of the error term.

Concerning the first hyperparameter, several translations are indeed applied to the patterns to be registered from their position in the MNI space.

Each of these translations allows to test a new initialization before applying the ICP algorithm. Three translation amplitudes are considered, $A = [5]$, $A = [10]$ and $A = [20]$, as well as the combinations $A = [5, 10]$ and $A = [10, 20]$. All translations $[t_x, t_y, t_z]$ with $(t_x, t_y, t_z) \in (-A \cup [0] \cup A)^3$ are tested. The A amplitudes leading to the best results in inner cross-validation are selected. Concerning the classifier parameters, the values [0.0001, 0.001, 0.01, 0.1] are tested for the γ parameter and [0.1, 1, 10, 100, 1000] for the C parameter.

In Figure 4.16, we observe that the balanced accuracies obtained for each hyperparameter are relatively stable between the cross-validation folds for ACC pattern recognition. They are much more variable regarding the detection of PBS, which is probably due to the small and unbalanced training database used.

SNIPE-based model Three hyperparameters are estimated using grid search: number of approximate nearest neighbors selected (5, 10, 15, 20, 25, 30), patch sizes (18, 26 or 34mm) and their combinations (all combinations of 1, 2, and 3 different patch sizes are tested). Note that two OPAL parameters were set a priori: the size of the search window at 14*14*14 mm and the number of iterations of the propagation and random search steps at 4.

In Figure 4.17, we observe that the balanced accuracy increases globally with the number of approximate nearest neighbours. This trend is particularly marked for the recognition of ACC patterns. Regarding patch sizes, small 18 mm patches seem more appropriate for PBS detection, while it is the combination of 18 mm and 26 mm patches that allows better recognition of ACC patterns.

CNN-based model In order to choose the hyperparameters (learning rate and momentum), several cross-validation loops are performed in turn to adjust the hyperparameters one after the other. First, the momentum is set at 0.9 and the learning rates 1e-2, 1e-3 and 1e-4 are tested in turn. Second, once the learning rates have been tested on all inner cross-validation folds, the learning rate is refined around the value lr that obtained the best E_{SI}: the values $lr/4$, $lr/2$, $lr * 2$ and $lr * 4$ are tested in turn. Finally, the best learning rate obtained is then tested with momentum 0.8, 0.7 and 0.6 to select the best value. We have chosen not to test values above 0.9 in order to reduce the risk of divergence due to excessive gradient memory.

In Figure 4.18, we observe that the values of the selected hyperparameters are quite variable, especially for momentum. In order to limit this variability, which probably impacts the results obtained, it would be necessary to test with a larger training database.

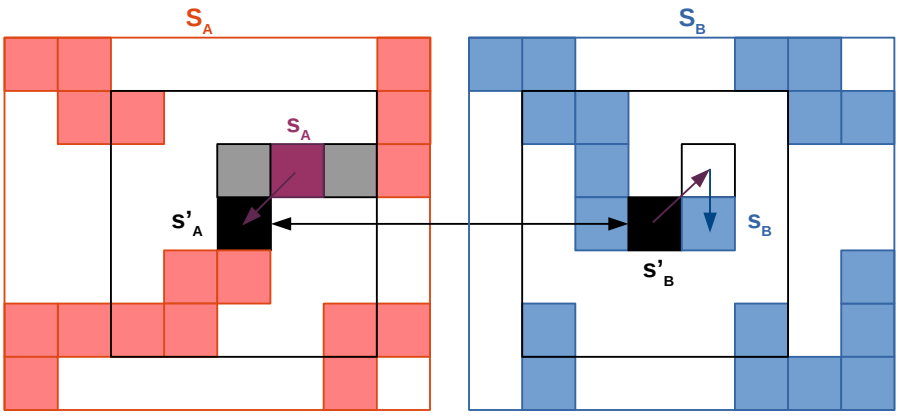


Figure 4.14 – Propagation step of the OPM algorithm adapted to our application. This step concerns the purple voxel s_A of the fold skeleton S_A. Its adjacent voxels are selected (grey or black voxels). Each of them is associated with a patch of one of the fold skeletons in the atlas library. For example, the black voxel s'_A is associated with the patch around the voxel s'_B of the skeleton S_B. In order to find the voxel s_B, whose patch will be compared to the s_A patch, we first search for the position adjacent to s'_B where the voxel s_A would be located relatively to s'_A (purple arrow). Then, we search for the voxel closest to this position (blue arrow).

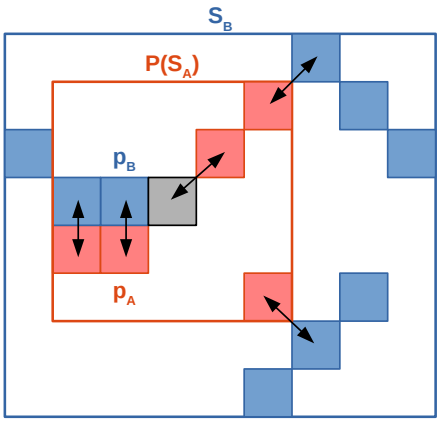


Figure 4.15 – Calculation of the distance from the patch $P(S_A)$ to the patch $P(S_B)$. The grey voxel represents the central voxel of the patch $P(S_A)$ which is superposed with a voxel of the skeleton S_B. For each voxel $p_A \in P(S_A)$, we look for its closest neighbor among the voxels of the skeleton S_B. The Euclidean distance between these two voxels is calculated. The distances over all the points $p_A \in P(S_A)$ and their nearest neighbors are then averaged to obtain $d(P(S_A) \rightarrow S_B)$.

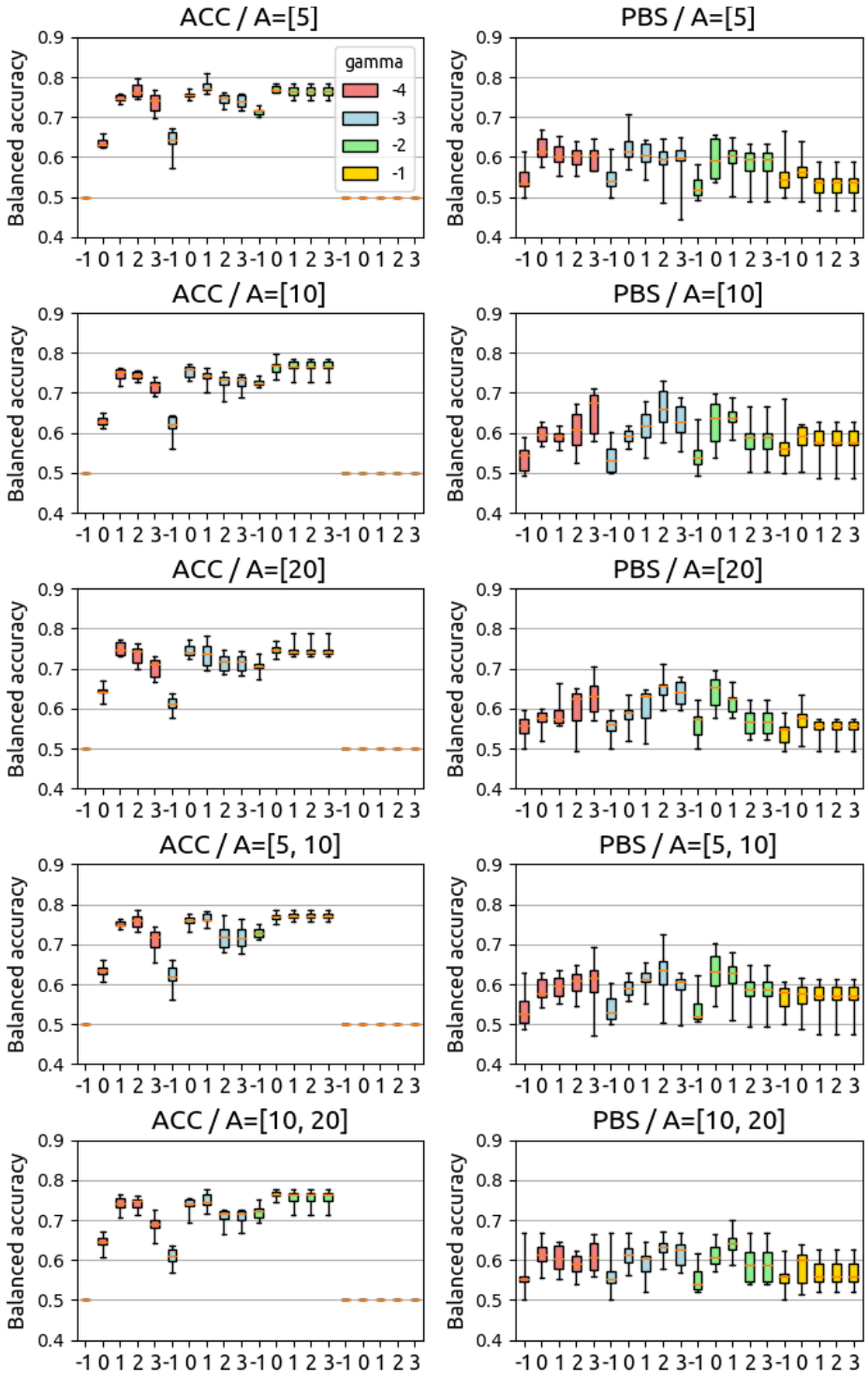


Figure 4.16 – Hyperparameters values for the SVM-based model. The graphs on the left correspond to the recognition of ACC patterns and the graphs on the right correspond to that of the PBS. Each graph corresponds to new amplitudes A of the translations tested for ICP initialization. The amplitudes A are indicated in the graph title. Each color corresponds to a different γ value (see legend on the first graph), while on the abscissa is the C values tested. Note that for the last two hyperparameters, the values indicated correspond to the power of 10 of the true tested value in order to facilitate visualization.

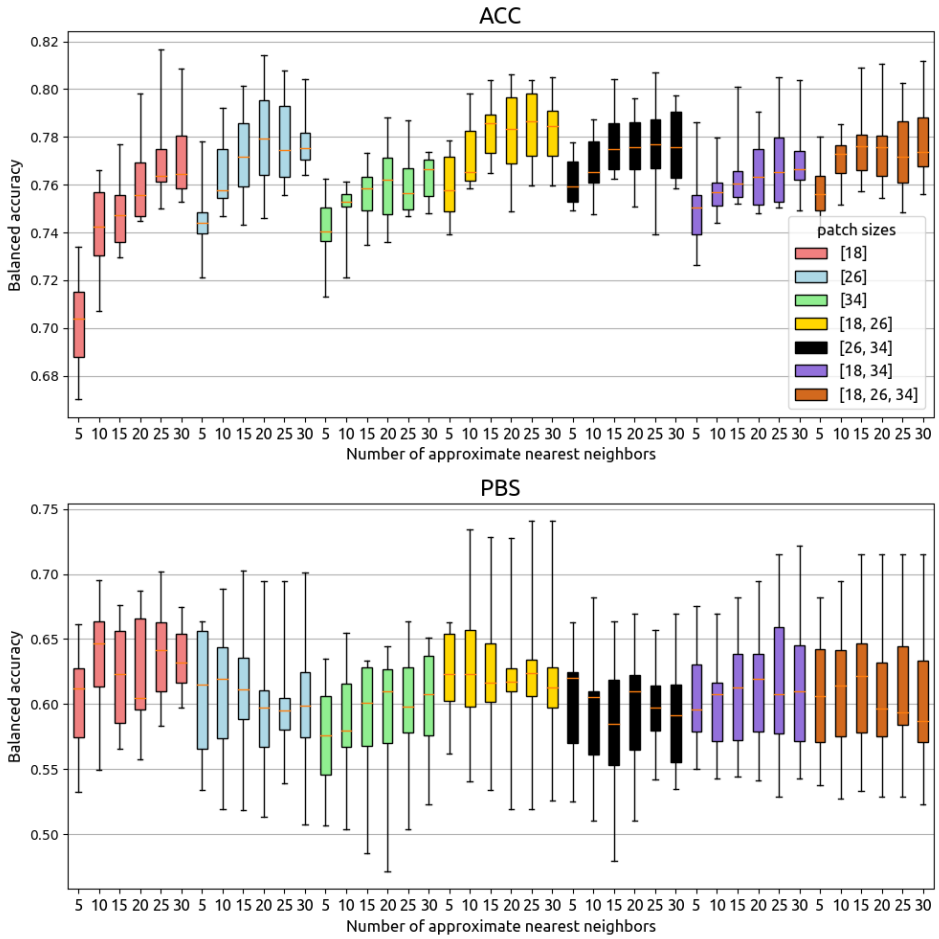


Figure 4.17 – Hyperparameters values for SNIPE-based model. The upper graph corresponds to the recognition of ACC patterns while the lower graph corresponds to the PBS recognition. During inner cross-validation, the hyperparameters are set using a grid search to select values that maximize the average balanced accuracy. The number of approximate nearest neighbors was set between 5 and 30. The patch sizes tested were 18, 26 and 34mm. All possible combinations of these patch sizes are represented in a different color (see the legend at the top graph). The box extends from the lower to upper quartile error values, with a line at the median. The whiskers extend from the box to show the minimum and maximum values.

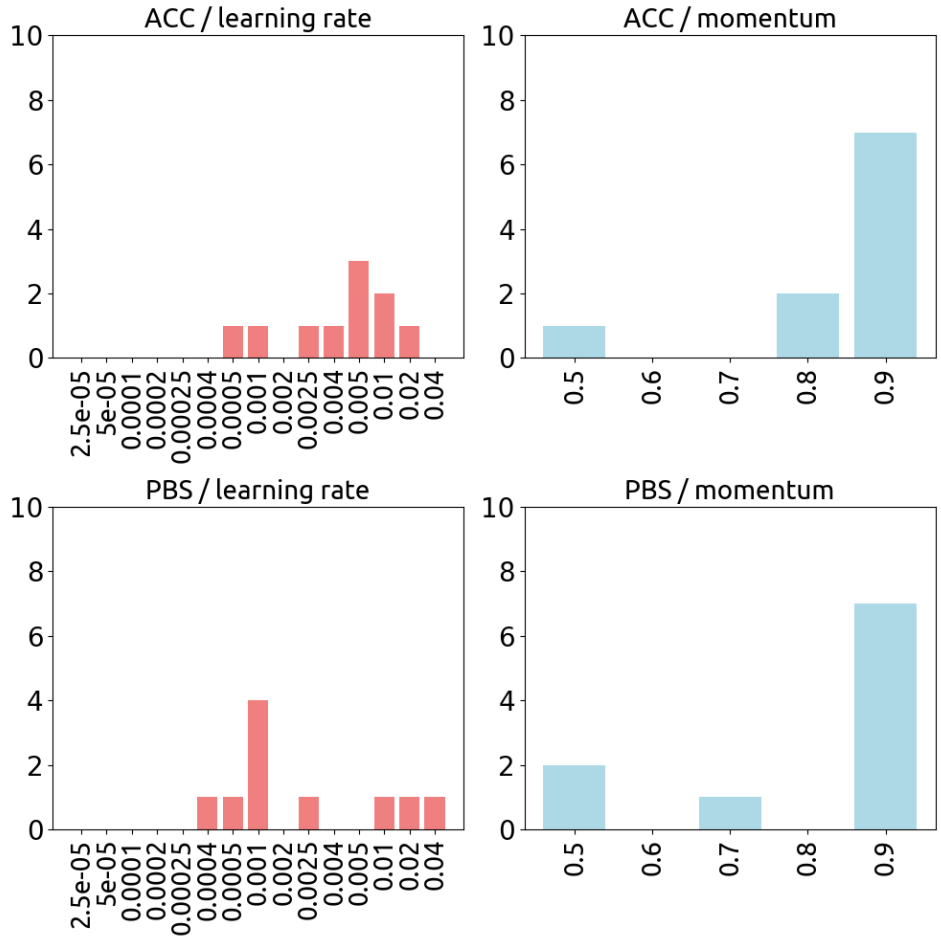


Figure 4.18 – Hyperparameters values for CNN-based model. The learning rate is represented in red and the momentum in blue. Each graph represents the number of times each tested value is selected. As an external cross validation of 10 folds is done, the values can be selected up to 10 times.

Chapter 5

Automatic labeling of cortical sulci

Abstract

The extreme variability of the folding pattern of the human cortex makes the recognition of cortical sulci, both automatic and manual, particularly challenging. Reliable identification of the human cortical sulci in its entirety, is extremely difficult and is practiced by only a few experts. Moreover, these sulci correspond to more than a hundred different structures, which makes manual labeling long and fastidious and therefore limits access to large labeled databases to train machine learning. Here, I compare two novel approaches to address this issue: patch-based Multi-Atlas Segmentation (MAS) techniques and Convolutional Neural Network (CNN)-based approaches. Both are currently applied for anatomical segmentations because they embed much better representations of inter-subject variability than approaches based on a single template atlas. However, these methods typically focus on voxel-wise labeling, disregarding certain geometrical and topological properties of interest for sulcus morphometry. Therefore, I propose to refine these approaches with domain specific bottom-up geometric constraints provided by the Morphologist toolbox, a widely used sulcus recognition toolbox included in the BrainVISA package. These constraints are utilized to provide a single sulcus label to each topologically elementary fold, the building blocks of the pattern recognition problem. To eliminate the shortcomings associated with the Morphologist's pre-segmentation into elementary folds, this regularization scheme is complemented using a top-down perspective which triggers an additional cleavage of the elementary folds when required. All the newly proposed models outperform the current BrainVISA/Morphologist model, the most efficient being a CNN U-Net-based approach which carries out sulcus recognition within a few seconds.

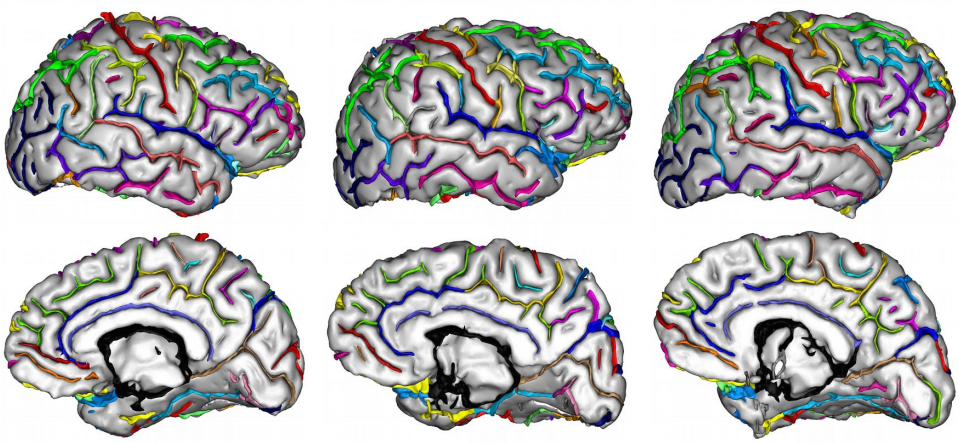


Figure 5.1 – Illustration of cortical folds variability. The manual labeling of the three right hemispheres represented here shows the variability of cortical sulci by their shape, size, and position.

The work presented in this chapter has been submitted to Medical Image Analysis as *Automatic labeling of cortical sulci using patch- or CNN-based segmentation techniques combined with bottom-up geometric constraints*, L. Borne, D. Rivière, M. Mancip, J.-F. Mangin.

5.1 Introduction

The surface of the brain is divided into many convolutions, called gyri, delimited by folds, called sulci. The main sulci are considered as the limits between functionally and architecturally different regions. Additionally, cortex morphometry is used to quantify brain development and degenerative diseases. Despite the many tools available for 3D visualization of sulci, sulci labeling is a long and fastidious process. It takes several hours for an expert to label all sulci in a single brain and reliable labeling requires the opinion of several experts. However, because of the large variability of the folding pattern in the general population, inferring developmental biomarkers requires the mining of data from a large number of brains. Therefore, automation of the sulcus recognition is essential.

Nevertheless, learning to label sulci is an extremely complex challenge for several reasons. First, as illustrated in Figure 5.1, sulci are highly variable structures, some sulci are even absent in more than 70% of brains. Additionally, each brain contains more than 120 different sulci and only a small number of segmentation algorithms are made for as many structures. Moreover, the average size of sulci can vary from a simple to a hundredfold, from one sulcus to another, thus, the problem is particularly unbalanced. Finally, the number of manually labeled subjects which can be used for supervised

learning is limited.

5.1.1 Overview of automatic sulci recognition methods

Algorithms dedicated to automatic sulci recognition are primarily based on graphical representations, which represents the relative positions of the sulci with respect to each other, as well as their position and their location in a standardized space (Royackkers et al., 1999; Riviere et al., 2002; Vivodtzev et al., 2006; Shi et al., 2007; Yang and Kruggel, 2009; Blida, 2014). To ensure their robust recognition, other methods have previously been experimented with to model inter-subject variability using several frameworks ranging from principal component analysis to Bayesian approaches (Lohmann and Von Cramon, 2000; Behnke et al., 2003; Fischl et al., 2004; Perrot et al., 2011). All of these methods are based on a segmentation algorithm followed by a classification algorithm, in which the sulci are first extracted, according to different representations, then labeled. In this thesis, the sulci are extracted by the BrainVISA/Morphologist pipeline and several labeling methods are compared. The new approaches presented are compared with the approach proposed in the BrainVISA/Morphologist package (Perrot et al., 2011), referred as the Statistical Probabilistic Anatomy Map (SPAM) approach in this thesis.

5.1.2 New approaches: MAS and CNN

Among the segmentation methods for biomedical applications, multi-atlas segmentation (MAS) and convolutional neural networks (CNNs) are among the most widely used today.

MAS techniques, initially introduced by Rohlfing et al. (2004) approximately fifteen years ago, use each manually segmented image as an atlas: the atlases are adjusted to the image to be segmented and the best matches are used to participate in the segmentation. Thus, MAS techniques make it possible to more accurately represent anatomical variability by not attempting to model a segmentation problem using an average model. These techniques are now widely used, but have a major disadvantage: the registration of the atlases to the images is particularly expensive.

Among the many variations of these techniques, the patch-based approach (Coupé et al., 2011; Rousseau et al., 2011) inspired by the non-local means method (Buades et al., 2005) have particularly attracted our attention. By using a patch-based search strategy to identify matches with the atlases, the image no longer needs to be aligned globally with all the atlases via expensive non-linear registration. Thus, the registration and selection of matching patches can be particularly accelerated thanks to the Optimized PatchMatch algorithm, (Ta et al., 2014; Giraud et al., 2016) based on the

PatchMatch algorithm (Barnes et al., 2009) adapted to segmentation of 3D images.

Inspired by these approaches, we propose two algorithms for cortical sulci recognition. The first is directly inspired by Romero et al. (2017a) where each patch is a cube in the image to be segmented. With regards to the second algorithm, we propose a new patch generation strategy based on a high level representation of the sulci, as the standard way of extracting patches does not seem capable optimally exploiting the sulci geometry and the relations between them, which we believe to be the discriminative features for their recognition. These two algorithms will be designated respectively by PMAS (for Patch-based MAS) and HPMAS (for Patch-based MAS with High level representation of the data).

The CNNs were initially developed to address problems in image classification and are now renowned for their formidable effectiveness in dealing with numerous computer vision problems. These techniques allow effective image analysis by learning an abstract representation of the image. Concerning segmentation problems, the first approach was proposed approximately ten years ago by Ciresan et al. (2012) where a neural network was trained to classify each voxel of the image to be segmented from its surrounding patch. Since then, new approaches allow the entire image segmentation using fully convolutional neural networks, such as the one initially proposed by Long et al. (2015) and dedicated to semantic segmentation. Concerning segmentation problems in medical imaging, the most commonly used architecture is the U-Net, which was initially proposed by Ronneberger et al. (2015) and whose adaptation to 3D images was proposed in (Çiçek et al., 2016; Milletari et al., 2016). Here, we propose to compare two approaches based on CNNs. The first is inspired by Ciresan et al. (2012), adapted to address problems associated with 3D imaging. The second uses the 3D U-Net architecture proposed in (Çiçek et al., 2016). These two approaches will be called PCNN (for Patch-based CNN) and UNET, respectively.

To the best of our knowledge, despite their current popularity, no MAS or CNN-based approach has yet been proposed for cortical sulci recognition. Note that these two approaches are generally used to segment the entire image while in this study only the presegmented folds need to be labeled, requiring several adjustments in the proposed models.

5.1.3 Bottom-up geometric constraints

There is no guarantee that the geometric definition of a sulcus, as a set of topologically simple surfaces, is respected in the case of MAS and CNN-based methods described above. This is particularly disadvantageous for morphometric studies whose measurements are based on the definition of sulci. To remedy this, the BrainVISA/Morphologist pipeline provides an

algorithm for bottom-up aggregation of voxels into elementary folds, which are the geometric building blocks of the problem. Once the voxels have been labeled by one of the methods proposed above, it is possible to regularize the results at the scale of the elementary folds. However, the upstream extraction of the elementary folds may sometimes be inaccurate. Although from the same MRI, vastly different fragmentations can be obtained because of stochastic optimizations embedded in the pipeline. This was previously a problem in the SPAM model proposed in (Perrot et al., 2011), which uses the same geometric entities to perform recognition, but is not capable of automatically re-dividing the elementary folds.

In this thesis, we propose to use voxel-wise labeling to give a top-down perspective to a traditional bottom-up pattern recognition system. Thus, the initial cutting into elementary folds proposed by BrainVISA/Morphologist is challenged by voxel-wise labeling, eliminating under-segmentation errors in the model. The proposed approach is particularly robust to the spatial inconsistencies that can occur during voxel labeling and to the potential incorrect definition of upstream geometric entities.

5.2 Database

The training base is composed of 62 healthy brains selected from different heterogeneous databases and labeled with a model containing 63 sulci for the right hemisphere and 64 for the left hemisphere. The "unknown" label is used to designate unidentified structures (usually small sulci). Most of the subjects are right-handed men, aged 25 to 35 years old. Compared to (Perrot et al., 2011), where 62 sulci were considered for the right hemisphere and 63 for the left hemisphere, the same MRI acquisitions were used but the sulci were re-labeled: four new sulci were used and the two ventricles were labeled but not considered as sulci anymore, which is why 63 sulci are considered for the right hemisphere and 64 for the left in this thesis.

Unfortunately, there is no gold standard for the sulci nomenclature. Even the boundaries of the well-known central sulcus can be indistinguishable from one nomenclature to another (Figure 5.2). Therefore, for this study, the elementary folds of each brain were manually labeled according to a sulcus nomenclature following a long iterative process to achieve a consensus across a panel of several experts on cortex morphology. The last iteration of the database labeling was performed using the TileViz visualization tool (Mancip et al., 2018). This tool allows the entire database to be visualized and labeled simultaneously on a wall of screens (Figure 5.3). Until now it was only possible to label and simultaneously evaluate a limited number of hemispheres, generally four, on a standard screen. Thus, this tool helps to limit the bias of labeling induced by a restricted view of the database. To support this new iteration, the elementary folds were manually cut when

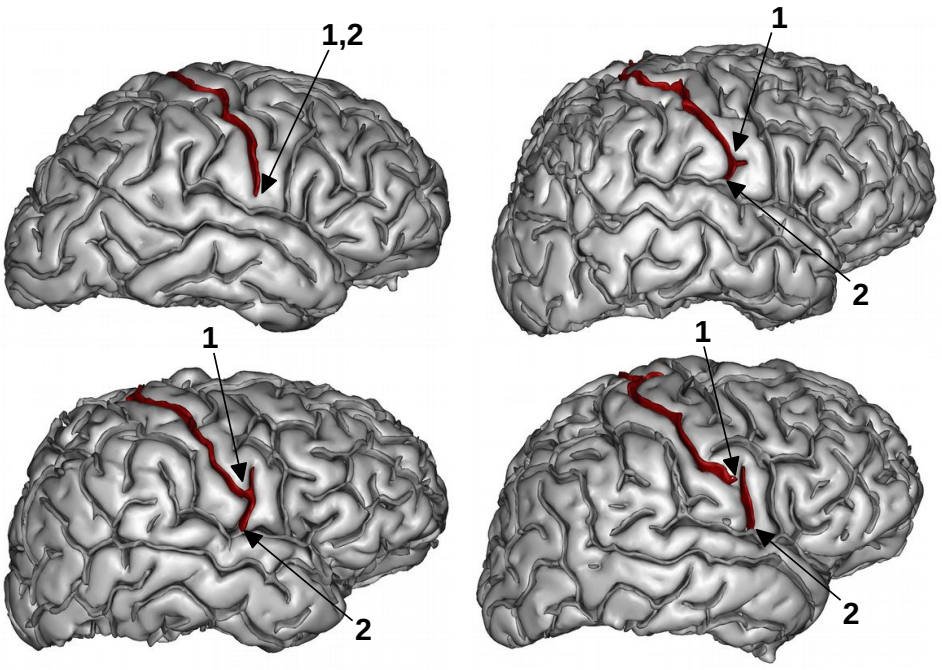


Figure 5.2 – Where should the central sulcus end? The folds that may belong to the central sulcus are shown in red. Limits 1 or 2 could be selected according to the nomenclature used.

necessary, which was not possible during the study of Perrot et al. (2011).

Note that compared to traditional labeling approaches where only one expert can label images, this database has been progressively labeled by several experts, both successively and simultaneously. This consensus-based labeling has sometimes led to the introduction of new sulci labels when it was considered necessary, making it essential to use the screen wall. However, the different experts have thus not produced independent labelings, which prevents us from assessing human-level performance on this dataset.

The new nomenclature used is described in the Figure 5.4. A more detailed description is provided in the following appendix section. The manually labeled database is now available on the BrainVISA website (*http://brainvisa.info/data/sulci_database/base_62/2019*).

5.3 Method

In this study, sulci labeling from an MRI is described in three steps (Figure 5.5). First, the folds are segmented from the MRI using the BrainVISA/Morphologist pipeline. Second, they are labeled using different algorithms. Third, the agglomeration of the voxels into elementary folds pro-

Figure 5.3 – Wall of screens of the Maison de la Simulation during a labeling session using the TileViz visualization tool. Fortunately, due to this wall of screens, brains can be visualized, moved/turned/zoomed all together, which allows to label them by checking the coherence with the entire database.

posed by the BrainVISA/Morphologist pipeline is used to regularize the results.

Note that the strategies used to set the method hyperparameters are detailed in the appendix.

5.3.1 Folds representation

Thanks to the BrainVISA/Morphologist pipeline (www.brainvisa.info), a widely used resource for studying cortical anatomy, the folds are represented by a set of voxels corresponding to a skeleton of the cerebrospinal fluid filling the fold. The representation of the folds, therefore corresponds to a negative mold of the brain. Once the fold skeleton has been segmented, the BrainVISA/Morphologist pipeline also proposes to fragment it into elementary folds based on topological and geometric constraints specific to the sulci's definition. This fragmentation is first based on the topological characterization of a simple surface proposed by Malandain et al. (1993) which isolates surface pieces that do not include any junction. The skeleton is also fragmented at the level of the buried gyri (Figure 5.6).

The skeleton representation has three main advantages. First, this 3D representation is essential during manual labeling because it allows the visualization of the relative position of the sulci between each other and the evaluation of their depth, size, etc. Additionally, the agglomeration of the voxels into elementary folds makes it possible to speed up labeling by giving

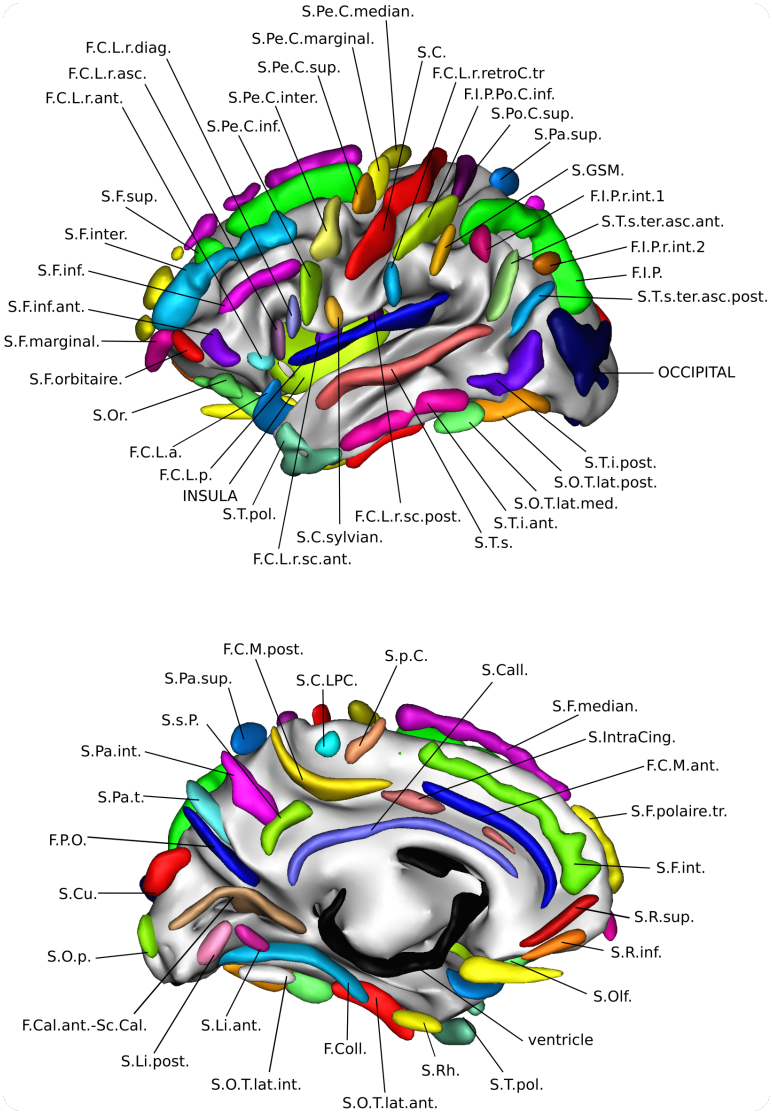


Figure 5.4 – New nomenclature used to label sulci. The visualization of the sulci labels is done thanks to the SPAM representation used by Perrot et al. (2011) which averages the position of the sulci as probability maps that are thresholded for this image. The new nomenclature includes 63 labels for the right hemisphere and 64 for the left hemisphere. Only the left hemisphere is represented in this figure. The right hemisphere has the same labels except the S.GSM. label. Compared to (Perrot et al., 2011), two new sulci are labeled (S.intraCing. and S.R.sup.). The ventricle label does not correspond to a sulcus label, but belongs to the fold skeleton extracted by the BrainVISA/Morphologist toolbox. Only the "unknown" label is not shown in this figure. Please refer to the appendix for English translations of each label.

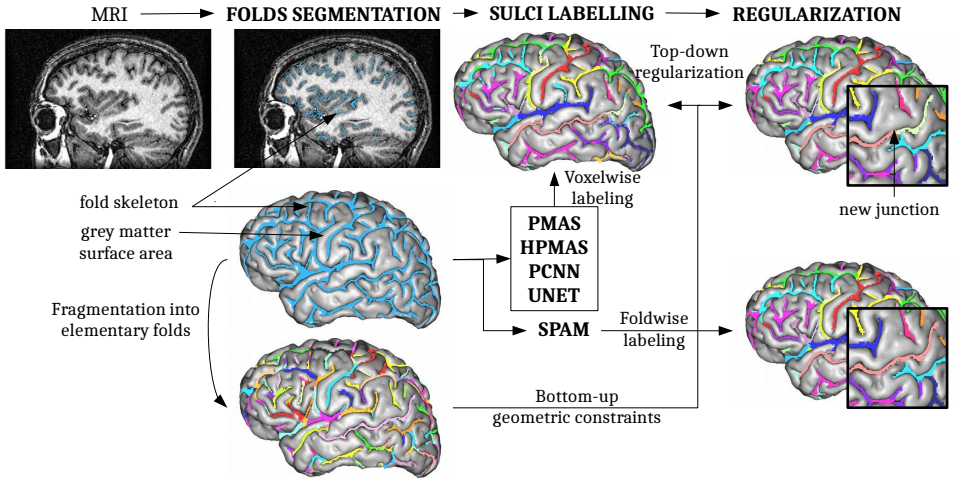


Figure 5.5 – MRI to labeled cortical sulci: a three-step pipeline. First, the fold skeleton is extracted using the BrainVISA/Morphologist toolbox. This toolbox also makes it possible to fragment the skeleton into elementary folds. Second, skeleton voxels are labeled by different algorithms. Algorithms based on MAS techniques (PMAS, HPMAS) and CNN-based algorithms (PCNN, UNET) label each skeleton voxel while the SPAM algorithm directly labels the elementary folds. Third, voxelwise labeling is regularized through the elementary folds, while automatically re-dividing them when the labeling indicates that it is required.

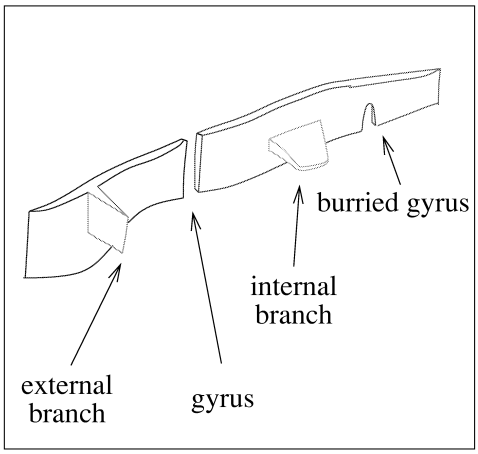


Figure 5.6 – Schematic representation of the fold skeleton. The fragmentation into elementary folds isolates the internal and external branches and cuts the skeleton at the level of the buried gyri. Image taken from Riviere et al. (2002).

a label to a set of voxels rather than individually. Second, as the data are particularly influenced by the type of MRI sequence, the age of the subjects (which has a significant impact on the opening of the sulci) or even their pathologies, this pre-processing enables optimal normalization of the data. Moreover, the algorithm can focus on labeling only after its segmentation. Finally, this representation has previously been used in other pipelines, making it possible to automate the calculation of measurements (depth, length, connectivity, etc.) used in morphometric studies or to realign the brains according to the major sulci (Auzias et al., 2011, 2013), which is why we have chosen to keep it. However, if we had chosen to construct a model to recognize the sulci, that carries out both their extraction and labeling without relying on this representation, it was highly probable that the results obtained would not conform to the representation used by these pipelines and that some significant post processing steps would be necessary.

Although the extraction of the fold skeleton is robust, its fragmentation into elementary folds demonstrates certain significant instabilities, such as vastly different fragmentations can be observed from the same MRI (Figure 5.7). Several stochastic optimizations were included in the segmentation pipeline (e. g. for bias correction, brain masking, skeletonization, etc.). These optimizations only have a slight impact on the shape of the resulting fold skeleton. However, for the topological fragmentation into elementary folds, a single voxel can then make the difference. Thus, these stochastic optimizations can have important consequences on the fragmentation of large simple surfaces. To remedy this, during manual labeling, the folds were cut manually when necessary. During automatic labeling, we propose a technique to automatically re-divide elementary folds when the labeling algorithm allows it.

5.3.2 Labeling methods

The methods described below seek to automatically label the voxels of the fold skeleton. Among the possible labels, while most correspond to cortical sulci, three other labels are used: those corresponding to the right and left ventricles and the "unknown" label. According to the methods presented here, the ventricles are treated as sulci, as they are relatively stable anatomical structures of the brain negative mold. However, the "unknown" label, corresponding to voxels that do not belong to any of the other labeled structures, must be treated differently in some cases.

5.3.2.1 Statistical Probabilistic Anatomy Map (SPAM) models

In this comparative study, the reference method corresponds to the one described in (Perrot et al., 2011), where they propose a coherent Bayesian framework to automatically identify sulci based on a probabilistic atlas (a

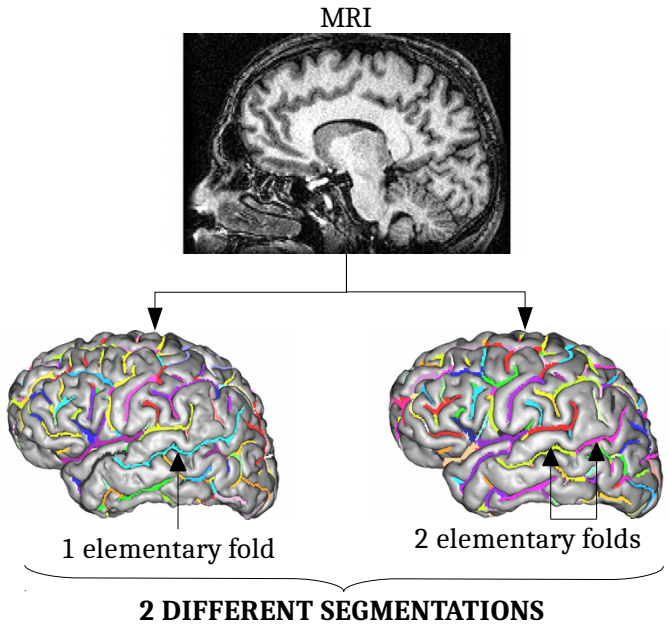


Figure 5.7 – Extraction of the elementary folds from the same MRI. In the two lower brains, each color represents a different elementary fold. We observed that the skeleton extraction is visually stable, but its division into elementary folds can produce very different results.

mixture of SPAM models) estimating simultaneously normalization parameters. This method, currently available in the BrainVISA/Morphologist pipeline, has been widely used on very large databases for large-scale morphometric studies (Le Guen et al., 2019). However, labeling errors are still too important to allow some studies to be reproduced. We believe that this is due to the fact that the SPAM models excessively generalize the high variability of sulci. Each sulcus can have several configurations, which may prove difficult to represent with a single average model.

5.3.2.2 MAS approaches

Two MAS approaches, PMAS and HPMAS, are compared in this section. The first approach is largely inspired by the one proposed in (Romero et al., 2017a) in which, unlike most MAS approaches, to similar atlases are searched between two cubic patches, instead of two full images. The second MAS algorithm presented here, and described in (Borne et al., 2018), aims to define a library of local patches embedding enough geometrical information to minimize ambiguities when searching for a high similarity hit in the unknown subject morphology. Therefore, instead of taking native cubic patches, this

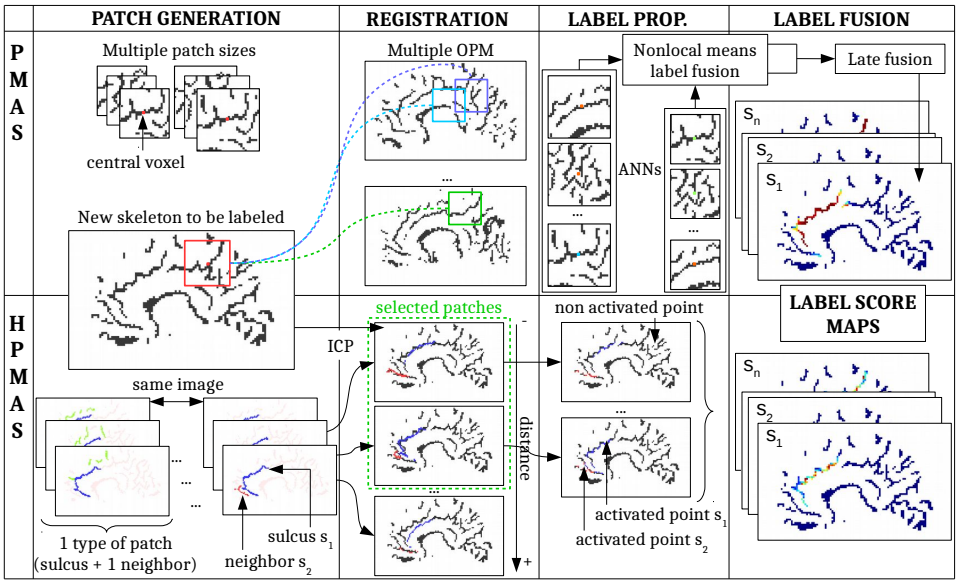


Figure 5.8 – Comparison of MAS approaches: PMAS vs. HPMAS. First of all, the patches are designed. Second, they are transferred to a new image to be labeled, where the fold skeleton has been extracted. Third, the best matches were selected and patch labels were propagated on the image to be labeled. Finally, the propagated labels are used to calculate the label score maps. In order to make the figures as readable as possible, we have chosen to represent the images in 2D while they are processed in 3D. All images are represented in 2*2*2 mm resolution, while for HPMAS, images are processed with the acquisition resolution. The acronym ANNs refers to the Approximate Nearest Neighbors patches obtained by the multiple run of the Optimized PatchMatch (OPM) algorithm.

algorithm builds virtual patches containing whole sulci.

These two approaches are described in four steps: first, the design of the patches (patch generation), second, the strategy of realigning the patches between them and selecting the best matches (registration), third, the strategy of propagating the labels from the patch to the brain to be labeled (label propagation) and finally the combination of the labels of the propagated patches (label fusion) (Figure 5.8).

Patch-based MAS approach (PMAS)

Patch generation The patches are extracted from images with a resolution of 2*2*2 mm, containing the fold skeleton that has been relocated in the well-known MNI space (Collins et al., 1994), which aligns the rough shapes of the brains through an affine transformation.

We chose to take into account only the patches with the central voxel belonging to the fold skeleton for two main reasons. First, it limits the number of patch matches that require optimization as the voxels belonging to the skeleton represent only a small part of the image's voxels. Second, since the patches are extracted from binarized images, the calculation of the distance between two patches can be successful only if the patches contain a minimum number of skeleton voxels.

As in (Giraud et al., 2016), we adopted a multi-scale approach, which involves the independent use of several patch sizes (determined by inner cross validation), to produce several score maps per label, which are then averaged.

Registration In order to find the most similar set of patches, we aimed to optimize the following distance d between two patches $P(S_A)$ and $P(S_B)$, respectively belonging to the fold skeletons S_A and S_B (superimposed by a simple translation):

$$d(P(S_A), P(S_B)) = \frac{d(P(S_A) \to S_B) + d(P(S_B) \to S_A)}{2} \tag{5.1}$$

The measurement from a patch $P(S_A)$ to a fold skeleton S_B corresponds to the average of quadratic Euclidean distances d_E of the skeleton voxels $p_A \in P(S_A)$ and their nearest neighbor in the fold skeleton S_B (Figure 5.9):

$$d(P(S_A) \to S_B) = \frac{1}{|P(S_A)|} \sum_{p_A \in P(S_A)} \min_{p_B \in S_B} [d_E^2(p_A, p_B)] \tag{5.2}$$

Note that, in order to avoid border effects, the closest neighbor of p_A is searched in the entire skeleton S_B and not only among the skeleton voxels contained in the patch $P(S_B)$.

Realigning and comparing all the patches in the database for each skeleton voxel to be labeled would be extremely expensive, making it impossible to label within a reasonable time. Additionally, it would increase the probability of spurious matching between remote areas in the brain while the images are already roughly aligned with each other. It is important to note that because we use binarized images, the risk of obtaining false positives is higher than usual.

In (Romero et al., 2017a), the Optimized Patch Match Label fusion (OPAL) (Ta et al., 2014; Giraud et al., 2016) was used. This segmentation method is based on the Optimized PatchMatch (OPM) algorithm which uses a cooperative and random strategy resulting in a very low computational burden. Compared to the PatchMatch algorithm (Barnes et al., 2009) from which it is inspired, OPM is adapted to 3D anatomical segmentation by taking into account the rough alignment of images. Here, as only patches with the central voxel belonging to the fold skeleton are considered, an

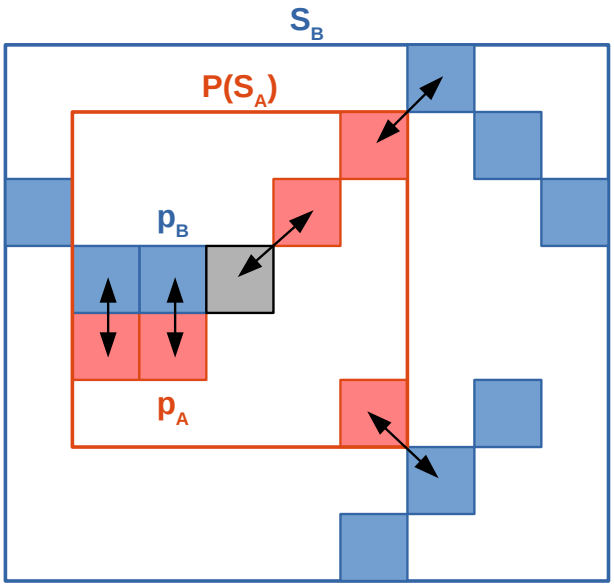


Figure 5.9 – Calculation of the distance from the patch $P(S_A)$ to the skeleton S_B for the PMAS method. The grey voxel represents the central voxel of the patch $P(S_A)$ which is superposed with a voxel of the skeleton S_B. For each voxel $p_A \in P(S_A)$, we look for its closest neighbor among the voxels of the skeleton S_B. The Euclidean distance between these two voxels is calculated. The distances over all the points $p_A \in P(S_A)$ and their nearest neighbors are then averaged to obtain $d(P(S_A) \rightarrow S_B)$.

adapted version of the OPM algorithm has been implemented. Please refer to the appendix of the previous chapter for more details.

Label propagation In order to select several Approximate Nearest Neighbors (ANNs) patch per skeleton voxel for a given patch size, multiple independent OPM were launched. The number of ANNs to be selected is determined by inner cross-validation. Once the ANNs have been selected, all the voxels of each ANN patch participates in the labeling, as done in (Rousseau et al., 2011; Giraud et al., 2016). However, there are only a few voxels belonging to the skeleton of the patch that overlap with the skeleton voxels to be labeled. Thus, we propose to propagate the label of each skeleton voxel of the patch to its nearest neighbor in the skeleton to be labeled.

Label fusion As in (Romero et al., 2017a), the non-local patch-based label fusion is used. In this strategy, the distance between patches is used to perform a robust weighted average of the labels. Once the non-local

means estimator has been calculated for all patch sizes, the final estimation is obtained by averaging these estimations thanks to a late fusion (Snoek et al., 2005). Thus, a score map is estimated for each label in the database.

Concerning the "unknown" label, present in the manually labeled database, it is treated like a sulcus label.

Patch-based MAS approach with High level representation of the data (HPMAS) As the standard way of extracting patches does not seem capable of exploiting the sulci geometry and the relations between them, which we believe to be the distinguishing features necessary for recognition, we have proposed a new virtual patch generation strategy based on a high level representation of the sulci (Borne et al., 2018). This framework is well adapted to leverage more information about the different folding configurations in the training dataset.

Note that this method is the only one of the proposed new methods to have been specifically developed for the recognition of cortical sulci. It includes many arrangements specific to this application. Its complex design gives an idea of the scores that can be obtained by pushing as far as possible in this direction. To facilitate the understanding of this ad-hoc method, Figure 5.10 represents the pipeline in 3D, which complements the 2D representation provided in Figure 5.8.

Patch generation In order to take into account as much geometric information as possible, the idea was to define virtual patches containing whole sulci. Some sulci are however too large to be contained in a cubic patch: it would then be too noisy. A virtual patch therefore only corresponds to a set of extracted sulci for this method. Note that the shape of small sulci is not specific enough to prevent spurious hits. Hence the we intended to aggregate two sulci to create discriminative local shapes. In the following, we define a type of virtual patches for each pair of sulci that are neighbors in the brain, extracted from MNI space at the image resolution.

In practice, a pair of sulci is selected in the circumstance that the two sulci are neighbors in at least one brain of the atlas dataset, according to the topology provided by the BrainVISA/Morphologist pipeline that produces the folds. This pipeline endows the list of folds with a graph structure corresponding to either direct connections or to the fact that two folds are separated by a piece of gyrus. Finally, each type is made up of the instances of the pair of sulci in the atlas dataset, most of the time as many shapes as atlases (some atlases miss a few small sulci) (Figure 5.10.1).

Note that only the unknown sulcus label is not selected to form virtual patches, as it does not constitute a coherent structure like the other labels. Thus, unlike the previous PMAS method, the unknown label is not treated like other sulcus labels.

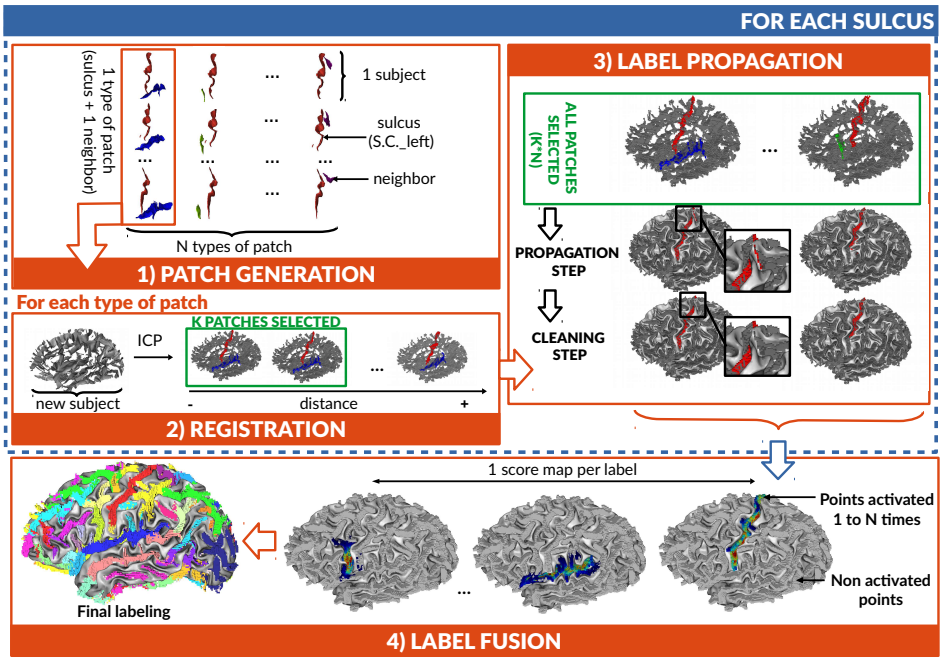


Figure 5.10 – 3D representation of the HPMAS method. As for the Figure 5.8 which represents the method in 2D, the approach is described in four steps: generating the virtual patches, registering them on the image to be labeled, propagating the labels of the selected virtual patches and finally merging the propagated labels to obtain the final labeling.

Registration For the registration step, the set of folds of the unknown brain and the virtual patches of the library are represented by point clouds. As in (Sun et al., 2012), we have applied the well-known iterative closest points algorithm (Besl and McKay, 1992), with the robust implementation of Holz et al. (2015), to find an optimal alignment of each virtual patch into the skeleton point cloud of the unknown brain. Note that compared to the PMAS approach which only uses translations to superimpose patches, the registration here allows rotations.

To build the measure used to rank the matches, the nearest voxels in the new fold skeleton S_B of each skeleton voxel $p_A \in P(S_A)$ are saved as activated voxels $p_B^* \in S^*_{B,P(S_A)}$. Then, the measure corresponds to the sum of the quadratic distances of the skeleton voxels and their corresponding activated voxels, divided by the number of different activated voxels :

$$d(P(S_A) \rightarrow S_B) = \frac{1}{|S^*_{B,P(S_A)}|} \sum_{p_A \in P(S_A)} \min_{p_B \in S_B} [d_E^2(p_A, p_B)] \qquad (5.3)$$

Note that by dividing by $|S^*_{B,P(S_A)}|$, we take into account the number

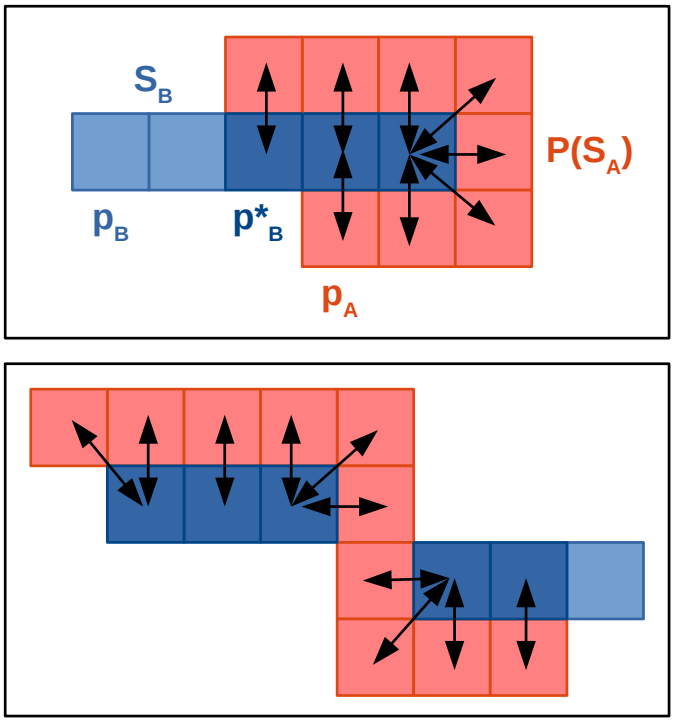


Figure 5.11 – Calculation of the distance from the virtual patch $P(S_A)$ to the skeleton S_B for the HPMAS method. For the sake of clarity, the skeletons S_A and S_B represented do not overlap in this Figure. For each voxel $p_A \in P(S_A)$, we look for its closest neighbor among the voxels of the skeleton S_B. The Euclidean distance between these two voxels is calculated. The distances over all the points $p_A \in P(S_A)$ and their nearest neighbor are then summed and divided by the number of different activated points p_B^* to obtain $d(P(S_A) \rightarrow S_B)$. The two configurations represented are penalized by the division by the number of different activated points rather than by the number of points in $P(S_A)$ as for a classical average. On the first configuration, we observe that the proposed distance penalizes the virtual patch more if its shape is more complex or if its size is larger than the structure on which it has been registered. On the second configuration, we observe a greater penalization of the virtual patch if it has only one connected component and if it is registered on two different components.

of different activated points. This allows the penalization of virtual patches where several points activate the same point of the skeleton to be labeled (Figure 5.11).

With regards to each type of virtual patch, all matches are ranked ac-

cording to the distance proposed above. A fixed number of matches (determined by inner cross-validation) leading to the shortest distances is selected to propagate the two parent sulci. All types of virtual patches are selected the same number of times even if they are not all equally informative. It is important to note that some sulcus instances are selected several times, because they win the competition for several virtual patch types, but their multiple contributions will be associated with slightly different alignments. Hence, sulcus instances maximizing regional similarity to the unknown subject get more weight.

Label Propagation Each selected virtual patch after the optimal alignment to the unknown subject, concomitantly propagates the label of each voxel to its nearest neighbor in the target brain. To consider the virtual patch structure, each connected set of voxels in the virtual patch should correspond to a unique connected set in the target brain: the smallest non-connected sets are excluded (Figure 5.10.3).

Label Fusion Post complete propagation of all the proposed virtual patches, the score maps are calculated by averaging the number of times the points are activated by different virtual patches. Compared to PMAS, where patches are weighted by their distance to the patch to be labeled, here each propagated point has the same weight in the label fusion. It is imperative that a distance should be determined that takes into account both the distance from the virtual patch point to the point it has activated and the distance from the entire virtual patch to the skeleton to be labeled. We were unable to converge on an appropriate distance from the problem at hand, so we chose to avoid weighting.

As the "unknown" label does not belong to any virtual patch, its score map is empty. This label will be selected only if the score maps of all other labels are also empty for a given elementary fold.

5.3.2.3 CNNs based approaches

As this is the first time that 3D CNNs are used for sulci labeling, we take inspiration from two models that have proven their efficacy in medical image segmentation (Figure 5.12): the first being a patch-based approach inspired by Ciresan et al. (2012) and the second an approach that treats the entire image with a 3D U-Net as in (Çiçek et al., 2016). First the common modalities used during training of these two networks are detailed followed by an individual description of each network. The models presented are implemented using the Pytorch library (Paszke et al., 2017).

Data All the fold skeletons are registered in MNI space and used as input: they correspond to 3D binary volumes with a common resolution of

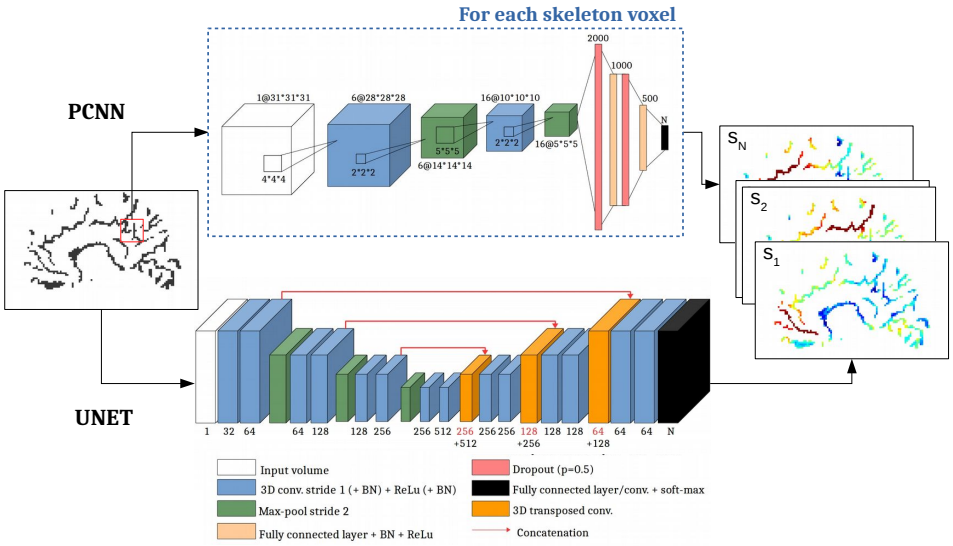


Figure 5.12 – Comparison of CNN-based approaches: PCNN vs. UNET. Boxes represent feature maps. The number of channels is denoted next to each feature map. The size of the feature map is indicated after the @ when appropriate. N is the number of different labels to be predicted. For clarity sake, input and output are represented in 2D rather than 3D.

2*2*2mm, where the voxels belonging to the skeleton are one and the others are zero. In order to augment the training dataset, a rotation in a random direction with a random angle (following a Gaussian distribution $\mathcal{N}(0, \frac{\pi}{16}^2)$) is applied to the images at each epoch.

At the output of the neural network, a score per label present in the database is obtained per voxel. Concerning the "unknown" label, it is treated like a sulcus label.

Training design Initialization of the weights of the neural networks was done as in (LeCun et al., 2012). Stochastic gradient descent was used for training, with learning rate and momentum determined by 3-folds inner cross validation. The learning rate was halved when the loss function had not improved for two consecutive epochs. After four consecutive epochs without improvement, training was stopped. The selected trained neural network corresponds to the epoch obtaining the lower error rate E_{SI}, described in (Perrot et al., 2011) and in the following section.

The loss function used is the cross- entropy loss. In most cases, for unbalanced problems, the loss function must be weighted to avoid favoring the labels most involved in backpropagation, due to their higher presence in the database. Although the average size of each sulcus is extremely unbalanced, we have chosen not to weigh this loss function because large sulci are also

the most interesting from a neuroanatomical point of view and need to be better recognized than small ones.

Patch-based model with a 3D CNN (PCNN) PCNN method adapts the approach proposed in (Ciresan et al., 2012), addressing a segmentation problem as a classification of each voxel based on its environment contained in a patch. Here, only voxels belonging to the skeleton are selected to participate in the classification.

We designed the architecture of the neural network so that it takes cubic patches of 31 side voxels in input, i.e. about 6 cm side, which we considered to be large enough to identify its central voxel. During training, the dropout strategy (Srivastava et al., 2014) with a probability of 0.5 is used on fully connected layers. Batch normalization (Ioffe and Szegedy, 2015) was also used on convolutional and fully connected layers. The batch size has been set at 100 to minimize learning time and fit in memory. In order to ensure that the inner cross-validation is not too time-consuming, only three epochs are calculated for each hyperparameter value tested.

3D U-Net based model (UNET) For the UNET method, the network architecture used is largely inspired by the one presented in (Çiçek et al., 2016). The particularity of this application of U-Net lies in the fact that all the voxels that do not belong to the fold skeleton, i.e. a large majority of the voxels in the image, do not need to be classified. Indeed, as the values predicted by U-Net are masked by the segmentation of the fold skeleton made upstream, the background voxels do not need to be predicted and therefore do not need to be learned. Thus, during training, all voxels that do not belong to sulci are not used for gradient backpropagation. The batch size has been set at 1 in order to fit in memory.

5.3.3 Bottom-up geometric constraints

In order to standardize the results, the voxels were agglomerated into elementary folds. However, these folds are not always sufficiently fragmented, so we propose to use the label score maps to reconsider their fragmentation.

The straightforward approach to regularize the results is to do a weighted majority vote. The scores of each elementary fold were averaged by label and the highest score label was selected. This strategy was used as a reference to evaluate the impact of the automatic re-division of elementary folds.

In this thesis, we propose to re-divide the elementary folds with help of the Ward's hierarchical agglomerative clustering method (Ward Jr, 1963). Clustering for each elementary fold was performed based on the label score maps. In order to ensure spatial consistency, a spatial connectivity constraint was imposed during cluster agglomeration. Then, the Calinski-Harabaz index (Caliński and Harabasz, 1974) was used to quantify the

quality of the proposed clustering. This score corresponded to the ratio of the between clusters dispersion mean and the within cluster dispersion. The ratio was higher when clusters are dense and well separated. If this score was higher than a threshold determined by inner cross validation, the partitioning was performed. When an elementary fold was split in two, each of the two clusters obtained were also challenged with the same manipulation, until all the elementary folds had a Calinski-Harabaz index below the threshold.

5.3.4 Performance evaluation of labeling models

As in (Perrot et al., 2011), two measures were used to compare the different models proposed above: E_{local} at the sulcus scale and E_{SI} at the subject scale. Error rates were assessed by 10-folds cross validation. One model was trained per hemisphere.

5.3.4.1 Mean/max error rates

To take into account the variability of the fragmentation into elementary folds and therefore the robustness of the labeling methods to this variability, each image was re-segmented ten times (Figure 5.13). Thus, if the image belonged to the training set, only the segmentation used for manual labeling was considered. However, if the image belonged to the test set, ten other segmentations (whose true labels have been transferred from manual segmentation) were labeled and used to quantify the error rates. Note that manual segmentation was not used to calculate error rates. Using ten different segmentations for each sulcus highlights the weaknesses of the BrainVISA/Morphologist preprocessing since we can compute errors from the worst result, typically associated to an issue of under-segmentation.

To quantify errors, for each new segmentation, the manual labeling on the initial segmentation must be transferred to the new one. Because of the variability of the segmentations obtained and the sparsity of the fold skeleton, the simple superposition of images was insufficient. To remedy this, a Voronoi diagram of the training segmentation was used to label the voxels of any other segmentation. Note that the elementary folds were not used to transfer the labeling and that the true labeling was on the voxel scale.

For each subject, from the ten segmentations, the average of the errors (E_{SI}^{mean} and E_{local}^{mean}) and the maximum error (E_{SI}^{max} and E_{local}^{max}) were calculated. Note that the training segmentation used for manual labeling was not used in the error calculation because it would bias our evaluation. By considering the maximum error rates, labeling errors due to model variability were highlighted. These errors in most models were related to an incorrect fragmentation of the fold skeleton into elementary folds. Only the PMAS

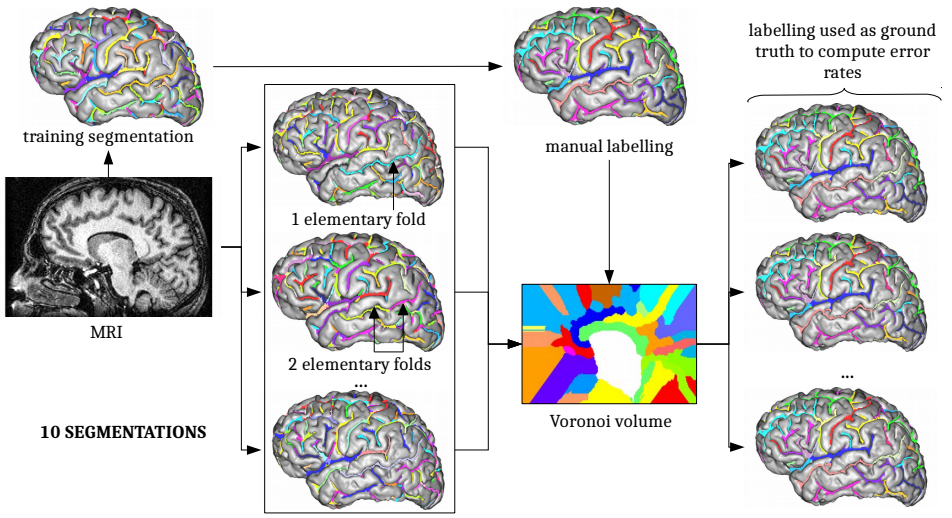


Figure 5.13 – Estimation of error rates taking into account the variability of the fragmentation into elementary folds. Using the same MRI, several different fold segmentations can be obtained because of stochastic optimizations embedded in the pipeline. In order to take this variability into account in the calculation of error rates, ten segmentations are performed from each image of the database. Then, a Voronoi diagram was used to propagate manual labeling to the ten new segmentations. The labels obtained were used to estimate the average and maximum error rate per image.

labeling model was not deterministic and includes stochastic optimizations that can penalize the calculation of maximum error rates.

5.3.4.2 Error at the sulcus scale: E_{local}

Given a sulcus l,

$$E_{local}(l) = \frac{FP_l + FN_l}{FP_l + FN_l + TP_l} \tag{5.4}$$

with TP_l, FP_l and FN_l, respectively the number of true positive, false positive and false negative voxels for the sulcus l.

It is important to note that the error rate was one, when the sulcus was absent and labeled by the model. Similarly, for when the sulcus was present but not labeled by the model. As small sulci are frequently absent, this explained why error rates can be highly variable when averaging the error rates per subject.

5.3.4.3 Error at the subject scale: E_{SI}

Given a set of sulci L,

$$E_{SI} = \sum_{l \in L} w_l * \frac{FP_l + FN_l}{FP_l + FN_l + 2 * TP_l} \tag{5.5}$$

with $w_l = s_l / \sum s_l$ and $s_l = FN_l + TP_l$, the sulcus l true size.

The error at the subject scale allows local errors to be generated in a single measurement. As explained in (Perrot et al., 2011), each component of the sum over labels differs on two points compared to $E_{local}(l)$. First, true positive measures are counted twice as compared to the false positive and negative measures, in order to remove errors shared by several labels, since each extra sulcal piece for a given label is a missing part for another label. Second, each component was weighted according to the sulcus true size so that each local component count as much as its size.

Compared to (Perrot et al., 2011), three labels were not included in the set of sulci (unknown and both ventricles). These labels were not particularly considered as sulcus labels, but correspond to other structures, not pertinent to our study. Thus, the scores presented here for the SPAM method are worse than presented in (Perrot et al., 2011) for four reasons. First, because removing the two labels considerably improved the scores. Second, because we cut the elementary folds during manual labeling while the SPAM model cannot automatically correct this kind of sub-segmentation errors. Third, because we are interested in the mean/max of the error rates. Finally, because the error rates are estimated by 10-folds cross-validation and not by leave-one-out cross-validation. Moreover, the addition of the four new sulci labels and our refined labeling of the training dataset may also have impacted the results.

5.3.4.4 Error rate comparison

During the 10-folds cross validation, each fold contained approximately 6 hemispheres labeled to test the model's performance. Error rates are calculated by hemisphere and then averaged over the entire database to obtain the mean error rates per model. When not specified, the average error rate includes the right and left hemispheres. In order to compare the models in pairs, a paired T-test was performed between the error rate lists for each hemisphere. If the p-value was less than 0.05, the error rates were considered significantly different.

5.4 Results

5.4.1 Which is the best model?

In order to compare the five models presented above, we were interested in the E_{SI}^{mean} and E_{SI}^{max} for each model, trained separately on each hemisphere (Figure 5.14). Please refer to the appendix for the numerical values of the error rates per hemisphere (Tables 5.1 and 5.2).

First, we observed that all of the new approaches proposed with regularization per elementary folds were significantly better than the SPAM approach (also based on this regularization), which suggests that a model based on an average template was not the most appropriate to represent the high variability of cortical folds.

Second, with regards to the four proposed methods, regularization by elementary folds of the label score maps significantly improved the results compared to voxel labeling. Most importantly, the automatic re-division of these elementary folds also significantly improved the four methods. Thus, the use of top-down refinement of bottom-up regularization is particularly relevant in this thesis.

Third, by comparing the new models in pairs, the models seem to demonstrate equivalent performance. Concerning E_{SI}^{mean}, only the UNET model is significantly better than the PMAS model, the other models were equivalent.

While observing E_{SI}^{max}, it is interesting to note that the HPMAS model was significantly better than the PMAS model whereas this was not the case with E_{SI}^{mean}. This is probably due to the initialization and constrained random search steps of the OPM algorithm. These steps, based on random selections, sometimes seem to generate worse labeling than usual. The robustness of this model could probably be improved by increasing the number of selected ANNs.

Concerning the PCNN and UNET models, this thesis consequently demonstrated the incredible efficiency of neural networks, even for the recognition of structures as variable as cortical folds. However, it is surprising that the UNET model was not better than the PCNN model due to its deeper architecture.

The fact that these four models do not stand out radically on this dataset suggests that these models may have reached the limit of what can be interpreted from this database, probably due to its insufficient size to represent the high variability of cortical folds. Therefore, the fold variability is such that manual labeling of a brain raises many questions and it may be possible that the models have reached the human-level performances. Unfortunately, since manual labeling is based on consensus among several experts, it is impossible for us to assess human-level performance on this database.

Finally, with regards to the computation time required to label a hemi-

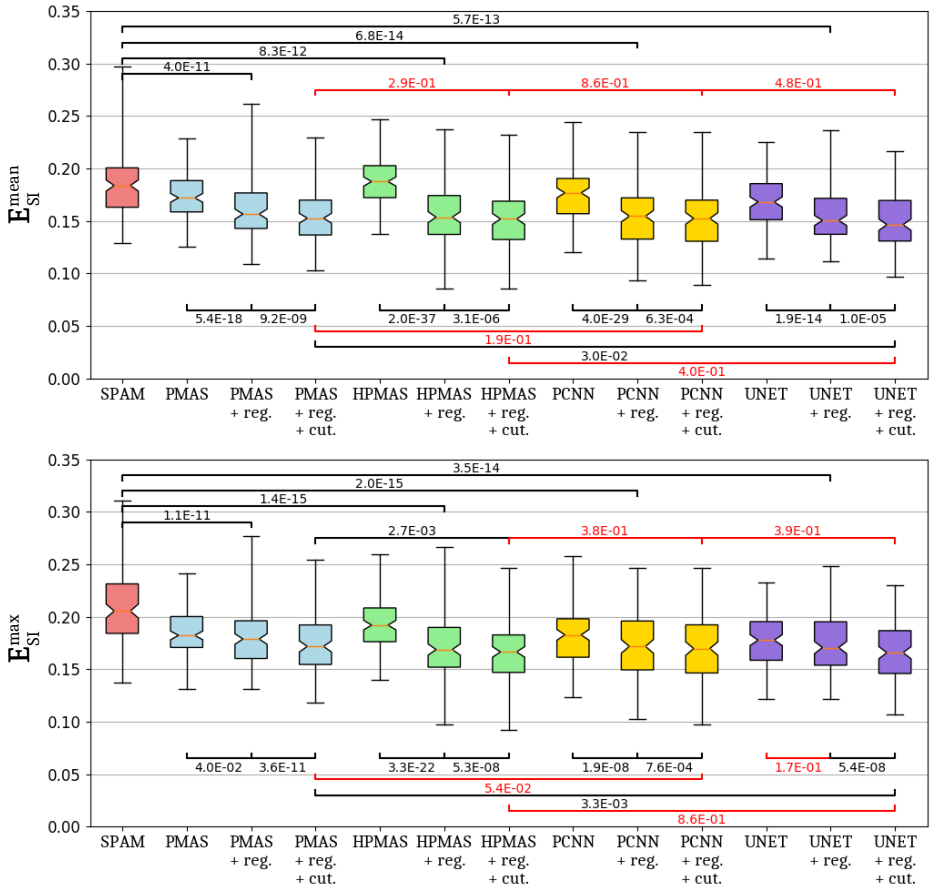


Figure 5.14 – Comparison of E_{SI} error rates by model. Once the 10 segmentations have been labeled by hemisphere, we consider the average error in the upper chart and the maximum error in the lower chart. The box extends from the lower to upper quartile error values, with a line at the median. The whiskers extend from the box to show the minimum and maximum limits of the error rates. The SPAM model is represented in red, the PMAS model in blue, the HPMAS model in green, the PCNN model in yellow and the UNET model in purple. For the four new models, three modalities are represented: first, labeling at the voxel scale, then labeling after regularization at the elementary fold scale (+ reg.), and finally the labeling obtained after automatic re-division of the elementary folds (+ reg. + cut.). The models are compared by matched T-test. The p-values of the differences in model performances are written above and below the compared models. The p-value is written in black if it is less than 0.05 and in red otherwise. Regularization by elementary folds significantly improves results. Automatic fold re-division also significantly improves results. All regularized models are significantly better than the SPAM model.

sphere, the SPAM model takes about 5 minutes, while the UNET model takes about 20 seconds, PCNN takes slightly more than a minute, PMAS and HPMAS take several hours. Although the PMAS model could be much faster by optimizing the codes as in (Giraud et al., 2016), the UNET model is currently by far the fastest. Thus, since the UNET model has the lowest error rates and is the fastest, we propose to study in more detail the differences between this model and the SPAM model in the following section. In the rest of this study, the UNET model will therefore refer to the model with regularization using elementary folds and automatic redivision of these, if necessary.

5.4.2 Which sulci are better recognized?

Concerning E_{local}^{mean} and E_{local}^{max} (Figure 5.15), the SPAM model has average/max error rates from 5% to 77% while the error rates of the UNET model vary between 2% and 68%. Comparing the E_{local}^{max} of each sulcus, we can see that the difference between the error rates of both model for a given sulcus reaches up to 25%. Finally, almost all sulci were better recognized by the UNET model, only about twenty sulci are less well recognized. Their comparison by paired T-test, by controlling the false discovery rate with the help of the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), showed that around 15% of sulci were significantly better recognized by UNET than SPAM (Figure 5.16), while none were significantly less well recognized. Please refer to the appendix for exact values of sulcus error rates (Tables 5.3, 5.4, 5.5 and 5.6). You will also find a graphical comparison of sulcus error rates between SPAM and UNET labeling in the appendix (Figure 5.24).

In the next section, we focus on the impact of the significant improvement in central sulcus recognition, in which the E_{local}^{max} value has gone from about 8% using the SPAM model to only 3% with the best UNET model.

5.4.3 Does this improvement have an impact in practice?

Here, the SPAM model and the UNET model were trained on the entire manually labeled database. The hyperparameters of the UNET model were estimated over the entire database, using the same procedures as during inner cross-validation, i.e. by performing a 3-folds cross-validation to select the hyperparameter values that minimize error rates. The database used by Sun et al. (2012) to study the effect of handedness on the shape of the central sulcus was labeled manually and automatically by these two models. This database contains 23 right-handed and 18 left-handed people.

We propose to investigate the asymmetry index I of the central sulcus length along the brain hull between the left $l_{S.C.\_left}$ and right hemispheres $l_{S.C.\_right}$:

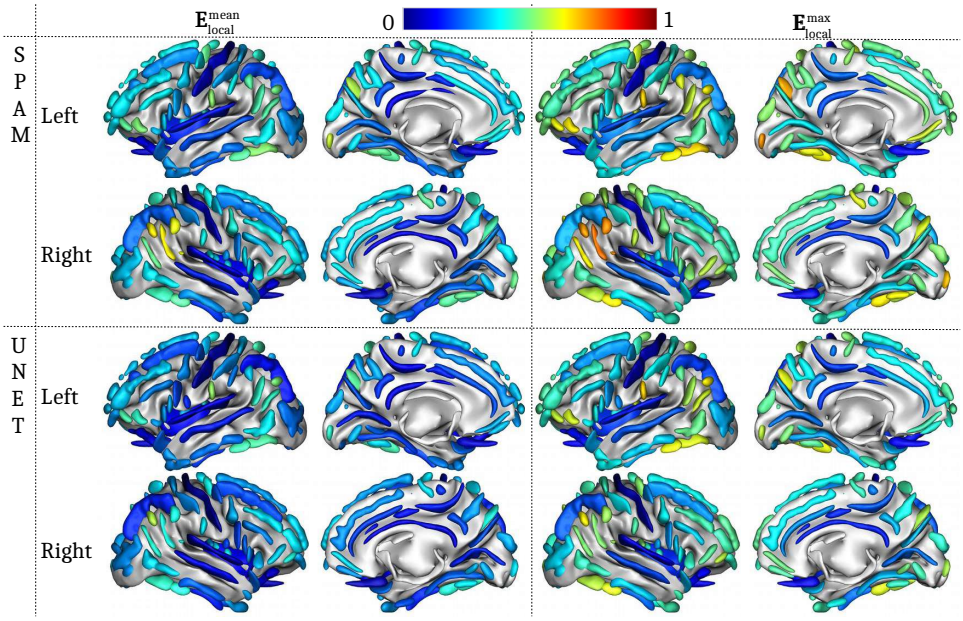


Figure 5.15 – E_{local} error rate per sulcus for SPAM and UNET models. The UNET model corresponds to the one after re-division of the elementary folds. Once the 10 segmentations have been labeled by hemisphere, we consider the average error per sulcus in the left column and the maximum error in the right column. The external and internal sides are represented for each of the right and left hemispheres.

$$I = \frac{l_{S.C.\_left} - l_{S.C.\_right}}{l_{S.C.\_left} + l_{S.C.\_right}} \tag{5.6}$$

Note that in the nomenclature proposed in this thesis, two sulci labels belong to the central sulcus: "S.C." and "S.C._sylvian.". Therefore, the lengths of these two "sub-sulci" are added together to obtain $l_{S.C.}$.

With manual labeling, there is a significant difference between left-handed and right-handed people (Figure 5.17). Therefore, left-handed people have on average a longer central sulcus in the right hemisphere than in the left, and vice versa for right-handed people. However, when focusing on the asymmetry index with SPAM labeling, no significant difference was found, whereas this difference was significant with UNET labeling.

Considering the worst labeling errors (Figure 5.18) of each model, we observe that the SPAM model can double the size of the central sulcus, by labeling completely unrelated large structures. However, the UNET model only adds small fragments.

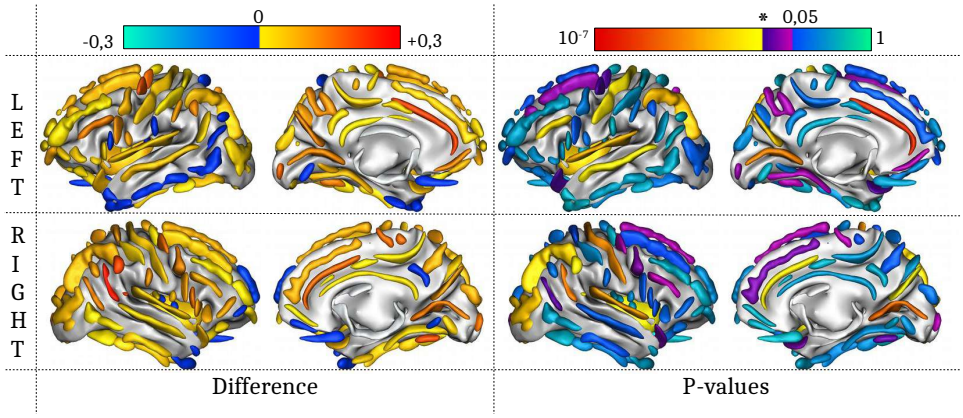


Figure 5.16 – Comparison of E_{local}^{max} error rates between the SPAM model and the UNET model. The left column represents the difference between the E_{local}^{max} of the SPAM model and of the UNET model. The right column shows the p-value of the paired T-test between each model. Note that the scale of the color palette used to represent p-values is logarithmic. In order to visualize the sulci significantly better recognized, the threshold 0.05 is indicated and the threshold at the star corresponds to the first sulci considered significantly better by controlling the false discovery rate through the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

5.5 Discussion

5.5.1 Patch-based MAS (PMAS)

Considering the hyperparameters selected during the inner cross validation (Figure 5.19), it seems that this method would benefit from testing new patch sizes and increasing the number of ANNs selected by voxel. Indeed, the 24 mm patches are the ones selected while being the smallest size tested. However, we concur that using patches less than 24 mm would make them insufficiently informative to capture sulci variability and would not fundamentally change the results. Note that the use of several patch sizes does not improve the results, which may be due to an oversized patch edge effect when propagating labels. Additionally, the number of ANNs is automatically set to 10, which is the upper limit of the values proposed in the inner cross-validation. A large number of ANNs probably compensate for the instability of the PMAS model due to the stochastic optimizations used by the OPM algorithm. However, testing a larger number of ANNs would require optimization of the codes currently in use and it is very likely that the model would not gain much in performance. Indeed, the evolution of the scores according to the number of ANNs suggests that a plateau is reached and that increasing this parameter would have little influence on the ranking of the methods obtained.

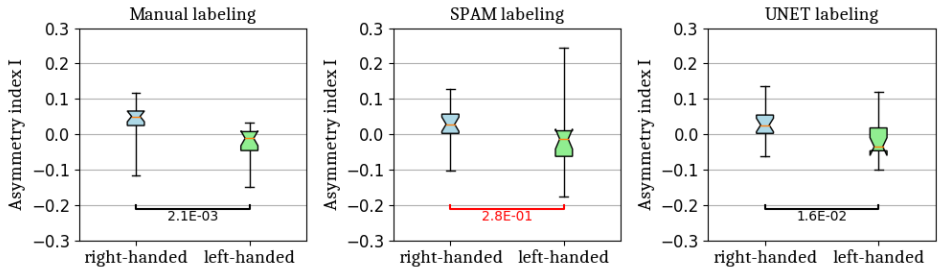


Figure 5.17 – Comparison of the asymmetry index I between right-handed and left-handed people. The left/middle/right graphs respectively show the results obtained with manual/SPAM/UNET labeling. The index for right-handed people is represented in blue and the one for left-handed people in green. The p-values of the T-test for the means of these two independent samples of scores are indicated on the graphs. With manual labeling, there is a significant difference: in left-handed people, the central sulcus is longer in the right hemisphere than in the left, while this is the opposite in right-handed people. The same significant difference is observed with the UNET model labeling but not with the SPAM model labeling. The box extends from the lower to upper quartile index I values, with a line at the median. The whiskers extend from the box to show the minimum and maximum values.

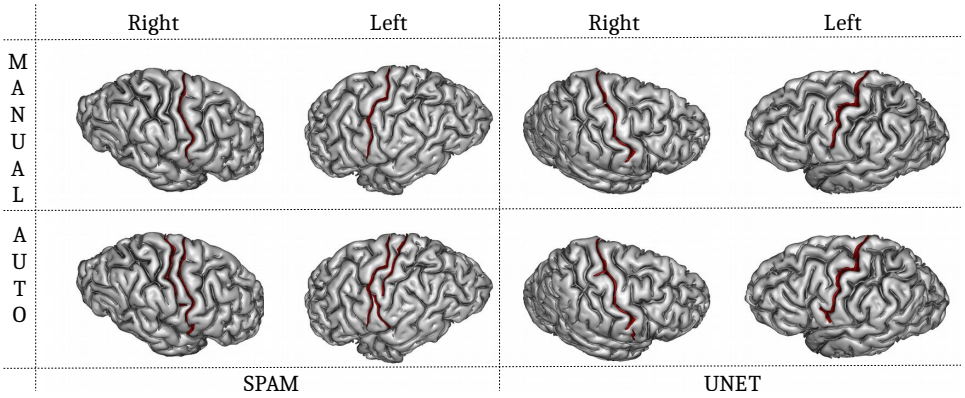


Figure 5.18 – Worst central sulcus labeling errors on the database used by Sun et al. (2012). The first line represents manual labeling. The second line represents the corresponding automatic labels, with the worst labeling errors of the SPAM model on the left and those of the UNET model on the right.

5.5.2 Patch-based MAS with High level representation of the data (HPMAS)

With regard to HPMAS, the choice to use sulci pairs to form patches was questionable, since there was no evidence suggesting that two sulci are suffi-

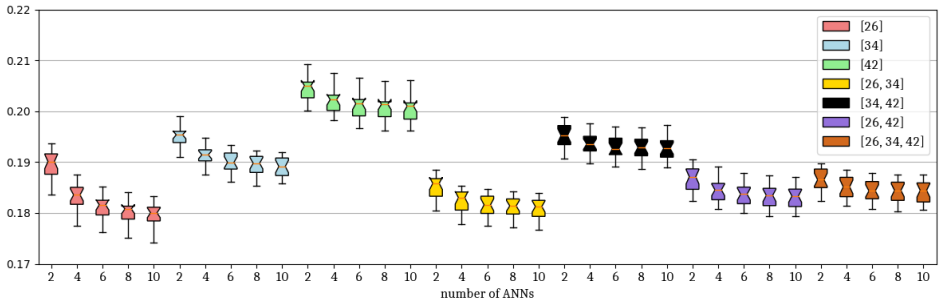


Figure 5.19 – Hyperparameters values for PMAS model. During inner cross-validation, the hyperparameters are set using a grid search to select values that minimize the average E_{SI}. This error rate is calculated based on the segmentation used to make the manual labeling. The number of Approximate Nearest Neighbors (ANNs) was set between 2 and 10. The patch sizes tested were 26, 34 and 42mm. All possible combinations of these patch sizes are represented in a different color (see the legend at the top right). In this graph, the average E_{SI} are represented for each pair of values tested. An external cross validation of 10 folds being done per hemisphere, 2*10=20 average E_{SI} are represented per possible value pair. The box extends from the lower to upper quartile error values, with a line at the median. The whiskers extend from the box to show the minimum and maximum values.

cient to prevent spurious hits, especially when two small sulci are associated. In order to create distinguishable local shapes, patches containing three or more sulci should also be considered. However, it would be too expensive to take into account all combinations of three neighboring sulci, as it is done for pairs of sulci. To remedy this, criteria for selecting relevant patch types should be determined, but none of the criteria we tested improves the results sufficiently to be considered here.

5.5.3 Patch-based CNN (PCNN) and UNET

Compared to the approach proposed by Ciresan et al. (2012), the PCNN approach has a major difference. In (Ciresan et al., 2012), several patch sizes, processed by several neural networks in parallel, were used to label each pixel, yet our PCNN approach is based on only one patch size. Moreover, the neural network used for PCNN is not deep (only one hidden layer) compared to (Ciresan et al., 2012). However, after trying to make the network architecture more complex by increasing the number of hidden layers or using multiple patch sizes, we did not observe significant improvements in the results. It is imperative to note that the PCNN model achieves performances comparable to the UNET model while the U-Net architecture is much deeper and previous studies show that it is supposed to achieve better

results (Ronneberger et al., 2015).

5.5.4 Unknown label

In this thesis, except for the HPMAS model, the "unknown" label in the manually labeled database is treated like the other sulci labels. However, although the "unknown" label represents about 0.5% of the skeleton voxels of manual labeling, this proportion is null if we consider the labels of the HPMAS model. Moreover, the PMAS and PCNN models label around 0.02% of voxels as "unknown" and the SPAM and UNET models 0.04%. These figures show that treating the "unknown" label as other labels is insufficient. Models should also assign the "unknown" label to structures where it is not sufficiently confident.

However, since all the new methods were compared to the SPAM model, which treated the unknown label as sulci labels, we chose not to address this point in this thesis.

5.6 Conclusion

To summarize, the new methods presented in this thesis outperform the current SPAM model provided by the Morphologist toolbox of BrainVISA. Compared to the SPAM model, the best models have a 4% higher recognition rate and 15% of sulci are significantly better recognized. By automatically re-dividing the elementary folds, the new models are considerably more robust to under-segmentation errors. In practice, these improvements make it possible to reproduce findings that were previously only possible with manual labeling. The UNET model will soon be available in the BrainVISA/Morphologist toolbox.

In this thesis, the application of methods based on MAS or CNNs give approximately the same results for the automatic recognition of cortical sulci. However, although CNN-based methods have a particularly long training process compared to MAS-based methods, which are significantly faster. Therefore, CNNs-based methods are far more productive in practice. The UNET method labels a brain in only twenty seconds, whereas the SPAM method takes about ten minutes. It is interesting to note that patch MAS approaches are also beginning to integrate deep learning techniques (Manjón et al., 2018), probably due to their ability to effectively summarize the data and for their rapidity of execution.

Furthermore, the top-down refinement of bottom-up regularization significantly improves the results. Indeed, voxel-wise labeling is used to give a top-down perspective to a traditional bottom-up pattern recognition process that agglomerates the voxels into elementary folds: these folds can therefore be automatically re-divided when necessary. Thus, the labeling is robust to

under-segmentation errors, unlike the SPAM method, which does not provide voxel-wise labeling. Note that despite the definition of elementary folds specific to the problem posed here, defining a coherent geometric entity is a legitimate concern addressed in many segmentation problems, for example by using super-pixels (Giraud et al., 2017; Soltaninejad et al., 2017) that group the most similar connected pixels together so that they have the same label.

In order to improve the current performance of the model, several options remain to be considered. First, the management of the "unknown" label is not optimal. Thus, in practice, it is preferable to have an "unknown" label on a structure where the model is not certain, rather than using an unreliable one. However, this aspect is not addressed by the models presented. Second, the inputs currently contain the fold skeleton in order to normalize the data for acquisition and age biases. However, the input can be enriched by taking into account grey/white matter segmentation or directly the normalized MRI. For instance, we could consider integrating this data into new input channels for CNN-based approaches. Finally, in order to take advantage of the large unlabeled databases currently available, a semi-supervised strategy would be particularly attractive to better represent the variability of the cortical folds.

In the near future, considering that the labeling model seems sufficiently reliable to us, we would like to reconsider the number of sulci used in the nomenclature on the basis of the sulci most often confused by the model. Indeed, the error rates of some small sulci are still too high to be used in morphological studies. By allowing the user to choose the level of granularity of the nomenclature, he will be able to use sufficiently stable labeling of the structures of interest to him.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 785907 (HBP SGA2), No. 720270 (HBP SGA1) and No. 604102 (HBP's ramp-up phase), and from the FRM DIC20161236445.

Appendix

Hyperparameters

SPAM SPAM training does not require learning hyperparameters in inner cross validation: the normalization parameters are directly evaluated on the training set.

PMAS Three hyperparameters are estimated using grid search with inner cross validation: number of ANNs selected (2, 4, 6, 8 or 10), patch sizes (26, 34 or 42 mm) and their combinations (all combinations of 1, 2, and 3 different patch sizes are tested). Note that two OPAL parameters were set a priori: the size of the search window at 14*14*14 mm and the number of iterations of the propagation and random search steps at 4.

The performances obtained by the different hyperparameter values tested during the inner cross validation were presented and commented earlier in the Figure 5.19.

HPMAS Here, only one hyperparameter seemed necessary to be tested with an inner cross-validation: the number of virtual patches selected by sulci (from 5 to 20). Note that the number of sulci per virtual patch could also be considered as an hyperparameter, but trying more than two sulci per virtual patch would be too expensive.

Concerning the selected values during inner cross-validation, the Figure 5.20 shows that the proposed range of values is sufficiently large. It might be interesting to set up a different number of selected Nearest Neighbors (NNs) per virtual patch type, which was too expensive to test in cross-validation for this thesis.

PCNN and UNET During inner cross validation, each fold contains about 62 * 9/10 (for the outer 10-folds cross validation training set) * 1/3 (for the inner 3-folds cross validation test set) $\approx$ 18 subjects. The training set therefore contains 2*18 = 36 subjects. Ideally, in order to perform early stopping, part of these 36 subjects should be reserved to evaluate the performance of the neural network at each epoch, and train on the others.

However, in order to maintain a training set of sufficient size, the validation fold containing 18 subjects is used to test the model at each epoch and the remaining 36 subjects are used for training.

In order to choose the hyperparameters (learning rate and momentum), several cross-validation loops are performed in turn to adjust the hyperparameters one after the other. First, the momentum is set at 0.9 and the learning rates 1e-2, 1e-3 and 1e-4 are tested in turn. Second, once the learning rates have been tested on all inner cross-validation folds, the learning rate is refined around the value lr that obtained the best E_{SI}: the values $lr/4$, $lr/2$, $lr*2$ and $lr*4$ are tested in turn. Finally, the best learning rate obtained is then tested with momentum 0.8, 0.7 and 0.6 to select the best value. We have chosen not to test values above 0.9 in order to reduce the risk of divergence due to excessive gradient memory.

Concerning the hyperparameters values selected for UNET and PCNN approaches (Figure 5.21 and Figure 5.22), we observe that the learning rate and momentum vary slightly from one fold to another of the cross-validation but the values retained are rather stable.

Cutting threshold Concerning the threshold on Calinski-Harabaz index used to either cut the elementary folds or not, this is determined once all the other hyperparameters have been fixed. Indeed, we consider the value of this threshold independent of the hyperparameters used for voxel-wise labeling. In order to set the cutting threshold, the following strategy was used: the thresholds 10, 100, 1000 and 10000 are tested in turn. Once these values have been tested on all inner cross-validation folds, the threshold is refined around the value th that obtained the best E_{SI}: the values $th/4$, $th/2$, $th*2$ and $th*4$ are tested in turn. For each model, the values selected per cross-validation fold are represented in Figure 5.23.

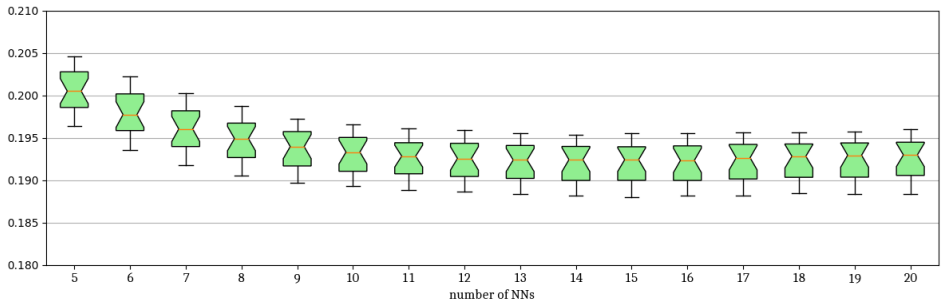


Figure 5.20 – Hyperparameter values for HPMAS model. During inner cross-validation, the number of Nearest Neighbors (NNs) selected by virtual patch type is set to minimize the average E_{SI}. This error rate is calculated based on the segmentation used to make the manual labeling. The number of NNs is set between 5 and 20. In this graph, the average E_{SI} are represented for each value tested. An external cross validation of 10 folds being done per hemisphere, 2*10=20 average E_{SI} are represented per value. The box extends from the lower to upper quartile error values, with a line at the median. The whiskers extend from the box to show the minimum and maximum values.

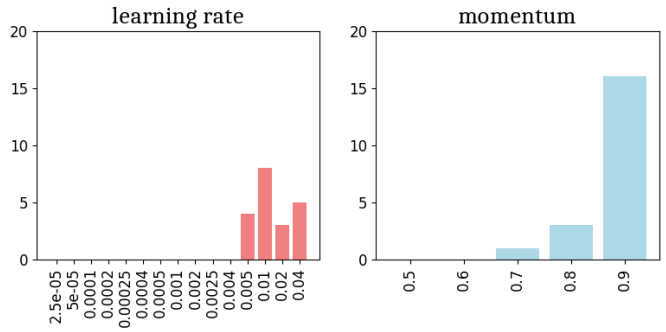


Figure 5.21 – Hyperparameters values for PCNN model. The learning rate is represented in red and the momentum in blue. Each graph represents the number of times each tested value is selected. An external cross validation of 10 folds being done per hemisphere, the values can be selected up to 2*10=20 times.

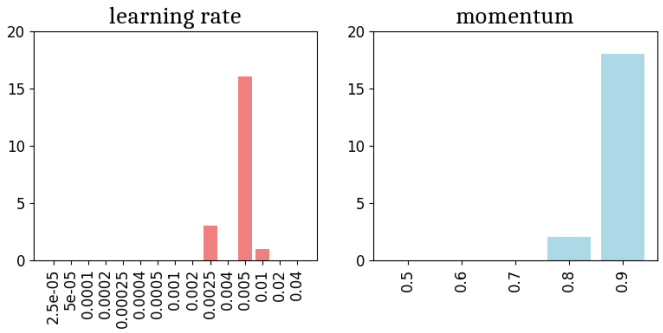


Figure 5.22 – Hyperparameters values for UNET model. The learning rate is represented in red and the momentum in blue. Each graph represents the number of times each tested value is selected. An external cross validation of 10 folds being done per hemisphere, the values can be selected up to 2*10=20 times.

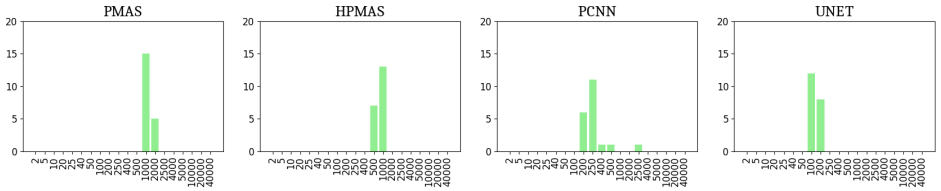


Figure 5.23 – Cutting threshold for each model (PMAS, HPMAS, PCNN, and UNET). Each graph represents the number of times each tested value is selected. An external cross validation of 10 folds being done per hemisphere, the values can be selected up to 2*10=20 times.

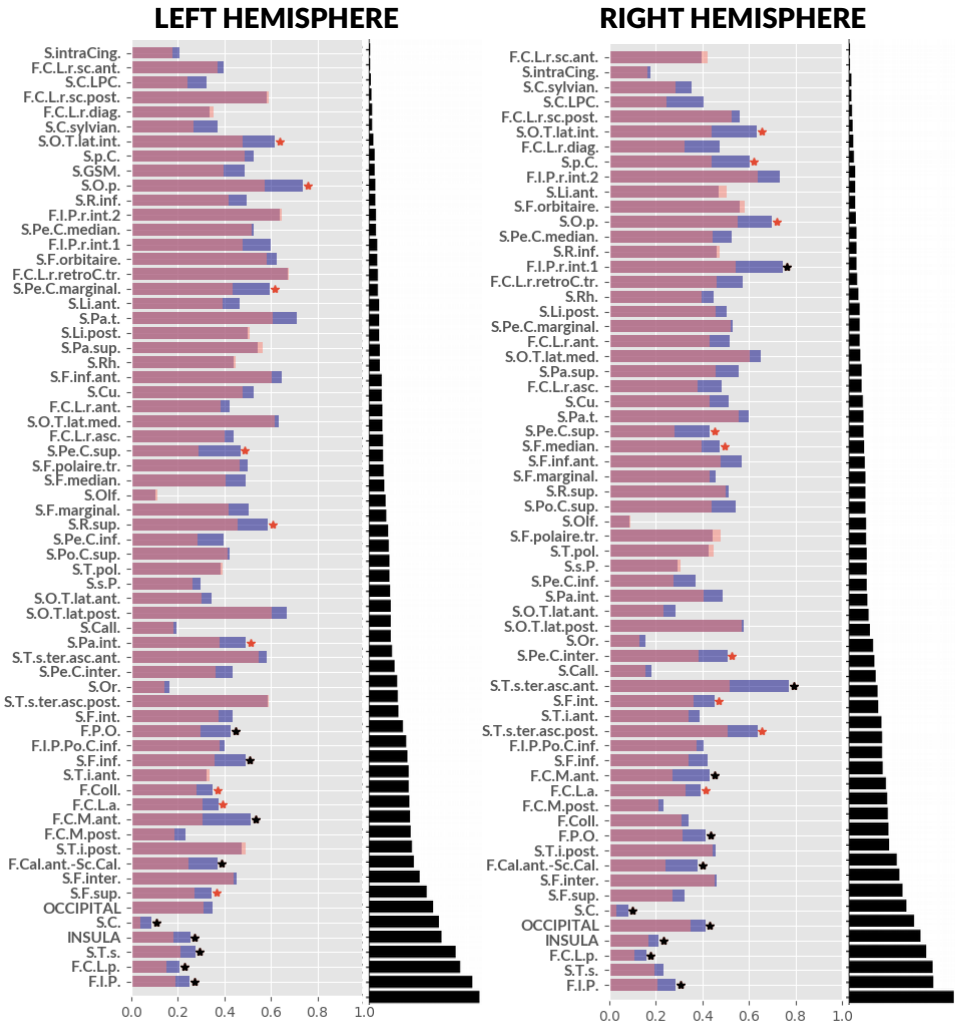


Figure 5.24 – E_{local}^{max} per sulcus. The graph on the left and the graph on the right present E_{local}^{max} for the sulci on the left hemisphere and on the right hemisphere, respectively. The SPAM model is represented in blue and the new model is represented in pink. The significant differences ($pvalue < 0.05$) are marked with a star. The star is black when the difference is still significant after controlling the false discovery rate through the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Sulci are sorted from top to bottom, from the smallest to the largest. The average sulci sizes, ranging from about 15 voxels to more than 2000 voxels per subject, are represented on the black graph.

Table 5.1 – E_{SI}^{mean} (standard deviation) in % for each model presented. The lowest error rates are shown in bold. Note that compared to (Perrot et al., 2011), the left and right hemispheres have similar error rates while the SPAM model was more accurate in recognizing the right hemisphere than the left hemisphere before the new iteration of the database labeling.

	Left hemisphere	Right hemisphere
SPAM	18.44 (2.90)	18.48 (3.13)
PMAS	17.25 (2.16)	17.68 (2.24)
PMAS + reg.	15.92 (2.26)	16.08 (2.83)
PMAS + reg. + cut.	15.32 (2.40)	15.61 (2.71)
HPMAS	18.64 (2.12)	19.12 (2.27)
HPMAS + reg.	15.74 (2.71)	15.66 (3.04)
HPMAS + reg. + cut.	15.09 (2.57)	15.39 (2.79)
PCNN	17.60 (2.29)	17.41 (2.48)
PCNN + reg.	15.61 (2.43)	15.19 (2.94)
PCNN + reg. + cut.	15.27 (2.38)	15.12 (2.93)
UNET	16.83 (2.35)	16.90 (2.34)
UNET + reg.	15.78 (2.67)	15.31 (2.79)
UNET + reg. + cut.	**15.07 (2.50)**	**15.04 (2.61)**

Table 5.2 – E_{SI}^{max} (standard deviation) in % for each model presented. The lowest error rates are shown in bold.

	Left hemisphere	Right hemisphere
SPAM	20.72 (3.00)	20.88 (3.63)
PMAS	18.22 (2.21)	18.66 (2.38)
PMAS + reg.	17.97 (2.39)	18.29 (2.97)
PMAS + reg. + cut.	17.24 (2.65)	17.53 (2.93)
HPMAS	19.14 (2.18)	19.64 (2.34)
HPMAS + reg.	17.26 (2.86)	17.34 (3.23)
HPMAS + reg. + cut.	**16.48 (2.76)**	16.89 (2.90)
PCNN	18.24 (2.43)	17.98 (2.55)
PCNN + reg.	17.37 (2.65)	16.96 (3.23)
PCNN + reg. + cut.	17.08 (2.69)	16.75 (3.13)
UNET	17.70 (2.46)	17.69 (2.39)
UNET + reg.	17.73 (2.82)	17.21 (3.08)
UNET + reg. + cut.	16.80 (2.67)	**16.65 (2.73)**

Table 5.3 – E_{local}^{mean} (standard deviation) in % for the SPAM and UNET model (first part). The UNET model presented corresponds to the most sophisticated model (UNET + reg. + cut.).

	Left hemisphere		Right hemisphere	
	SPAM	UNET	SPAM	UNET
F.C.L.a.	25.66 (10.36)	23.13 (7.38)	27.32 (10.67)	24.59 (9.56)
F.C.L.p.	15.16 (7.58)	11.19 (5.99)	11.52 (5.67)	8.02 (3.96)
F.C.L.r.ant.	32.68 (31.01)	28.89 (28.66)	36.78 (30.70)	30.33 (27.90)
F.C.L.r.asc.	33.71 (33.02)	28.77 (28.02)	32.77 (28.02)	26.81 (23.14)
F.C.L.r.diag.	28.82 (41.69)	30.41 (42.91)	35.06 (41.71)	22.96 (35.89)
F.C.L.r.retroC.tr.	52.27 (37.45)	53.44 (38.85)	39.59 (36.01)	36.33 (35.57)
F.C.L.r.sc.ant.	31.63 (42.24)	30.78 (42.72)	33.42 (43.00)	36.63 (44.58)
F.C.L.r.sc.post.	40.72 (38.98)	39.75 (37.62)	38.43 (36.68)	30.18 (30.70)
F.C.M.ant.	36.69 (23.48)	20.72 (13.19)	32.70 (31.24)	19.04 (14.52)
F.C.M.post.	16.07 (15.04)	11.86 (10.39)	16.33 (13.48)	15.30 (12.46)
F.Cal.ant.-Sc.Cal.	24.13 (15.91)	15.44 (7.30)	22.77 (12.58)	16.22 (9.08)
F.Coll.	23.77 (13.37)	19.21 (12.47)	23.21 (15.33)	22.20 (16.04)
F.I.P.	20.11 (8.79)	14.32 (8.93)	22.40 (10.42)	16.01 (8.72)
F.I.P.Po.C.inf.	29.16 (18.59)	28.05 (15.19)	27.62 (16.20)	27.51 (18.29)
F.I.P.r.int.1	46.75 (38.23)	33.44 (37.25)	61.44 (35.38)	40.87 (40.52)
F.I.P.r.int.2	50.25 (41.79)	51.97 (42.58)	64.66 (41.14)	56.33 (44.15)
F.P.O.	32.48 (22.36)	21.31 (15.91)	27.21 (19.18)	21.11 (15.22)
INSULA	16.10 (7.17)	13.38 (5.69)	13.54 (5.36)	12.30 (5.53)
OCCIPITAL	27.99 (12.08)	24.53 (12.02)	33.44 (10.37)	27.61 (11.45)
S.C.	5.77 (7.17)	2.24 (4.25)	5.12 (5.65)	1.68 (2.82)
S.C.LPC.	22.22 (37.00)	17.95 (34.36)	30.58 (42.05)	17.67 (33.28)
S.C.sylvian.	25.87 (36.17)	17.00 (29.65)	27.84 (39.74)	19.54 (32.65)
S.Call.	9.60 (13.42)	8.57 (9.24)	9.22 (15.75)	8.58 (15.70)
S.Cu.	37.30 (27.12)	29.63 (24.63)	35.46 (24.64)	29.61 (28.79)
S.F.inf.	38.16 (22.14)	26.51 (16.47)	32.67 (21.55)	26.06 (20.01)
S.F.inf.ant.	51.40 (34.22)	43.30 (32.83)	41.08 (28.87)	34.47 (30.21)
S.F.int.	31.70 (18.47)	26.43 (15.30)	35.04 (21.95)	26.96 (17.88)
S.F.inter.	36.53 (16.52)	33.44 (14.26)	38.98 (19.73)	36.24 (17.90)
S.F.marginal.	35.58 (25.33)	28.21 (21.92)	32.60 (24.70)	29.90 (21.22)
S.F.median.	39.34 (24.65)	32.15 (23.84)	34.91 (22.10)	28.63 (23.00)
S.F.orbitaire.	47.63 (38.56)	45.63 (40.02)	39.77 (34.66)	41.84 (36.85)
S.F.polaire.tr.	37.90 (22.02)	32.16 (20.69)	31.51 (20.11)	34.19 (20.19)
S.F.sup.	26.86 (15.84)	19.59 (12.21)	25.40 (14.10)	19.94 (12.66)
S.GSM.	39.26 (43.46)	26.11 (35.68)		

Table 5.4 – E_{local}^{mean} (standard deviation) in % for the SPAM and UNET model (second part). The UNET model presented corresponds to the most sophisticated model (UNET + reg. + cut.).

	Left hemisphere		Right hemisphere	
	SPAM	UNET	SPAM	UNET
S.Li.ant.	27.14 (23.68)	23.97 (25.57)	28.59 (29.77)	31.74 (32.25)
S.Li.post.	34.92 (24.12)	30.83 (23.68)	31.19 (24.36)	29.61 (24.14)
S.O.T.lat.ant.	21.40 (19.38)	18.87 (17.44)	18.61 (21.37)	16.14 (20.83)
S.O.T.lat.int.	48.61 (38.83)	34.57 (37.24)	48.71 (38.85)	36.54 (40.81)
S.O.T.lat.med.	47.27 (28.51)	45.66 (28.36)	47.57 (31.52)	42.29 (26.33)
S.O.T.lat.post.	52.24 (26.21)	46.54 (27.75)	44.62 (25.00)	43.31 (22.06)
S.O.p.	58.36 (37.25)	43.34 (40.11)	49.22 (35.83)	35.85 (35.23)
S.Olf.	5.71 (6.58)	6.33 (7.34)	4.85 (5.82)	4.32 (5.00)
S.Or.	10.88 (8.60)	9.02 (6.81)	10.03 (8.47)	8.27 (7.81)
S.Pa.int.	36.08 (26.53)	29.09 (27.24)	34.13 (22.13)	27.47 (23.63)
S.Pa.sup.	42.59 (39.58)	45.54 (38.85)	43.89 (39.23)	36.34 (39.65)
S.Pa.t.	56.15 (37.69)	47.52 (39.65)	46.04 (35.71)	41.30 (34.50)
S.Pe.C.inf.	29.45 (35.57)	19.75 (24.68)	25.70 (34.09)	19.45 (20.58)
S.Pe.C.inter.	34.06 (27.53)	27.40 (23.75)	41.41 (25.79)	30.50 (24.12)
S.Pe.C.marginal.	48.64 (36.36)	31.35 (29.89)	39.60 (31.07)	40.23 (31.90)
S.Pe.C.median.	42.36 (38.24)	38.73 (38.55)	37.96 (38.29)	35.45 (36.69)
S.Pe.C.sup.	36.07 (37.95)	21.42 (27.72)	33.23 (33.99)	21.44 (30.66)
S.Po.C.sup.	30.41 (25.01)	31.60 (23.57)	40.69 (31.26)	29.23 (22.56)
S.R.inf.	36.06 (38.23)	31.46 (38.93)	29.20 (34.95)	27.12 (31.48)
S.R.sup.	41.75 (33.43)	30.51 (27.95)	36.87 (34.94)	34.54 (29.84)
S.Rh.	26.86 (25.33)	26.31 (25.10)	24.22 (23.97)	21.45 (23.58)
S.T.i.ant.	24.58 (18.99)	25.21 (20.72)	28.66 (16.55)	23.83 (17.23)
S.T.i.post.	37.82 (19.97)	37.28 (21.67)	36.75 (14.78)	34.75 (17.81)
S.T.pol.	27.87 (22.89)	27.25 (22.28)	30.05 (22.22)	29.56 (21.13)
S.T.s.	20.73 (11.79)	15.46 (8.83)	15.66 (10.99)	13.60 (9.06)
S.T.s.ter.asc.ant.	46.12 (35.80)	43.00 (29.56)	64.44 (34.97)	41.41 (29.32)
S.T.s.ter.asc.post.	46.90 (30.21)	43.58 (29.81)	49.04 (29.41)	41.93 (28.55)
S.intraCing.	16.79 (35.76)	15.78 (35.03)	16.76 (36.26)	14.22 (33.59)
S.p.C.	40.72 (38.83)	36.36 (38.00)	48.13 (39.89)	34.94 (39.42)
S.s.P.	20.31 (17.64)	17.87 (15.84)	18.96 (15.04)	20.23 (16.87)

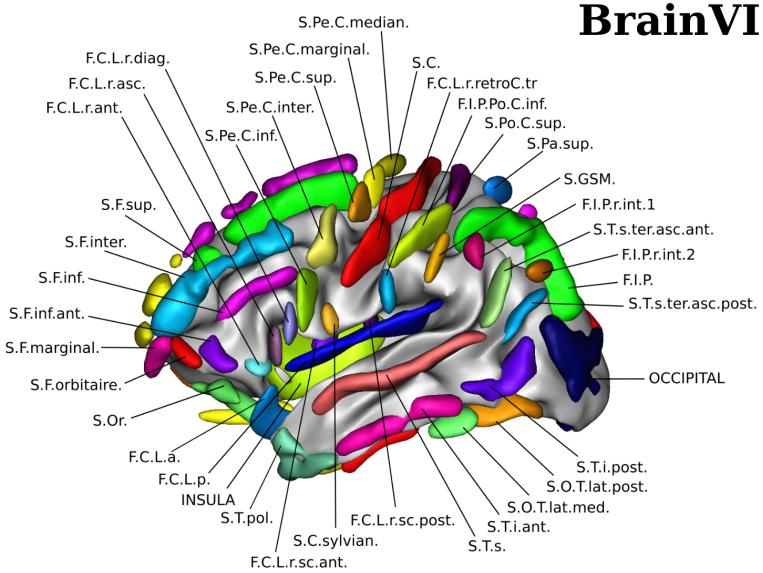
Table 5.5 – E_{local}^{max} (standard deviation) in % for the SPAM and UNET model (first part). The UNET model presented corresponds to the most sophisticated model (UNET + reg. + cut.).

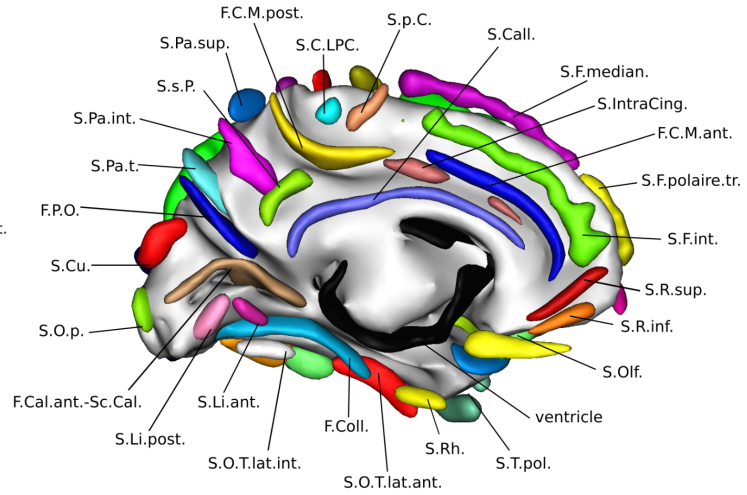
	Left hemisphere		Right hemisphere	
	SPAM	UNET	SPAM	UNET
F.C.L.a.	37.44 (18.33)	30.67 (8.97)	39.37 (17.51)	32.79 (11.25)
F.C.L.p.	20.96 (8.68)	15.23 (7.68)	15.96 (7.06)	10.89 (5.25)
F.C.L.r.ant.	42.48 (32.88)	38.44 (31.93)	51.92 (35.33)	43.03 (33.75)
F.C.L.r.asc.	43.88 (33.16)	39.97 (31.88)	48.36 (33.72)	37.82 (27.33)
F.C.L.r.diag.	33.64 (45.93)	35.60 (46.38)	47.43 (48.21)	32.44 (42.94)
F.C.L.r.retroC.tr.	67.47 (37.66)	67.74 (37.01)	57.49 (40.48)	46.32 (38.32)
F.C.L.r.sc.ant.	39.79 (47.79)	37.32 (46.81)	39.90 (47.16)	42.39 (49.09)
F.C.L.r.sc.post.	58.10 (42.97)	59.32 (46.02)	56.18 (40.25)	52.56 (40.83)
F.C.M.ant.	51.49 (26.45)	30.69 (16.01)	43.29 (34.88)	27.18 (16.63)
F.C.M.post.	23.32 (19.68)	18.71 (12.82)	23.46 (17.26)	21.27 (14.13)
F.Cal.ant.-Sc.Cal.	36.96 (19.94)	24.67 (10.65)	38.02 (20.10)	24.05 (11.53)
F.Coll.	35.17 (17.73)	28.26 (15.70)	34.35 (19.96)	31.21 (17.37)
F.I.P.	25.14 (9.28)	19.01 (9.55)	28.54 (11.93)	20.95 (10.51)
F.I.P.Po.C.inf.	39.98 (20.94)	38.11 (16.81)	40.54 (20.41)	37.47 (17.88)
F.I.P.r.int.1	60.13 (39.38)	48.03 (41.51)	74.52 (31.71)	54.16 (40.62)
F.I.P.r.int.2	63.94 (41.94)	64.79 (43.51)	73.34 (41.27)	63.66 (43.62)
F.P.O.	42.69 (22.20)	30.05 (17.20)	41.51 (24.14)	31.35 (18.60)
INSULA	25.37 (12.09)	18.24 (7.60)	21.28 (9.65)	16.83 (6.90)
OCCIPITAL	34.93 (12.60)	31.23 (13.48)	41.32 (10.78)	34.97 (11.50)
S.C.	8.93 (10.37)	3.87 (5.93)	8.25 (8.08)	3.11 (3.88)
S.C.LPC.	32.61 (45.54)	24.31 (41.49)	40.75 (47.21)	24.83 (41.10)
S.C.sylvian.	37.01 (41.30)	26.83 (38.71)	35.52 (42.87)	28.59 (39.71)
S.Call.	19.67 (21.39)	18.10 (18.06)	18.14 (22.39)	15.42 (18.20)
S.Cu.	52.47 (26.15)	47.79 (29.62)	51.57 (27.74)	42.98 (33.64)
S.F.inf.	49.08 (24.32)	35.71 (17.06)	42.19 (24.50)	34.22 (21.53)
S.F.inf.ant.	64.77 (31.99)	60.22 (31.52)	57.04 (30.38)	47.86 (32.70)
S.F.int.	43.45 (19.27)	37.71 (16.49)	45.21 (22.87)	36.37 (19.06)
S.F.inter.	45.36 (16.72)	44.01 (15.95)	46.02 (19.90)	45.38 (19.00)
S.F.marginal.	50.58 (28.20)	41.81 (25.18)	45.93 (27.81)	43.09 (22.19)
S.F.median.	49.22 (26.35)	40.60 (24.96)	47.65 (22.45)	39.53 (23.32)
S.F.orbitaire.	62.47 (39.57)	58.14 (41.25)	56.02 (37.85)	58.22 (38.93)
S.F.polaire.tr.	50.11 (21.88)	46.64 (19.88)	44.27 (20.17)	47.82 (22.48)
S.F.sup.	34.55 (17.81)	27.47 (14.06)	32.47 (15.85)	27.26 (12.67)
S.GSM.	48.89 (46.66)	39.70 (44.85)		

Table 5.6 – E_{local}^{max} (standard deviation) in % for the SPAM and UNET model (second part). The UNET model presented corresponds to the most sophisticated model (UNET + reg. + cut.).

	Left hemisphere		Right hemisphere	
	SPAM	UNET	SPAM	UNET
S.Li.ant.	46.77 (31.32)	39.23 (32.40)	47.17 (34.90)	50.45 (37.41)
S.Li.post.	50.02 (25.06)	50.91 (29.18)	50.40 (29.25)	45.57 (28.76)
S.O.T.lat.ant.	34.74 (24.66)	30.46 (22.24)	28.67 (25.67)	23.57 (24.57)
S.O.T.lat.int.	61.82 (36.85)	47.89 (41.19)	63.44 (39.75)	44.06 (42.91)
S.O.T.lat.med.	63.47 (26.72)	61.78 (27.77)	65.01 (29.23)	60.29 (28.97)
S.O.T.lat.post.	66.93 (25.45)	60.52 (27.96)	57.61 (24.63)	56.84 (22.26)
S.O.p.	73.90 (37.20)	57.32 (43.75)	69.86 (35.12)	55.17 (39.86)
S.Olf.	10.38 (9.27)	11.49 (10.17)	8.92 (8.09)	8.98 (7.99)
S.Or.	16.65 (9.89)	14.35 (9.12)	15.81 (10.56)	13.09 (9.23)
S.Pa.int.	49.23 (30.40)	37.83 (28.98)	48.88 (25.01)	40.57 (26.46)
S.Pa.sup.	54.34 (42.12)	56.46 (42.46)	55.61 (39.15)	45.75 (41.18)
S.Pa.t.	71.09 (38.08)	61.04 (41.36)	60.11 (38.63)	55.71 (37.68)
S.Pe.C.inf.	39.54 (40.51)	28.41 (27.95)	37.35 (39.79)	27.66 (25.44)
S.Pe.C.inter.	43.50 (30.11)	36.21 (28.06)	50.71 (28.83)	38.33 (26.02)
S.Pe.C.marginal.	59.65 (37.21)	43.76 (32.19)	53.28 (35.45)	52.24 (33.05)
S.Pe.C.median.	52.72 (39.69)	51.91 (40.16)	52.79 (41.55)	44.48 (41.53)
S.Pe.C.sup.	46.98 (40.14)	29.06 (31.49)	43.38 (40.26)	27.97 (33.22)
S.Po.C.sup.	42.37 (29.42)	41.30 (25.69)	54.54 (30.70)	44.22 (25.46)
S.R.inf.	49.75 (41.11)	42.08 (43.04)	46.04 (37.06)	47.38 (37.82)
S.R.sup.	58.83 (35.13)	45.71 (31.38)	51.32 (36.39)	49.98 (34.36)
S.Rh.	44.24 (29.34)	44.69 (30.83)	44.79 (31.43)	39.83 (30.67)
S.T.i.ant.	32.21 (20.12)	33.92 (21.44)	38.68 (15.17)	34.15 (18.73)
S.T.i.post.	47.59 (21.94)	49.11 (22.06)	45.90 (15.37)	44.54 (18.57)
S.T.pol.	38.62 (22.84)	39.10 (24.45)	42.90 (24.29)	45.09 (24.03)
S.T.s.	27.60 (13.01)	21.41 (10.31)	23.32 (14.08)	19.60 (10.28)
S.T.s.ter.asc.ant.	58.07 (35.61)	54.68 (29.33)	77.34 (32.05)	51.74 (29.64)
S.T.s.ter.asc.post.	58.85 (31.47)	59.24 (30.70)	63.76 (31.00)	51.08 (27.90)
S.intraCing.	20.97 (40.71)	17.74 (38.20)	17.74 (38.20)	16.36 (36.72)
S.p.C.	52.84 (43.61)	48.98 (43.22)	60.26 (42.34)	44.19 (44.58)
S.s.P.	29.88 (20.71)	26.42 (16.87)	29.52 (19.34)	30.81 (22.40)

BrainVISA Sulci Nomenclature v.2019





	F.C.L.a.	anterior lateral f ssure
	F.C.L.p.	posterior lateral f ssure
	F.C.L.r.ant.	anterior ramus of the lateral f ssure
	F.C.L.r.asc.	ascending ramus of the lateral f ssure
	F.C.L.r.diag.	diagonal ramus of the lateral f ssure
	F.C.L.r.retroC.tr.	retro central transverse ramus of the lateral f ssure
	F.C.L.r.sc.ant.	anterior sub-central ramus of the lateral f ssure
	F.C.L.r.sc.post.	posterior sub-central ramus of the lateral f ssure
	F.C.M.ant.	calloso-marginal anterior f ssure
	F.C.M.post.	calloso-marginal posterior f ssure
	F.Cal.ant.-Sc.Cal.	calcarine f ssure
	F.Coll.	collateral f ssure
	F.I.P.Po.C.inf.	superior postcentral intraparietal superior sulcus
	F.I.P.	intraparietal sulcus
	F.I.P.r.int.1	primary intermediate ramus of the intraparietal sulcus
	F.I.P.r.int.2	secondary intermediate ramus of the intraparietal sulcus
	F.P.O.	parieto-occipital f ssure
	INSULA	insula
	OCCIPITAL	lobe occipital
	S.C.LPC.	paracentral lobule central sulcus
	S.C.	central sulcus
	S.C.sylvian.	central sylvian sulcus

	S.Call.	subcallosal sulcus
	S.Cu.	cuneal sulcus
	S.F.inf.	inferior frontal sulcus
	S.F.inf.ant.	anterior inferior frontal sulcus
	S.F.int.	internal frontal sulcus
	S.F.inter.	intermediate frontal sulcus
	S.F.marginal.	marginal frontal sulcus
	S.F.median.	median frontal sulcus
	S.F.orbitaire.	orbital frontal sulcus
	S.F.polaire.tr.	polar frontal sulcus
	S.F.sup.	superior frontal sulcus
	S.GSM.	sulcus of the supra-marginal gyrus
	S.intraCing.	intracingular sulcus
	S.Li.ant.	anterior intralingual sulcus
	S.Li.post.	posterior intra-lingual sulcus
	S.O.T.lat.ant.	anterior occipito-temporal lateral sulcus
	S.O.T.lat.int.	internal occipito-temporal lateral sulcus
	S.O.T.lat.med.	median occipito-temporal lateral sulcus
	S.O.T.lat.post.	posterior occipito-temporal lateral sulcus
	S.O.p.	occipito-polar sulcus
	S.Olf.	olfactory sulcus
	S.Or.	orbital sulcus

	S.Pa.int.	internal parietal sulcus
	S.Pa.sup.	superior parietal sulcus
	S.Pa.t.	transverse parietal sulcus
	S.Pe.C.inf.	inferior precentral sulcus
	S.Pe.C.inter.	intermediate precentral sulcus
	S.Pe.C.marginal.	marginal precentral sulcus
	S.Pe.C.median.	median precentral sulcus
	S.Pe.C.sup.	superior precentral sulcus
	S.Po.C.sup.	superior postcentral sulcus
	S.R.inf.	inferior rostral sulcus
	S.R.sup.	superior rostral sulcus
	S.Rh.	rhinal sulcus
	S.T.i.ant.	anterior inferior temporal sulcus
	S.T.i.post.	posterior inferior temporal sulcus
	S.T.pol.	polar temporal sulcus
	S.T.s.	superior temporal sulcus
	S.T.s.ter.asc.ant.	anterior terminal ascending branch of the superior temporal sulcus
	S.T.s.ter.asc.post.	posterior terminal ascending branch of the superior temporal sulcus
	S.p.C.	paracentral sulcus
	S.s.P.	sub-parietal sulcus
	ventricle	ventricle
	unknown	unknown cortical folds

L. Borne, D. Rivière, M. Mancip, J.-F. Mangin. Medical Image Analysis (2019)

Conclusion

Contributions

Throughout this thesis, I developed two types of tools: the first for the automatic recognition of local patterns of cortical folds and the second for the automatic sulci labeling.

Concerning the automatic recognition of local folding patterns, two types of patterns were studied: the paracingulate sulcus, a pattern present in nearly half of the population, and the Power Button Sign (PBS), a considerably rarer pattern. The proposed tools allow to recognize hemispheres with paracingulate sulcus or PBS with a balanced accuracy of about 80% and 60% respectively. In the first case, the proposed tools enable the automatic classification of the pattern on very large databases. Additionally, the tools provide a reliability score to manually verify the hemisphere on which the model is least confident. In the second case, the tools help to search for the rare pattern on large databases in order to enlarge the current database or to associate them with specific cognitive characteristics.

Concerning the automatic labeling of cortical sulci, the new model is more efficient (with a recognition rate of around 85% instead of 81%) and faster (it labels a hemisphere in only 20 seconds instead of 5 minutes). I have demonstrated that this improvement is sufficient to reproduce observations previously only possible with manual labeling. But above all, the great novelty of this new model is its robustness to under-segmentation errors. Indeed, the main errors of the previous model are due to an insufficient fragmentation of the fold skeleton into elementary folds, forcing the model to label an elementary fold with a single label when it actually contains two labels. The new model is capable of automatically re-dividing the elementary folds when implied by voxelwise labeling.

Limitations

- **Data** Currently, the manual and automatic labeling of sulci or of local folds patterns, is only based on structural MRIs. However, central sulcus labeling sometimes benefits the use of functional data for ambigu-

ous configurations. Additionally, the underlying architecture of neural networks is now available through tractography, which could particularly help to make labeling more anatomically consistent. Therefore, the combination of structural, functional and diffusion MRIs would be particularly interesting, both for manual and automatic labeling.

- **Deep learning** As a first application of 3D convolutional neural networks for the study of cortical folds, the results obtained are particularly impressive. However, deep learning is a hot topic nowadays and new techniques are continuously being published. Therefore, the use of these recent techniques could also improve current performance (e.g., using deeper networks, using atrous convolutions for segmentation (Chen et al., 2018), using tranfer learning to better represent folds variability, etc.).

- **Unsupervised learning** The extreme variability of sulci makes them fascinating structures but also particularly difficult to study. The definition of folding patterns and the implementation of a nomenclature are indeed subject to interpretation. In this thesis, a new iteration of labeling of the database was carried out in order to make it coherent thanks to the screen wall, allowing to visualize all brains simultaneously. However, this task was particularly difficult, especially for small and more variable sulci. This problem suggests that humans alone are not able to understand their complex organization because of their limited ability to represent all possible configurations. Therefore, supervised learning is probably limited by a human interpretation of this variability, which is sometimes inconsistent. However, current advances in unsupervised learning suggest that it is time to let the computer, with its memory and computer capabilities far superior than ours, understand this organization for us.

Perspectives

In the continuity of my thesis work, approaches based on unsupervised learning seem to be the most promising to understand cortical fold organization. Indeed, with the thousands of brain MRIs now available, this colossal amount of data leaves the door wide open for ambitious studies.

First, some of the proposed methods for supervised learning of local sulcal pattern classification would, in my opinion, be easily adaptable to the search for new patterns that characterize a group. For example, fold skeletons could be used to train a model to distinguish left-handed from right-handed, women from men, etc. If the model achieves significant classification performance, it becomes particularly interesting to look at the

patterns that allow it to perform a correct classification. For example, since the SNIPE method is based on the similarity of each patch surrounding the voxels of a fold skeleton to both populations in the learning base, it is then easy to visually identify, for a given skeleton, the sulcal pattern specific to the assigned group. Similarly, by training a CNN to make such classifications, different techniques could help identify areas of the fold skeleton that allow correct classification, such as the use of saliency maps (Simonyan et al., 2013) or occlusions (Zeiler and Fergus, 2014), already used to understand how neural networks, used for 2D image classification, work. Such tools would help neuroanatomy experts to identify fold patterns specific to a given group.

Concerning the unsupervised definition of sulci, this ambitious project raises some unusual questions. Indeed, it is simultaneously necessary to group the voxels of the same subject together to form a sulcus and find the correspondence of a group of voxels from one subject to another to create a sulcus label. One alternative would be to use manual labeling as a proxy. For example, it would be interesting to see if the UNET model trained during this thesis gets better performance by being re-trained from scratch on a database that it has labeled itself. Ideally, this approach would eliminate the inconsistencies of manual labeling and achieve a new consistent labeling across the entire database. Nevertheless, this method depends heavily on the proxy used and would probably be subject to overfitting.

In order to avoid these two problems, I think this project could be carried out with a patch clustering approach. The idea is intuitive: if you need to group voxels together in the same subject to form a sulcus and between subjects to define a sulcus label, why not cluster the voxels concerned according to the patches surrounding them? The chosen clustering algorithm will have to group together voxels belonging to the same elementary folds and those that have similar patches between subjects. These two conditions make it difficult to define the similarity between two patches. To this end, deep learning can be used to learn the optimal representation of patches, guided by a chosen unsupervised objective. For example, the methods proposed in (Xie et al., 2016; Yang et al., 2016, 2017) jointly optimizes dimension reduction and clustering. One of these methods has already been adapted to the unsupervised segmentation of medical images (Moriya et al., 2018), by clustering the images patches. Given the promising results obtained with the use of deep learning and patch learning approaches for the supervised study of cortical folds, the adaptation of these methods to the unsupervised study of cortical folds also seems to be a promising approach.

Closing remarks

The models developed in this thesis are part of a range of tools for studying cortical folds. These tools are mainly based on the manual or automatic labeling of sulci, allowing for example the study of their characteristics (depth, length, opening, etc.), the registration of the main sulci (Auzias et al., 2009), etc. The models compared in this thesis for the automatic classification of fold patterns and the automatic recognition of sulci, once again show the incredible effectiveness of deep learning. As this area is constantly evolving, it is highly probable that recent developments could improve the performance achieved in this thesis. However, the use of a nomenclature for sulci labeling has its limitations when it comes to representing all possible folding configurations. This is why there is currently no consensus among experts on a nomenclature. Hence, it seems essential to consider the use of unsupervised techniques to fully understand their organization, without relying on an too subjective human interpretation.

Summary in French

Introduction

Considéré comme le siège de notre intelligence, le cerveau est une énigme fascinante. L'une des inconnues de cette énigme est le lien entre anatomie et organisation fonctionnelle. Avec l'apparition de l'imagerie non invasive, comme l'Imagerie par Résonance Magnétique (IRM), il y a une quarantaine d'années, il a été possible de visualiser le cerveau, d'abord anatomiquement, puis fonctionnellement. Ces outils puissants ont mis en évidence des zones correspondant à des fonctions cognitives spécifiques. Le fonctionnement de ces zones est-il lié à l'organisation anatomique du cerveau ? Concernant l'anatomie de la surface corticale, les caractéristiques les plus remarquables sont ses convolutions, dont la variabilité est telle que chaque individu a des motifs de plissement uniques. Cependant, cette caractéristique frappante est encore mal comprise aujourd'hui.

D'un point de vue darwinien, les plissements corticaux sont considérés comme une astuce de l'évolution pour augmenter la surface corticale sans modifier le volume de la cavité crânienne. Cependant, si on s'intéresse à leur mécanisme de formation, on observe que le cortex se plisse au cours du dernier trimestre de grossesse. Or, cette période correspond à l'émergence de l'architecture corticale : les neurones migrent dans différentes zones prédéfinies juste avant la formation des plis. Cette observation suggère un lien entre la formation des plis et l'organisation fonctionnelle du cortex. En effet, des motifs de plissements inhabituels sont généralement liés à des développements anormaux pouvant mener à des syndromes psychiatriques comme l'épilepsie ou la schizophrénie. De plus, un lien étroit a été démontré entre la forme de certains sillons et des fonctions cognitives telles que la latéralité manuelle ou la lecture. Par conséquent, les motifs de plis semblent être des signatures de l'organisation fonctionnelle du cerveau de chaque individu. Cette hypothèse est maintenant relativement acceptée pour les sillons les plus importants, mais elle reste controversée pour les sillons secondaires.

Quoi qu'il en soit, l'étude des plis corticaux exige un haut niveau d'expertise que peu de neuroanatomistes possèdent à l'heure actuelle. Afin de faciliter l'étude des plissements, les outils développés durant cette thèse

utilisent des algorithmes artificiellement intelligents pour automatiser les tâches nécessitant des connaissances avancées. Aujourd'hui, l'intelligence artificielle a fait des progrès significatifs, notamment dans le domaine de la vision par ordinateur, révolutionnant notre quotidien. Ceci est dû à l'émergence d'une nouvelle méthode inspirée par le fonctionnement du cerveau : l'apprentissage profond. Cette technique a permis aux smartphones de parler, aux ordinateurs de battre les champions d'échecs, aux voitures de conduire seules, etc. Permettront-ils également de déchiffrer nos lignes sulcales ?

Contributions de la thèse: Les travaux menés au cours de cette thèse ont permis l'implémentation de deux types d'outils dédiés à l'étude des plis corticaux, qui seront bientôt disponibles dans la boîte à outils BrainVISA/Morphologist (http://brainvisa.info). Le premier outil automatise la classification des motifs locaux de plissements corticaux. Cette tâche n'avait jamais été abordée auparavant. Le deuxième outil automatise l'étiquetage des sillons en modélisant des mécanismes de reconnaissance "top-down" nécessaires pour pallier les faiblesses des démarches "bottom-up" développées jusqu'à présent. Ainsi, en plus d'améliorer les performances et la rapidité d'exécution par rapport au modèle précédemment proposé par la boîte à outils BrainVISA/Morphologist, le nouveau modèle proposé est robuste aux erreurs de sous-segmentation, ce qui est l'une des plus grandes faiblesses de l'ancien système.

Classification automatique des motifs locaux de plissements corticaux

Le mécanisme de formation des plis et l'impact de cette caractéristique anatomique sur le fonctionnement du cerveau sont encore mal compris aujourd'hui. Cependant, il a été démontré que la configuration spatiale de certains sillons sulcaux, marqueurs du développement précoce du cerveau (Cachia et al., 2016), sont associés à des fonctions cérébrales spécifiques. En effet, plusieurs études ont rapporté une corrélation entre des motifs sulcaux reconnus visuellement et des caractéristiques cognitives - par exemple, l'efficacité du contrôle cognitif (Fornito et al., 2004; Cachia et al., 2014) ou la latéralité manuelle (Sun et al., 2012) - ou des maladies neuropsychiatriques - par exemple, l'épilepsie (Mellerio et al., 2014) ou la schizophrénie (Plaze et al., 2015). L'automatisation de la reconnaissance de ces motifs sulcaux permettra d'étendre et de confirmer ces études sur des bases de données plus importantes et de mieux comprendre les liens subtils entre les formes sulcales et l'architecture fonctionnelle. Dans cette thèse, l'identification automatique de deux types de motifs est abordée : les motifs du Cortex Cingulaire Antérieur (CCA) et le signe *Power Button* (Figure 1 et 2).

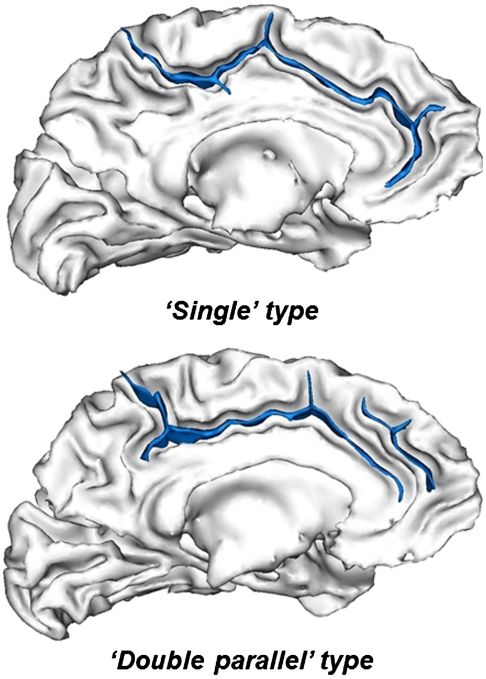


Figure 1 - Deux motifs de plis identifiés dans la région cingulaire. Le motif '*single*' est caractérisé par la présence du sillon cingulaire seul. Le motif '*double parallel*' présente un sillon paracingulaire supplémentaire, qui est parallèle au sillon cingulaire et suffisamment long pour être identifié. (Cachia et al., 2014)

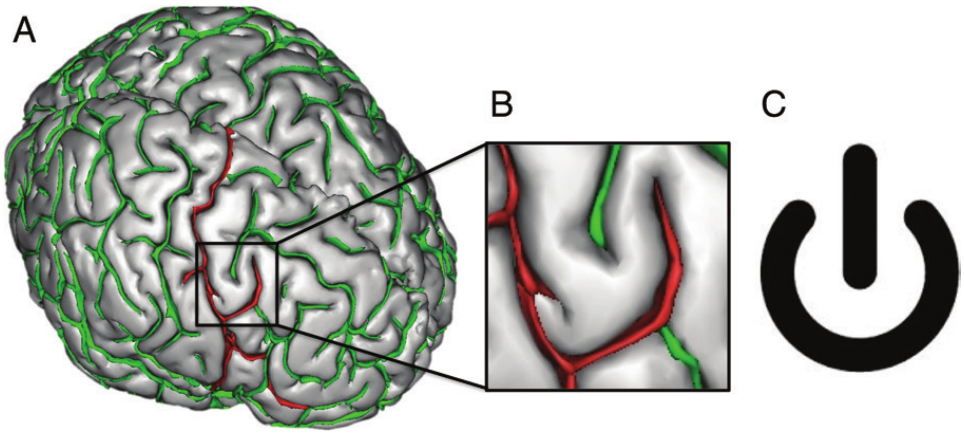


Figure 2 - Le signe *Power Button* : un motif de plis lié à l'épilepsie. Ce motif se caractérise par l'interposition d'un segment sulcal précentral (en vert) entre le sillon central (en rouge) et l'une de ses branches ascendantes antérieures en forme de crochet (en rouge). (Mellerio et al., 2014)

Base de données : Le pipeline BrainVISA/Morphologist a été utilisé pour segmenter et fournir une représentation graphique 3D à partir de données IRM brutes. Cette reconstruction 3D des plis corticaux à l'aide d'un maillage a été utilisée pour visualiser et étiqueter manuellement les cerveaux. Concernant la région du CCA, on distingue deux motifs : le motif *'single'* et le motif *'double parallel'* (Figure **??**). La base de données d'apprentissage contient 348 hémisphères, dont 141 motifs *'single'* et 207 motifs *'double parallel'*. Le signe *Power Button* est beaucoup plus rare. La base de données d'apprentissage contient 114 hémisphères, dont seulement 28 possèdent ce signe. Ce désiquilibre et la taille limitée de la base de données sont deux challenges pour les algorithmes de classification décrits ci-dessous.

Méthode : Dans cette thèse, trois méthodes ont été implémentées et comparées pour la reconnaissance automatique de ces motifs.

La première méthode compare chaque hémiphère de la base d'apprentissage possédant un sillon paracingulaire, ou un signe *Power Button*, avec l'hémisphère à classer. Si le même motif est présent dans le nouvel hémisphère, les distances des motifs de la base d'apprentissage au nouvel hémisphère devraient être plus faibles. En se basant sur cet a priori, la première méthode utilise ces distances pour entrainer un classifieur *Support Vector Machine*.

La seconde approche se base sur le *Scoring by Non-local Image Patch Estimator* (SNIPE) proposé dans (Coupé et al., 2012). Pour classer des images, le SNIPE utilise une mesure de classement basée sur un système de comparaison de patchs. Cette mesure permet d'estimer la similitude des patchs entourant les voxels de l'image à classer avec les patchs présents dans les différentes populations de la base d'apprentissage. Dans cette thèse, les "populations" de la base de données d'apprentissage correspondent à des motifs différents.

La troisième méthode utilise l'apprentissage profond, qui a permis des avancées fulgurantes dans le monde de la vision par ordinateur. Ici, un réseau de neurones récurrent 3D est entraîné à classer les motifs.

Résultats : Les outils proposés permettent de classer les hémisphères en fonction de leur motif dans la région du CCA et de la présence du signe *Power Button* avec une précision équilibrée d'environ 80% et 60% respectivement. Les performances des différents outils sont équivalentes pour les deux problèmes de reconnaissance posés, néanmoins le réseau de neurones est beaucoup plus rapide à utiliser en pratique. Concernant les motifs de la région cingulaire, les outils proposés permettent la classification automatique des motifs sur de très grandes bases de données. De plus, les outils fournissent un score de fiabilité pour vérifier manuellement les hémisphères

sur lesquels le modèle est le moins confiant. Concernant le signe *Power Button*, les outils permettent de rechercher ce motif rare sur de grandes bases de données afin d'agrandir la base de données actuelle ou de les associer à des caractéristiques cognitives spécifiques.

Étiquetage automatique des sillons corticaux

L'extrême variabilité des plissements du cortex humain rend l'étiquetage des sillons, à la fois automatique et manuel, particulièrement difficile. L'identification fiable des sillons corticaux dans leur intégralité est extrêmement difficile et n'est pratiquée que par quelques experts. De plus, ces sillons correspondent à plus d'une centaine de structures différentes, ce qui rend l'étiquetage manuel fastidieux et limite donc la possibilité d'accès à des grandes bases de données étiquetées pour entraîner des algorithmes d'apprentissage automatique. Dans cette thèse, deux nouveaux types d'approches d'étiquetage automatique sont comparés : les techniques de segmentation multi-atlas par patch et les approches basées sur les réseaux neuronaux convolutifs. Ces deux méthodes sont actuellement largement utilisées pour la segmentation anatomique en imagerie médicale car elles permettent de mieux représenter la variabilité entre les sujets que les approches basées sur un modèle moyen unique. Cependant, ces méthodes permettent généralement une segmentation par voxel, sans tenir compte des propriétés géométriques et topologiques propres à la morphométrie des sillons. Par conséquent, je propose dans cette thèse de régulariser ces approches avec des contraintes géométriques "bottom-up" spécifiques à la morphologie sulcale, fournies par le pipeline BrainVISA/Morphologist. Ces contraintes regroupent les voxels en plis élémentaires, chaque plis étant associé à une unique étiquette. Afin d'éliminer les erreurs de pré-segmentation des plis élémentaires, le schéma de régularisation est complété par une perspective "top-down" qui permet le découpage automatique des plis élémentaires si nécessaire.

Base de données : La base de données est composée de 62 cerveaux sains sélectionnés à partir de différentes bases de données hétérogènes et étiquetés avec une nomenclature contenant 63 sillons pour l'hémisphère droit et 64 pour l'hémisphère gauche. Malheureusement, il n'existe pas de nomenclature standard pour l'étiquetage des sillons. Pour cette étude, les plis élémentaires de chaque cerveau ont été étiquetés manuellement selon une nomenclature de sillons issue d'un long processus itératif aboutissant à un consensus au sein d'un panel de plusieurs experts de la morphologie du cortex. La dernière itération de l'étiquetage de la base de données a été effectuée durant cette thèse en utilisant l'outil de visualisation TileViz (Mancip et al., 2018). Cet outil permet de visualiser et d'étiqueter simultanément

l'ensemble de la base de données sur le mur d'écrans situé à la Maison de la Simulation. Jusqu'à présent, il n'était possible d'étiqueter et d'évaluer simultanément qu'un nombre limité d'hémisphères, généralement quatre, sur un écran standard. Ainsi, cet outil permet de limiter le biais d'étiquetage induit par une vue restreinte de la base de données. De plus, lors de cette nouvelle itération, les plis élémentaires ont été découpés manuellement si nécessaire, ce qui était impossible auparavant.

Méthode : Les techniques de segmentation par multi-atlas, initialement introduites par Rohlfing et al. (2004) il y a une quinzaine d'années, utilisent chaque image segmentée manuellement comme atlas : les atlas sont recalés sur l'image à segmenter et les meilleures correspondances sont sélectionnées pour participer à la segmentation. Parmi les nombreuses variantes de ces techniques, les approche par patch (Coupé et al., 2011; Rousseau et al., 2011) permettent de s'affranchir d'un recalage non linéaire généralement coûteux et compliqué à mettre en place. Inspirés par ces approches, deux algorithmes de reconnaissance des sillons corticaux ont été implémentés. Le premier est directement inspirée de (Romero et al., 2017a) où chaque patch est un cube dans l'image à segmenter. Concernant le second algorithme, une nouvelle stratégie de génération de patchs est proposée : celle-ci se base sur une représentation à plus grande échelle des sillons, car la méthode standard d'extraction des patchs ne semble pas adéquate pour exploiter de manière optimale la géométrie des sillons et leurs positions relatives, alors que ce sont des caractéristiques discriminantes pour leur reconnaissance. Ces deux algorithmes seront désignés respectivement par PMAS (pour *Patch-based Multi-Atlas Segmentation*) et HPMAS (pour *Patch-based Multi-Atlas Segmentation with High level representation of the data*).

Les réseaux de neurones convolutifs ont d'abord été développés pour la classification d'images et sont aujourd'hui reconnus pour leur formidable efficacité à traiter de nombreux problèmes de vision par ordinateur. Ces techniques permettent une analyse efficace en apprenant une représentation abstraite de l'image. Concernant les problèmes de segmentation, la première approche a été proposée il y a une dizaine d'années par Ciresan et al. (2012) où un réseau de neurones est entraîné à classer chaque voxel de l'image à segmenter à partir de son patch environnant. Depuis, de nouvelles approches permettent la segmentation complète de l'image en utilisant des réseaux neuronaux entièrement convolutifs, comme celui initialement proposé par Long et al. (2015), dédié à la segmentation sémantique. En imagerie médicale, l'architecture la plus couramment utilisée pour les problèmes de segmentation est le U-Net, initialement proposé par Ronneberger et al. (2015), et adapté à la segmentation d'images 3D par Çiçek et al. (2016) et Milletari et al. (2016). Dans cette thèse, deux approches basées sur les réseaux de neurones convolutifs sont proposées. La première s'inspire de (Ciresan et al.,

2012), adapté à l'analyse d'images 3D. La seconde utilise l'architecture U-Net 3D proposée par Çiçek et al. (2016). Ces deux approches seront appelées PCNN (pour *Patch-based Convolutional Neural Network*) et UNET, respectivement.

Les quatre méthodes proposées permettent d'étiqueter chaque voxel du squelette de plis précédemment segmenté par le pipeline BrainVISA/Morphologist. Toutefois, ces méthodes ne garantissent pas que la définition géométrique d'un sillon, comme un ensemble de surfaces topologiquement simples, soit respectée. Ceci est particulièrement ennuyeux pour les études morphométriques dont les mesures sont basées sur cette définition. Pour y remédier, le pipeline BrainVISA/Morphologist fournit un algorithme d'agrégation "bottom-up" des voxels en plis élémentaires, qui sont directement étiquetés pour le modèle SPAM (*Statistical Probabilistic Anatomy Map*) proposé dans ce même pipeline. Une fois les voxels étiquetés par l'une des méthodes proposées ci-dessus, il est possible de régulariser les résultats à l'échelle des plis élémentaires. Cependant, l'extraction en amont des plis élémentaires peut parfois être inexacte. Dans cette thèse, l'étiquetage des voxels est utilisé pour donner une perspective "top-down" à un système traditionnel de reconnaissance de formes "bottom-up" (Figure 3). Ainsi, la découpe initiale en plis élémentaires proposée par BrainVISA/Morphologist est remise en question par l'étiquetage des voxels, éliminant les erreurs de sous-segmentation. L'approche proposée est donc particulièrement robuste aux incohérences spatiales qui peuvent survenir lors de l'étiquetage par voxel et à la définition potentiellement incorrecte des plis élémentaires en amont.

Résultats : Les nouvelles méthodes présentées dans cette thèse surpassent le modèle SPAM actuel fourni par le pipeline Morphologist de BrainVISA (Perrot et al., 2011). Par rapport à l'ancien système, les nouveaux modèles sont plus efficaces (avec des taux de reconnaissance d'environ 85 % au lieu de 81 %) et le modèle UNET est beaucoup plus rapide (il étiquette un hémisphère en seulement 20 secondes au lieu de 5 minutes). De plus, grâce à la technique de régularisation "top-down" permettant le redécoupage automatique des plis élémentaires, l'ensemble des nouveaux modèles proposés sont considérablement plus robustes aux erreurs de sous-segmentation. En pratique, ces améliorations permettent de reproduire des résultats d'analyses morphométriques sur des populations qui n'étaient auparavant possibles qu'avec un étiquetage manuel.

Conclusion

Les modèles développés dans cette thèse font partie d'une gamme d'outils pour l'étude des plissements corticaux. Ces outils sont principalement basés

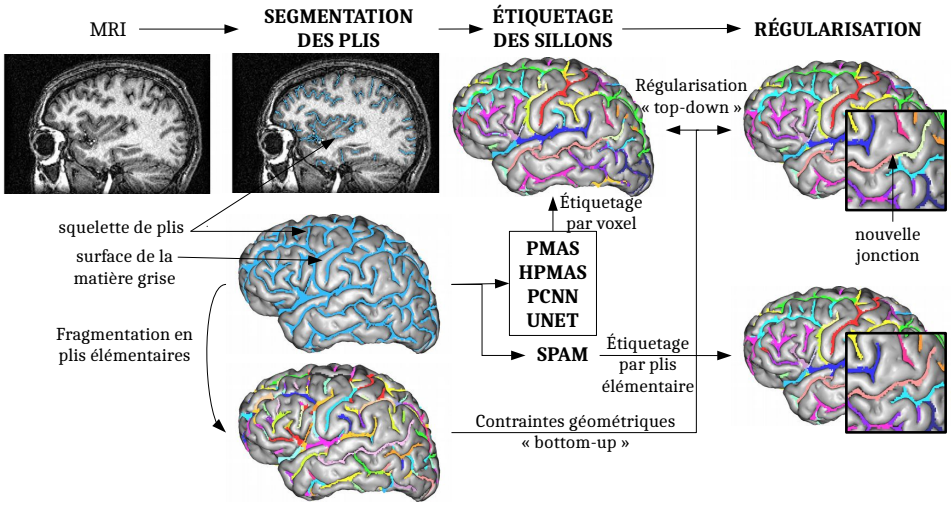


Figure 3 - De l'IRM aux sillons corticaux étiquetés : un pipeline en trois étapes. Tout d'abord, le squelette de plis est extrait à l'aide de la boîte à outils BrainVISA/Morphologist. Cette boîte à outils permet également de fragmenter le squelette en plis élémentaires. Ensuite, les voxels du squelette sont étiquetés par différents algorithmes. Les algorithmes basés sur les techniques de segmentation par multi-atlas (PMAS, HPMAS) et sur les réseaux de neurones convolutifs (PCNN, UNET) étiquettent chaque voxel du squelette tandis que l'algorithme SPAM étiquette directement les plis élémentaires. Enfin, l'étiquetage des voxels est régularisé par plis élémentaires, tout en les redivisant automatiquement en fonction de l'étiquetage par voxel proposé.

sur l'étiquetage manuel ou automatique des sillons, permettant par exemple l'étude de leurs caractéristiques (profondeur, longueur, ouverture, etc.), le recalage non linéaire de la surface corticale en se basant sur les principaux sillons (Auzias et al., 2009), etc. Les modèles comparés dans cette thèse pour la classification automatique des motifs de plissement et pour l'étiquetagee automatique des sillons, montrent une fois de plus l'incroyable efficacité de l'apprentissage profond. Comme ce domaine est en constante évolution, il est fort probable que les récents développements liés à ces techniques permettent d'améliorer les performances obtenues dans cette thèse. Cependant, l'utilisation d'une nomenclature pour l'étiquetage des sillons a ses limites lorsqu'il s'agit de représenter l'ensemble des configurations de plissement possibles. C'est pourquoi il n'y a pas actuellement de consensus entre experts sur une nomenclature. Il semble donc essentiel d'envisager l'utilisation de techniques non supervisées pour comprendre leur organisation sans s'appuyer sur une interprétation humaine trop subjective.

Bibliography

Alemán-Gómez, Y., J. Janssen, H. Schnack, E. Balaban, L. Pina-Camacho, F. Alfaro-Almagro, J. Castro-Fornieles, S. Otero, I. Baeza, D. Moreno, et al.
2013. The human cerebral cortex flattens during adolescence. *Journal of Neuroscience*, 33(38):15004–15010.

Amiez, C., P. Kostopoulos, A.-S. Champod, and M. Petrides
2006. Local morphology predicts functional organization of the dorsal premotor region in the human brain. *Journal of Neuroscience*, 26(10):2724–2731.

Amunts, K. and K. Zilles
2015. Architectonic mapping of the human brain beyond brodmann. *Neuron*, 88(6):1086–1107.

Asman, A. J. and B. A. Landman
2013. Non-local statistical label fusion for multi-atlas segmentation. *Medical image analysis*, 17(2):194–208.

Auzias, G., O. Colliot, J. A. Glaunes, M. Perrot, J.-F. Mangin, A. Trouve, and S. Baillet
2011. Diffeomorphic brain registration under exhaustive sulcal constraints. *IEEE transactions on medical imaging*, 30(6):1214–1227.

Auzias, G., J. Glaunes, O. Colliot, M. Perrot, J.-F. Mangin, A. Trouvé, and S. Baillet
2009. Disco: A coherent diffeomorphic framework for brain registration under exhaustive sulcal constraints. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Pp. 730–738. Springer.

Auzias, G., J. Lefevre, A. Le Troter, C. Fischer, M. Perrot, J. Régis, and O. Coulon
2013. Model-driven harmonic parameterization of the cortical surface: Hip-hop. *IEEE transactions on medical imaging*, 32(5):873–887.

Bai, W., W. Shi, D. P. O'regan, T. Tong, H. Wang, S. Jamil-Copley, N. S. Peters, and D. Rueckert
2013. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac mr images. *IEEE transactions on medical imaging*, 32(7):1302–1315.

Barnes, C., D. B. Goldman, E. Shechtman, and A. Finkelstein
2011. The patchmatch randomized matching algorithm for image manipulation. *Communications of the ACM*, 54(11):103–110.

Barnes, C., E. Shechtman, A. Finkelstein, and D. B. Goldman
2009. Patchmatch: A randomized correspondence algorithm for structural image editing. In *ACM Transactions on Graphics (ToG)*, volume 28, P. 24. ACM.

Behnke, K. J., M. E. Rettmann, D. L. Pham, D. Shen, S. M. Resnick, C. Davatzikos, and J. L. Prince
2003. Automatic classification of sulcal regions of the human brain cortex using pattern recognition. In *Medical Imaging 2003: Image Processing*, volume 5032, Pp. 1499–1511. International Society for Optics and Photonics.

Bénézit, A., L. Hertz-Pannier, G. Dehaene-Lambertz, K. Monzalvo, D. Germanaud, D. Duclap, P. Guevara, J.-F. Mangin, C. Poupon, M.-L. Moutard, et al.
2015. Organising white matter in a brain without corpus callosum fibres. *Cortex*, 63:155–171.

Benjamini, Y. and Y. Hochberg
1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Besl, P. J. and N. D. McKay
1992. Method for registration of 3-d shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, Pp. 586–607. International Society for Optics and Photonics.

Besson, P., F. Andermann, F. Dubeau, and A. Bernasconi
2008. Small focal cortical dysplasia lesions are located at the bottom of a deep sulcus. *Brain*, 131(12):3246–3255.

Blida, A.
2014. Ontology driven graph matching approach for automatic labeling brain cortical sulci. *IT4OD*, P. 162.

Borne, L., J.-F. Mangin, and D. Rivière
2018. A patch-based segmentation approach with high level representation of the data for cortical sulci recognition. In *International Workshop on Patch-based Techniques in Medical Imaging*, Pp. 114–121. Springer.

Borst, G., A. Cachia, J. Vidal, G. Simon, C. Fischer, A. Pineau, N. Poirel, J.-F. Mangin, and O. Houdé
2014. Folding of the anterior cingulate cortex partially explains inhibitory control during childhood: a longitudinal study. *Developmental cognitive neuroscience*, 9:126–135.

Brodersen, K. H., C. S. Ong, K. E. Stephan, and J. M. Buhmann
2010. The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, Pp. 3121–3124. IEEE.

Buades, A., B. Coll, and J.-M. Morel
2005. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, Pp. 60–65. IEEE.

Cachia, A., G. Borst, C. Tissier, C. Fisher, M. Plaze, O. Gay, D. Rivière, N. Gogtay, J. Giedd, J.-F. Mangin, et al.
2016. Longitudinal stability of the folding pattern of the anterior cingulate cortex during development. *Developmental cognitive neuroscience*, 19:122–127.

Cachia, A., G. Borst, J. Vidal, C. Fischer, A. Pineau, J.-F. Mangin, and O. Houdé
2014. The shape of the acc contributes to cognitive control efficiency in preschoolers. *Journal of cognitive neuroscience*, 26(1):96–106.

Caliński, T. and J. Harabasz
1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.

Chen, H., Y. Li, F. Ge, G. Li, D. Shen, and T. Liu
2017. Gyral net: A new representation of cortical folding organization. *Medical image analysis*, 42:14–25.

Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille
2018. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.

Çiçek, Ö., A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger
2016. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, Pp. 424–432. Springer.

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

Collins, D. L., P. Neelin, T. M. Peters, and A. C. Evans
1994. Automatic 3d intersubject registration of mr volumetric data in standardized talairach space. *Journal of computer assisted tomography*, 18(2):192–205.

Coupé, P., S. F. Eskildsen, J. V. Manjón, V. S. Fonov, D. L. Collins, A. disease Neuroimaging Initiative, et al.
2012. Simultaneous segmentation and grading of anatomical structures for patient's classification: application to alzheimer's disease. *NeuroImage*, 59(4):3736–3747.

Coupé, P., J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins
2011. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954.

Dale, A. M., B. Fischl, and M. I. Sereno
1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194.

De Guio, F., J.-F. Mangin, D. Rivière, M. Perrot, C. D. Molteno, S. W. Jacobson, E. M. Meintjes, and J. L. Jacobson
2014. A study of cortical morphology in children with fetal alcohol spectrum disorders. *Human brain mapping*, 35(5):2285–2296.

de Juan Romero, C., C. Bruder, U. Tomasello, J. M. Sanz-Anquela, and V. Borrell
2015. Discrete domains of gene expression in germinal layers distinguish the development of gyrencephaly. *The EMBO journal*, 34(14):1859–1874.

Destrieux, C., B. Fischl, A. Dale, and E. Halgren
2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15.

Eskildsen, S. F., P. Coupé, V. Fonov, J. V. Manjón, K. K. Leung, N. Guizard, S. N. Wassef, L. R. Østergaard, D. L. Collins, A. D. N. Initiative, et al.
2012. Beast: brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373.

Fei-Fei, L., K. Andrej, and J. Justin
2015. Cs231n: Convolutional neural networks for visual recognition.

Fernández, V., C. Llinares-Benadero, and V. Borrell
2016. Cerebral cortex expansion and folding: what have we learned? *The EMBO journal*, 35(10):1021–1044.

Fischl, B., N. Rajendran, E. Busa, J. Augustinack, O. Hinds, B. T. Yeo, H. Mohlberg, K. Amunts, and K. Zilles
2007. Cortical folding patterns and predicting cytoarchitecture. *Cerebral cortex*, 18(8):1973–1980.

Fischl, B., A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, et al.
2004. Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22.

Fonov, V., P. Coupé, S. Eskildsen, J. Manjon, and L. Collins
2012. Multi-atlas labeling with population-specific template and non-local patch-based label fusion. In *MICCAI 2012 Workshop on multi-atlas labeling*, Pp. 63–66.

Fornito, A., M. Yücel, S. Wood, G. W. Stuart, J.-A. Buchanan, T. Proffitt, V. Anderson, D. Velakoulis, and C. Pantelis
2004. Individual differences in anterior cingulate/paracingulate morphology are related to executive functions in healthy males. *Cerebral cortex*, 14(4):424–431.

Fukushima, K.
1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202.

Giraud, R., V.-T. Ta, A. Bugeau, P. Coupé, and N. Papadakis
2017. Superpatchmatch: An algorithm for robust correspondences using superpixel patches. *IEEE Transactions on Image Processing*, 26(8):4068–4078.

Giraud, R., V.-T. Ta, N. Papadakis, J. V. Manjón, D. L. Collins, P. Coupé, A. D. N. Initiative, et al.
2016. An optimized patchmatch for multi-scale and multi-feature label fusion. *NeuroImage*, 124:770–782.

Glasser, M. F., T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, et al.
2016. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171.

Glorot, X. and Y. Bengio
2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, Pp. 249–256.

Hamelin, L., M. Bertoux, M. Bottlaender, H. Corne, J. Lagarde, V. Hahn, J.-F. Mangin, B. Dubois, M. Chupin, L. C. De Souza, et al.
2015. Sulcal morphology as a new imaging marker for the diagnosis of early onset alzheimer's disease. *Neurobiology of aging*, 36(11):2932–2939.

Hara, K., H. Kataoka, and Y. Satoh
2018. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Pp. 6546–6555.

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

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

Heckemann, R. A., J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers
2006. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126.

Holz, D., A. E. Ichim, F. Tombari, R. B. Rusu, and S. Behnke
2015. Registration with the point cloud library: A modular framework for aligning in 3-d. *IEEE Robotics & Automation Magazine*, 22(4):110–124.

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

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger
2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Pp. 4700–4708.

Iglesias, J. E. and M. R. Sabuncu
2015. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219.

Im, K. and P. E. Grant
2019. Sulcal pits and patterns in developing human brains. *Neuroimage*, 185:881–890.

Im, K., H. J. Jo, J.-F. Mangin, A. C. Evans, S. I. Kim, and J.-M. Lee
2009. Spatial distribution of deep sulcal landmarks and hemispherical asymmetry on the cortical surface. *Cerebral cortex*, 20(3):602–611.

Ioffe, S. and C. Szegedy
2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Janssen, J., Y. Alemán-Gómez, H. Schnack, E. Balaban, L. Pina-Camacho, F. Alfaro-Almagro, J. Castro-Fornieles, S. Otero, I. Baeza, D. Moreno, et al.
2014. Cortical morphology of adolescents with bipolar disorder and with schizophrenia. *Schizophrenia research*, 158(1-3):91–99.

Jouvent, E., Z. Y. Sun, F. De Guio, E. Duchesnay, M. Duering, S. Ropele, M. Dichgans, J.-F. Mangin, and H. Chabriat
2016. Shape of the central sulcus and disability after subcortical stroke: a motor reserve hypothesis. *Stroke*, 47(4):1023–1029.

Kao, C.-Y., M. Hofer, G. Sapiro, J. Stern, K. Rehm, and D. A. Rottenberg
2007. A geometric method for automatic extraction of sulcal fundi. *IEEE transactions on medical imaging*, 26(4):530–540.

Klein, A., B. Mensh, S. Ghosh, J. Tourville, and J. Hirsch
2005. Mindboggle: automated brain labeling with multiple atlases. *BMC medical imaging*, 5(1):7.

Konukoglu, E., B. Glocker, D. Zikic, and A. Criminisi
2013. Neighbourhood approximation using randomized forests. *Medical image analysis*, 17(7):790–804.

Kostovic, I. and P. Rakic
1990. Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *Journal of Comparative Neurology*, 297(3):441–470.

Krizhevsky, A., I. Sutskever, and G. E. Hinton
2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, Pp. 1097–1105.

Le Goualher, G., C. Barillot, and Y. Bizais
1997. Modeling cortical sulci with active ribbons. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(08):1295–1315.

Le Goualher, G., D. L. Collins, C. Barillot, and A. C. Evans
1998. Automatic identificaiton of cortical sulci using a 3d probabilistic atlas. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Pp. 509–518. Springer.

Le Guen, Y., C. Philippe, D. Riviere, H. Lemaitre, A. Grigis, C. Fischer, G. Dehaene-Lambertz, J.-F. Mangin, and V. Frouin
2019. eqtl of kcnk2 regionally influences the brain sulcal widening: evidence from 15,597 uk biobank participants with neuroimaging data. *Brain Structure and Function*, 224(2):847–857.

Le Provost, J.-B., D. Bartrés-Faz, M.-L. Paillère-Martinot, E. Artiges, S. Pappata, C. Recasens, M. Pérez-Gómez, M. Bernardo, I. Baeza, F. Bayle, et al.
2003. Paracingulate sulcus morphology in men with early-onset schizophrenia. *The British Journal of Psychiatry*, 182(3):228–232.

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

LeCun, Y. A., L. Bottou, G. B. Orr, and K.-R. Müller
2012. Efficient backprop. In *Neural networks: Tricks of the trade*, Pp. 9–48. Springer.

Leroy, F., Q. Cai, S. L. Bogart, J. Dubois, O. Coulon, K. Monzalvo, C. Fischer, H. Glasel, L. Van der Haegen, A. Bénézit, et al.
2015. New human-specific brain landmark: the depth asymmetry of superior temporal sulcus. *Proceedings of the National Academy of Sciences*, 112(4):1208–1213.

Li, G., L. Guo, J. Nie, and T. Liu
2010. An automated pipeline for cortical sulcal fundi extraction. *Medical image analysis*, 14(3):343–359.

Lohmann, G. and D. Y. Von Cramon
2000. Automatic labelling of the human cortical surface using sulcal basins. *Medical image analysis*, 4(3):179–188.

Lohmann, G., D. Y. Von Cramon, and A. C. Colchester
2007. Deep sulcal landmarks provide an organizing framework for human cortical folding. *Cerebral Cortex*, 18(6):1415–1420.

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

Lyu, I., S. H. Kim, N. D. Woodward, M. A. Styner, and B. A. Landman
2018. Trace: a topological graph representation for automatic sulcal curve extraction. *IEEE transactions on medical imaging*, 37(7):1653–1663.

Malandain, G., G. Bertrand, and N. Ayache
1993. Topological segmentation of discrete surfaces. *International journal of computer vision*, 10(2):183–197.

Mancip, M., R. Spezia, Y. Jeanvoine, and C. Balsier
2018. Tileviz: Tile visualization for direct dynamics applied to astrochemical reactions. *Electronic Imaging*, 2018(16):286–1.

Mangin, J.-F.
1995. *Mise en correspondance d'images médicales 3 D multi-modalités multi-individus pour la corrélation anatomo-fonctionnelle cérébrale*. PhD thesis.

Mangin, J.-F., G. Auzias, O. Coulon, Z. Y. Sun, D. Rivière, and J. Régis
2015a. Sulci as landmarks. *Brain Mapping: An Encyclopedic Reference, Elsevier Inc*, Pp. 45–52.

Mangin, J.-F., V. Frouin, I. Bloch, J. Régis, and J. López-Krahe
1995. From 3d magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical Imaging and Vision*, 5(4):297–318.

Mangin, J.-F., E. Jouvent, and A. Cachia
2010. In-vivo measurement of cortical morphology: means and meanings. *Current opinion in neurology*, 23(4):359–367.

Mangin, J.-F., M. Perrot, G. Operto, A. Cachia, C. Fischer, J. Lefèvre, D. Rivière, and C. Neurospin
2015b. Sulcus identification and labeling. *Brain mapping: an encyclopedic reference. Academic Press: Elsevier, New York*, Pp. 365–371.

Mangin, J.-F., D. Riviere, A. Cachia, E. Duchesnay, Y. Cointepas, D. Papadopoulos-Orfanos, D. L. Collins, A. C. Evans, and J. Régis
2004. Object-based morphometry of the cerebral cortex. *IEEE transactions on medical imaging*, 23(8):968–982.

Manjón, J., S. Eskildsen, P. Coupé, J. Romero, L. Collins, and M. Robles
2014a. Nice: non-local intracranial cavity extraction. *International Journal of Biomedical Imaging*.

Manjón, J. V., P. Coupé, P. Raniga, Y. Xia, P. Desmond, J. Fripp, and O. Salvado
2018. Mri white matter lesion segmentation using an ensemble of neural

networks and overcomplete patch-based voting. *Computerized Medical Imaging and Graphics*, 69:43–51.

Manjón, J. V., S. F. Eskildsen, P. Coupé, J. E. Romero, D. L. Collins, and M. Robles
2014b. Nonlocal intracranial cavity extraction. *Journal of Biomedical Imaging*, 2014:10.

Mann, H. B. and D. R. Whitney
1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, Pp. 50–60.

Mellerio, C., P. Roca, F. Chassoux, F. Danière, A. Cachia, S. Lion, O. Naggara, B. Devaux, J.-F. Meder, and C. Oppenheim
2014. The power button sign: a newly described central sulcal pattern on surface rendering mr images of type 2 focal cortical dysplasia. *Radiology*, 274(2):500–507.

Miller, B. L. and J. L. Cummings
1999. The human frontal lobes: Functions and disorders.

Milletari, F., N. Navab, and S.-A. Ahmadi
2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, Pp. 565–571. IEEE.

Moriya, T., H. R. Roth, S. Nakamura, H. Oda, K. Nagara, M. Oda, and K. Mori
2018. Unsupervised segmentation of 3d medical images based on clustering and deep representation learning. In *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10578, P. 1057820. International Society for Optics and Photonics.

Muellner, J., C. Delmaire, R. Valabrégue, M. Schüpbach, J.-F. Mangin, M. Vidailhet, S. Lehéricy, A. Hartmann, and Y. Worbe
2015. Altered structure of cortical sulci in gilles de la tourette syndrome: Further support for abnormal brain development. *Movement disorders*, 30(5):655–661.

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

Ono, M., S. Kubik, and C. D. Abernathey
1990. *Atlas of the cerebral sulci*. Tps.

Parvathaneni, P., V. Nath, M. McHugo, Y. Huo, S. M. Resnick, N. D. Woodward, B. A. Landman, and I. Lyu
2019. Improving human cortical sulcal curve labeling in large scale cross-sectional mri using deep neural networks. *Journal of neuroscience methods*, P. 108311.

Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer
2017. Automatic differentiation in pytorch. In *NIPS-W*.

Perrot, M.
2009. *Reconnaissance automatique des sillons corticaux*. PhD thesis, École normale supérieure de Cachan-ENS Cachan.

Perrot, M., D. Rivière, and J.-F. Mangin
2011. Cortical sulci recognition and spatial normalization. *Medical image analysis*, 15(4):529–550.

Platt, J. et al.
1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74.

Plaze, M., J.-F. Mangin, M.-L. Paillere-Martinot, E. Artiges, J.-P. Olie, M.-O. Krebs, R. Gaillard, J.-L. Martinot, and A. Cachia
2015. "who is talking to me?"—self–other attribution of auditory hallucinations and sulcation of the right temporoparietal junction. *Schizophrenia research*, 169(1-3):95–100.

Rademacher, J., A. Galaburda, D. Kennedy, P. Filipek, and V. Caviness Jr
1992. Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of cognitive neuroscience*, 4(4):352–374.

Régis, J., J.-F. Mangin, T. Ochiai, V. Frouin, D. Riviére, A. Cachia, M. Tamura, and Y. Samson
2005. "sulcal root" generic model: a hypothesis to overcome the variability of the human cortex folding patterns. *Neurologia medico-chirurgica*, 45(1):1–17.

Régis, J., M. Tamura, M. C. Park, A. McGonigal, D. Riviere, O. Coulon, F. Bartolomei, N. Girard, D. Figarella-Branger, P. Chauvel, et al.
2011. Subclinical abnormal gyration pattern, a potential anatomic marker of epileptogenic zone in patients with magnetic resonance imaging–negative frontal lobe epilepsy. *Neurosurgery*, 69(1):80–94.

Reillo, I., C. de Juan Romero, M. Á. García-Cabezas, and V. Borrell
2010. A role for intermediate radial glia in the tangential expansion of the mammalian cerebral cortex. *Cerebral Cortex*, 21(7):1674–1694.

Rettmann, M. E., D. Tosun, X. Tao, S. M. Resnick, and J. L. Prince
2005. Program for assisted labeling of sulcal regions (pals): description and reliability. *NeuroImage*, 24(2):398–416.

Riviere, D.
2000. Apprentissage de la variabilité inter-individuelle de l'anatomie corticale cérébrale pour la reconnaissance automatique des sillons. *These de doctorat, Université d'Evry.*

Rivière, D., D. Geffroy, I. Denghien, N. Souedet, and Y. Cointepas
2009. Brainvisa: an extensible software environment for sharing multimodal neuroimaging data and processing tools. *Neuroimage*, 47.

Riviere, D., J.-F. Mangin, D. Papadopoulos-Orfanos, J.-M. Martinez, V. Frouin, and J. Régis
2002. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Medical image analysis*, 6(2):77–92.

Roca, P., C. Mellerio, F. Chassoux, D. Rivière, A. Cachia, S. Charron, S. Lion, J.-F. Mangin, B. Devaux, J.-F. Meder, et al.
2015. Sulcus-based mr analysis of focal cortical dysplasia located in the central region. *PloS one*, 10(3):e0122252.

Rohlfing, T., R. Brandt, R. Menzel, and C. R. Maurer Jr
2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4):1428–1442.

Romero, J. E., P. Coupé, R. Giraud, V.-T. Ta, V. Fonov, M. T. M. Park, M. M. Chakravarty, A. N. Voineskos, and J. V. Manjón
2017a. Ceres: a new cerebellum lobule segmentation method. *NeuroImage*, 147:916–924.

Romero, J. E., P. Coupé, and J. V. Manjón
2017b. Hips: A new hippocampus subfield segmentation method. *NeuroImage*, 163:286–295.

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

Rosenblatt, F.
1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Rousseau, F., P. A. Habas, and C. Studholme
2011. A supervised patch-based approach for human brain labeling. *IEEE transactions on medical imaging*, 30(10):1852–1862.

Royackkers, N., M. Desvignes, H. Fawal, and M. Revenu
1999. Detection and statistical analysis of human cortical sulci. *NeuroImage*, 10(6):625–641.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al.
2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.

Seong, J.-K., K. Im, S. W. Yoo, S. W. Seo, D. L. Na, and J.-M. Lee
2010. Automatic extraction of sulcal lines on cortical surfaces based on anisotropic geodesic distance. *Neuroimage*, 49(1):293–302.

Shattuck, D. W., A. A. Joshi, D. Pantazis, E. Kan, R. A. Dutton, E. R. Sowell, P. M. Thompson, A. W. Toga, and R. M. Leahy
2009. Semi-automated method for delineation of landmarks on models of the cerebral cortex. *Journal of neuroscience methods*, 178(2):385–392.

Shi, Y., P. M. Thompson, I. Dinov, and A. W. Toga
2008. Hamilton–jacobi skeleton on cortical surfaces. *IEEE transactions on medical imaging*, 27(5):664–673.

Shi, Y., Z. Tu, A. L. Reiss, R. A. Dutton, A. D. Lee, A. M. Galaburda, I. Dinov, P. M. Thompson, and A. W. Toga
2007. Joint sulci detection using graphical models and boosted priors. In *Biennial International Conference on Information Processing in Medical Imaging*, Pp. 98–109. Springer.

Simonyan, K., A. Vedaldi, and A. Zisserman
2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

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

Snoek, C. G., M. Worring, and A. W. Smeulders
2005. Early versus late fusion in semantic video analysis. In *Proceedings of the 13th annual ACM international conference on Multimedia*, Pp. 399–402. ACM.

Soltaninejad, M., G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye
2017. Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair mri. *International journal of computer assisted radiology and surgery*, 12(2):183–203.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov
2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Sun, Z., P. Pinel, D. Rivière, A. Moreno, S. Dehaene, and J.-F. Mangin
2016. Linking morphological and functional variability in hand movement and silent reading. *Brain Structure and Function*, 221(7):3361–3371.

Sun, Z. Y., S. Klöppel, D. Rivière, M. Perrot, R. Frackowiak, H. Siebner, and J.-F. Mangin
2012. The effect of handedness on the shape of the central sulcus. *Neuroimage*, 60(1):332–339.

Sur, M. and J. L. Rubenstein
2005. Patterning and plasticity of the cerebral cortex. *Science*, 310(5749):805–810.

Sutskever, I., J. Martens, G. Dahl, and G. Hinton
2013. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, Pp. 1139–1147.

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

Ta, V.-T., R. Giraud, D. L. Collins, and P. Coupé
2014. Optimized patchmatch for near real time and accurate label fusion. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Pp. 105–112. Springer.

Tallinen, T., J. Y. Chung, J. S. Biggins, and L. Mahadevan
2014. Gyrification from constrained cortical expansion. *Proceedings of the National Academy of Sciences*, 111(35):12667–12672.

Tallinen, T., J. Y. Chung, F. Rousseau, N. Girard, J. Lefèvre, and L. Mahadevan
2016. On the growth and form of cortical convolutions. *Nature Physics*, 12(6):588.

Vaillant, M. and C. Davatzikos
1997. Finding parametric representations of the cortical sulci using an active contour model. *Medical Image Analysis*, 1(4):295–315.

Van Essen, D. C.
1997. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature*, 385(6614):313.

Vivodtzev, F., L. Linsen, B. Hamann, K. I. Joy, and B. A. Olshausen
2006. Brain mapping using topology graphs obtained by surface segmentation. In *Scientific Visualization: The Visual Extraction of Knowledge from Data*, Pp. 35–48. Springer.

Wang, H., J. W. Suh, S. Das, J. Pluta, M. Altinay, and P. Yushkevich
2011. Regression-based label fusion for multi-atlas segmentation. In *CVPR 2011*, Pp. 1113–1120. IEEE.

Wang, L., F. Shi, G. Li, Y. Gao, W. Lin, J. H. Gilmore, and D. Shen
2014. Segmentation of neonatal brain mr images using patch-driven level sets. *NeuroImage*, 84:141–158.

Ward Jr, J. H.
1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244.

Watson, J. D., R. Myers, R. S. Frackowiak, J. V. Hajnal, R. P. Woods, J. C. Mazziotta, S. Shipp, and S. Zeki
1993. Area v5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral cortex*, 3(2):79–94.

Weiner, K. S., G. Golarai, J. Caspers, M. R. Chuapoco, H. Mohlberg, K. Zilles, K. Amunts, and K. Grill-Spector
2014. The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage*, 84:453–465.

Welker, W.
1990. Why does cerebral cortex fissure and fold? In *Cerebral cortex*, Pp. 3–136. Springer.

Werbos, P. J. et al.
1990. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560.

Wolz, R., C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert
2013. Automated abdominal multi-organ segmentation with subject-

specific atlas generation. *IEEE transactions on medical imaging*, 32(9):1723–1730.

Wu, G., Q. Wang, D. Zhang, F. Nie, H. Huang, and D. Shen
2014. A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Medical image analysis*, 18(6):881–890.

Xie, J., R. Girshick, and A. Farhadi
2016. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, Pp. 478–487.

Yang, B., X. Fu, N. D. Sidiropoulos, and M. Hong
2017. Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, Pp. 3861–3870. JMLR. org.

Yang, F. and F. Kruggel
2009. A graph matching approach for labeling brain sulci using location, orientation, and shape. *Neurocomputing*, 73(1-3):179–190.

Yang, J., D. Parikh, and D. Batra
2016. Joint unsupervised learning of deep representations and image clusters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Pp. 5147–5156.

Yücel, M., S. J. Wood, L. J. Phillips, G. W. Stuart, D. J. Smith, A. Yung, D. Velakoulis, P. D. Mcgorry, and C. Pantelis
2003. Morphology of the anterior cingulate cortex in young men at ultra-high risk of developing a psychotic illness. *The British Journal of Psychiatry*, 182(6):518–524.

Zeiler, M. D. and R. Fergus
2014. Visualizing and understanding convolutional networks. In *European conference on computer vision*, Pp. 818–833. Springer.

Zeng, X., L. H. Staib, R. T. Schultz, H. Tagare, L. Win, and J. S. Duncan
1999. A new approach to 3d sulcal ribbon finding from mr images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Pp. 148–157. Springer.

Zhang, D., Q. Guo, G. Wu, and D. Shen
2012. Sparse patch-based label fusion for multi-atlas segmentation. In *International Workshop on Multimodal Brain Image Analysis*, Pp. 94–102. Springer.

Zhou, Y., P. M. Thompson, and A. W. Toga
1999. Extracting and representing the cortical sulci. *IEEE Computer Graphics and Applications*, 19(3):49–55.

Zoph, B., V. Vasudevan, J. Shlens, and Q. V. Le
2018. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Pp. 8697–8710.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Physique et ingénierie : électrons, photons, sciences du vivant (EOBE)

Titre : Conception d'un algorithme de vision par ordinateur "top-down" dédié à la reconnaissance des sillons corticaux

Mots clés : sillons corticaux, apprentissage profond, apprentissage par patch, segmentation, reconnaissance de formes, vision par ordinateur

Résumé : Les plissements du cortex caractérisent de manière unique chaque être humain. Ils apparaissent pendant le dernier trimestre de grossesse, c'est-à-dire pendant la mise en place de l'architecture cérébrale. Les motifs de ces plis sont impactés par les spécificités de cette architecture propres à chaque individu. Ils pourraient donc dévoiler les signatures de certaines anomalies du développement à l'origine de pathologies psychiatriques. Le laboratoire d'analyse d'images de Neurospin développe depuis 25 ans un programme de recherche visant à mettre en évidence de telles signatures grâce à la conception d'outils de vision par ordinateur dédiés qu'il diffuse à la communauté (http://brainvisa.info). Cette thèse a permis l'émergence d'une nouvelle génération d'outils basés sur des techniques d'apprentissage automatique. Le premier outil proposé classifie automatiquement des motifs locaux de plissements du cortex, un problème qui n'avait jamais été abordé jusqu'ici. Le second outil vise l'étiquetage automatique des sillons corticaux en modélisant des mécanismes de reconnaissance "top-down" nécessaires pour pallier les faiblesses des démarches "bottom-up" développées jusqu'à présent. Ainsi, en plus d'avoir des taux de reconnaissances plus élevés et un temps d'exécution plus court, le nouveau modèle proposé est robuste aux erreurs de sous-segmentation, ce qui est l'une des plus grandes faiblesses de l'ancien système. Pour réaliser ces deux outils, plusieurs algorithmes d'apprentissage automatique ont été implémentés et comparés. Ces algorithmes s'inspirent d'une part des méthodes multi-atlas, en particulier de l'approche par patch, qui sont largement utilisées pour la segmentation anatomique d'images médicales et d'autre part des méthodes d'apprentissage profond qui révolutionnent aujourd'hui le monde de la vision par ordinateur. Les travaux de cette thèse confirment l'incroyable efficacité des techniques d'apprentissage profond pour s'adapter à des problèmes complexes. Cependant, les performances obtenues avec ces techniques sont généralement équivalentes à celles des approches par patch, voire moins bonnes si la base de données d'apprentissage est restreinte. Ce qui fait de l'apprentissage profond un outil particulièrement intéressant en pratique n'est autre que sa rapidité d'exécution, d'autant plus pour l'analyse des bases de données colossales aujourd'hui disponibles.

Title : Design of a top-down computer vision algorithm dedicated to the recognition of cortical sulci

Keywords : cortical sulci, deep learning, patch learning, segmentation, pattern recognition, computer vision

Abstract : We are seven billion humans with unique cortical folding patterns. The cortical folding process occurs during the last trimester of pregnancy, during the emergence of cortical architecture. The folding patterns are impacted by architectural features specific to each individual. Hence, they could reveal signatures of abnormal developments that can lead to psychiatric syndroms. For the last 25 years, the image analysis lab of Neurospin has been designing dedicated computer vision tools to tackle the research of such signatures. The resulting tools are distributed to the community (http://brainvisa.info). This thesis has resulted in the emergence of a new generation of tools based on machine learning techniques. The first proposed tool automatically classifies local patterns of cortical folds, a problem that had never been addressed before. The second tool aims at the automatic labeling of cortical sulci by modeling the top-down recognition mechanisms necessary to overcome weaknesses of the current bottom-up systems. Thus, in addition to having higher recognition rates and shorter execution time, the proposed new model is robust to sub-segmentation errors, which is one of the greatest weaknesses of the old system. To realize these two tools, several machine learning algorithms were implemented and compared. These algorithms are inspired on the one hand by multi-atlas methods, in particular the patch approaches, which are widely used for the anatomical segmentation of medical images and on the other hand by the deep learning methods that are revolutionizing the world of computer vision. The work of this thesis confirms the incredible effectiveness of deep learning techniques to adapt well to complex problems. However, the performances obtained with these techniques are generally equivalent to those of patch approaches, or even worse if the training database is limited. What makes deep learning a particularly interesting tool in practice is its fast execution, especially for the analysis of the huge databases now available.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France