A unified framework for focal intensity change detection and deformable image registration : application to the monitoring of multiple sclerosis lesions in longitudinal 3D brain MRI

Eléonore Dufresne

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE MSII

Laboratoire ICube UMR 7357 CNRS – Université de Strasbourg

THÈSE présentée par :

Eléonore DUFRESNE

soutenue le : 20 06 **2023**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Signal, Image, Automatique, Robotique

Cadre unifié pour la détection de changements en IRM cérébrale. Application au suivi longitudinal de patients atteints de sclérose en plaques.

THÈSE dirigée par :

Mr NOBLET Vincent	Ingénieur de Recherche, CNRS, ICube, Université de Strasbourg
Mr KREMER Stéphane	PU-PH, HDR, Université de Strasbourg, Hôpitaux Universitaires de Strasbourg

RAPPORTEURS :

Mr DOJAT Michel	Directeur de Recherche, INSERM, Grenoble Institut des Neurosciences
Mr ROUSSEAU François	Professeur, IMT Atlantique

AUTRES MEMBRES DU JURY :

Mme MERVEILLE Odyssée	Maîtresse de Conférence, INSA Lyon - Examinatrice
Mr FORTUN Denis	Chargé de Recherche, CNRS, ICube, Université de Strasbourg - Encadrant

Contents

Introduction

Context

Magnetic resonance imaging (MRI) plays an essential role for the diagnosis, the monitoring of disease course and the therapeutic assessment of neurodegenerative diseases such as multiple sclerosis (MS). MS causes degeneration of the myelin coating that surrounds the nerves. Its evolution is characterized by the apparition of focal lesions in the brain and in the spinal cord, and by a progressive atrophy of brain tissues. Both phenomena can be monitored thanks to MRI [Kaunzner and Gauthier, 2017]. In clinical routine, the evolution of lesion load and brain atrophy is generally assessed qualitatively. The precise quantification of lesion changes over time would be of great interest to finely characterize the course of the pathology and to evaluate at an early stage the effect of a therapeutic strategy [McNamara et al., 2017]. However, the manual delineation of MS lesion changes in MRI is a tedious and time consuming task, prone to intra- and inter-observer variability. Therefore, there is a great need for efficient and reliable automated tools [Altay et al., 2013].

Subject of the thesis

The area of interest of this thesis is the longitudinal analysis of brain MRI collected from patients suffering from Multiple Sclerosis (MS). In longitudinal MRI scans of a given subject, several factors can cause discrepancies between two images, such as changes in brain lesion load, deformation of brain structures due to atrophy, or acquisition artifacts. The main challenge is to disentangle these changes in order to quantify their individual impact on the overall changes. Most methods for detecting appearing or evolving lesions rely on a first preliminary step that consist in correcting for all the other kinds of changes sequentially. First, global intensity changes caused by the difference of MRI acquisition setups are eliminated thanks to bias field correction and image intensity normalization algorithms. Then, geometrical discrepancies due to difference in patient positioning, geometrical distortions and brain atrophy are estimated and corrected through registration. Finally, the evolution of lesions are identified by detecting remaining focal intensity changes. This last step generally consists in thresholding a map of intensity- or deformation-based features (see [Lladó et al., 2012] for a survey). However, correcting each source of changes separately without taking into account their intertwining may result in biased estimates.

Registration is a particularly crucial step that can greatly impact the performance of the detection of MS lesion changes. The most common practice is to use either rigid or affine transformation models to compensate for different patient positioning without altering the appearance of lesions. However, such linear models are not able to capture the complex deformations induced by brain atrophy, which typically occurs in MS. These remaining deformations may yield spurious detections in atrophied areas, especially in the cortex and around the ventricles. Deformable registration is a solution to estimate more complex deformations, but it tends to make new lesions disappear while minimizing the dissimilarity between the two images, resulting in a detection that lacks sensitivity. Therefore, affine registration is prefered over deformable approaches in practice, despite producing a high number of false positive detections. The main goal of this thesis is to conciliate deformable registration and high sensitivity, in order to combine proper brain atrophy compensation and accurate lesion change detection.

Contributions

To overcome the limitations of the sequential approach and improve the detection of MS lesion evolution, we propose a unified framework for joint registration and change detection. To this end, we formulate these two separate tasks as a single optimization problem involving a unique energy that models their coupling. It results in a difficult non-convex problem for which we develop an efficient and versatile optimization strategy based on an alternating scheme. We demonstrate experimentally on synthetic and real MS data that our joint formulation improves the accuracy of lesion change detection over the standard sequential approach in the presence of brain atrophy. This is possible thanks to the combination of accurate atrophy compensation provided by the deformable registration model, with the preservation of the lesions shape during registration thanks to our joint formulation. We also provide preliminary investigations of extensions of this model. In particular, we integrate the intensity normalization step into our joint formulation, and we also propose to take into account multiple modalities and a symmetric registration model in our method. Our work has been published in an international conference paper [Dufresne et al., 2020], a national conference paper [Dufresne et al., 2022b], and a journal paper [Dufresne et al., 2022a].

Document outline

- Chapter 1 sets the scientific background of the thesis. It first provides an overview of multiple sclerosis and its evolution. Then, the basic principles of MRI and its relevance for diagnosis and patient monitoring in a clinical setting are presented. The chapter also addresses the difficulties that arise with automated MRI analysis. Finally, the standard clinical workflow is explained, and we highlight the importance of automated processing for longitudinal MRI examinations.

- Chapter 2 outlines the classical sequential process for automated lesion change detection in brain MRI. It covers the techniques used to eliminate non-relevant

sources of change caused by factors such as MRI acquisition parameters, patient positioning, and false detections outside of the brain. The chapter also presents the basic principles of longitudinal brain MRI registration, reviews the state of the art of lesion change detection methods, and identifies the drawbacks of the sequential approaches.

- Our hypothesis is that the limitations of sequential approaches stem from treating registration and change detection as separate tasks. In Chapter 3 we propose a unified framework that performs these tasks jointly. First we formulate registration and change detection as minimization problems and explain how they can be combined into a single minimization problem. We solve this problem with alternating minimization with respect to the deformation field and the change detection map. We used and approach based on alternated direction method of multiplier (ADMM) for the registration subproblem, and graph-cut for the change detection subproblem.

- In Chapter 4, a comparison between the results of the proposed model and conventional sequential approaches is made. The evaluation framework is composed of one synthetic and two publicly available longitudinal MRI multiple sclerosis patients datasets. The assessment of the performance of the joint model is conducted using voxel-wise and lesion-wise metrics. The chapter also outlines the process carried out to select appropriate values of the hyperparameters governing the model. Quantitative outcomes for both the synthetic and real patients datasets are presented, indicating that the joint model yields better overall sensitivity in comparison to the sequential model with deformable registration, and overall specificity compared to the sequential model with affine registration. Additionally, limitations of the proposed model are discussed.

- Chapter 5 illustrates the adaptability of the proposed unified framework by introducing variants of the data-term that exhibit increased noise resilience, incorporate multi-channel information, symmetric registration, or integrate intensity normalization task within the joint model. The implementation of these variants and their potential to address some of the limitations of the existing model are discussed.

Chapter 1

Medical Context

This chapter gives an overview of the context of the thesis, which is the automated analysis of longitudinal brain MRI for the diagnosis and monitoring of patients suffering from multiple sclerosis. It is divided in three parts. Multiple sclerosis is first described in terms of prevalence, symptoms, and phenotypes. A brief overview of the basic principles of MRI is presented in the second part. Finally, the last section outlines the standard clinical practices for the diagnosis and the follow-up of multiple sclerosis and emphasizes the need for automated tools for analyzing longitudinal brain MRI.

1.1 Multiple Sclerosis

Multiple sclerosis (MS) is an auto-immune neurodegenerative disease that affects over 2.3 million people worldwide[1] and 110 000 people in France[2]. It is usually diagnosed at ages 25 to 35 with a higher prevalence among women (sex-ratio of 1 man for 3 women).

The auto-immune mechanisms attack the myelin coating that surrounds the nerves, as schematized in Figure 1.1, leading to the appearance of lesions called plaque. These lesions correspond to inflammation, de-myelinization and often axonal degradation. The myelin coating ensures that nervous impulses are properly transferred from neurons to nerve endings. Damage caused by inflammation can disrupt the transmission of signals in the body, resulting in various types of disabilities, including motor (such as numbness or weakness in limbs, electric shock sensations, tremors, lack of coordination, and unsteady gait), sensorial (such as partial or complete loss of vision, prolonged double vision, and blurry vision), or cognitive impairments.

The progression of MS and its symptoms vary greatly among individuals. The diagnosis is made by observing clinical signs and brain lesions through Magnetic Resonance Imaging (MRI). The symptoms can change over time and vary from one person to another depending on the areas of the brain or spinal cord affected by the lesions. Additionally, brain atrophy progresses at a faster rate than normal aging.

There are four patterns of progression and the three progressive forms illustrated in Figure 1.2:

[1] https://www.nationalmssociety.org/What-is-MS/MS-FAQ-s

[2] https://www.inserm.fr/dossier/sclerose-en-plaques-sep/

- Clinically Isolated Syndrome (CIS)
- Relapsing-Remitting MS (RRMS)
- Primary Progressive MS (PPMS)
- Secondary progressive MS (SPMS)

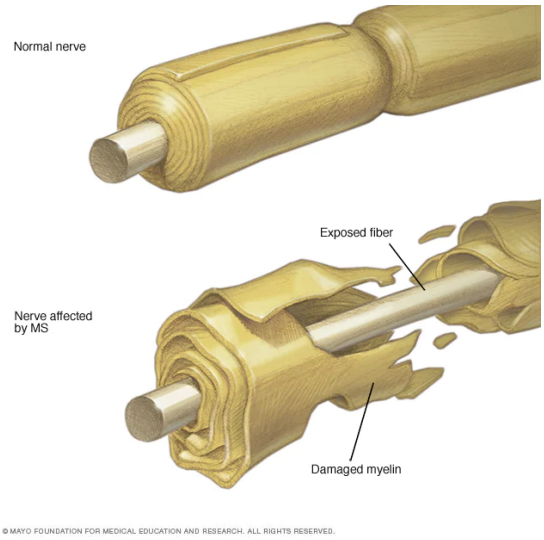


Figure 1.1: Schematic view of demyelinization (Source: https://www.mayoclinic.org/diseases-conditions/multiple-sclerosis/symptoms-causes/syc-20350269)

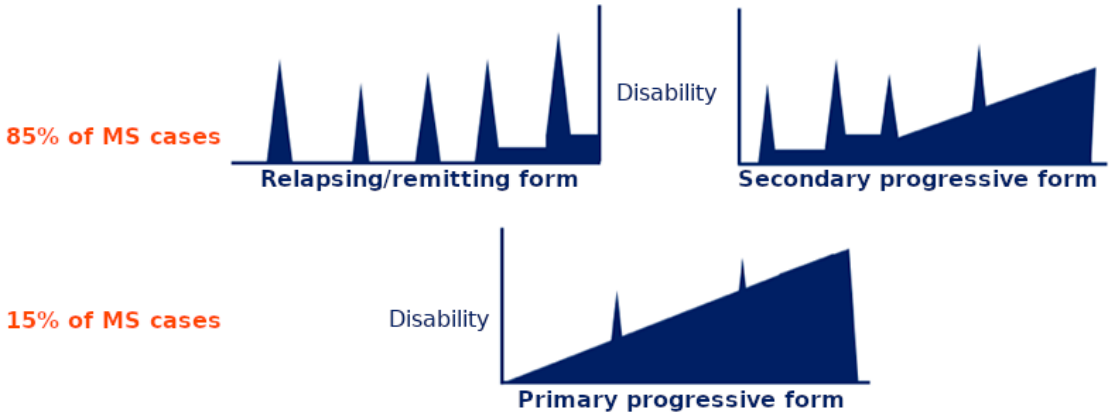


Figure 1.2: Patterns of progression of MS (Source: https://institutducerveau-icm.org/fr/sclerose-en-plaques/)

The most prevalent type of MS is relapsing-remitting MS (RRMS), accounting for 85% of cases. It typically begins with a clinically isolated syndrome (CIS), and

is characterized by unpredictable relapses with symptoms appearing suddenly and often accompanied by fatigue. These symptoms may partially or fully resolve within a few weeks, or persist in a milder form for months or years before another relapse occurs. In most cases, those with RRMS will eventually develop secondary progressive MS (SPMS), where symptoms persist without any periods of remission. In contrast, primary progressive MS (PPMS) occurs in 15% of cases, and is characterized by a lack of remission after initial symptoms appear.

Currently, MS cannot be cured but symptomatic and disease modifying treatments are available. The administration of corticosteroids during relapses is proposed to reduce the duration of symptoms by soothing the inflammatory response. To reduce the frequency of relapses and the occurrence of new lesions, medication that modulate the immune system is prescribed.

Early diagnosis of MS allows for the early initiation of treatment, which can help prevent disability. Designing an appropriate treatment plan is complex due to the numerous options available. A balance has to be found between effectiveness, safety, and adverse effects. The optimal treatment plan may vary as the disease progresses. Efforts are getting underway to identify biomarkers that can predict treatment response.

In addition to the evaluation of clinical symptoms, magnetic resonance imaging (MRI) is used to diagnose and monitor progression of the disease by assessing the spatial dissemination and temporal evolution of lesions.

In France, the Observatoire Français de la Sclérose en Plaques (OFSEP) is an organism that is collecting data to aid research for MS; so far, they have a national cohort of 75,000 people with MS. To monitor the disease, a recommendation of OFSEP is to acquire new MRI every 6 months[3]. When over two to three new lesions appear in MRI, the treatment is changed to a more efficient drug to prevent disability. The objective is to achieve the state of NEDA (No Evidence of Disease Activity) [Wattjes et al., 2015].

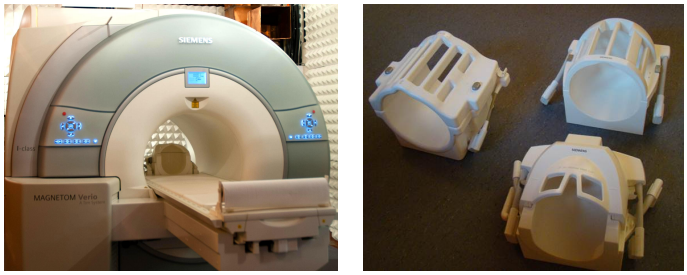
1.2 Magnetic Resonance Imaging (MRI)

MRI is a non-invasive medical imaging technique used to produce detailed images of the internal structures of the body. It plays a key role in diagnosing and monitoring Multiple Sclerosis. This section explains the formation of MRI images and highlights key characteristics to consider in automated image processing.

The main components of an MRI scanner are the magnet, gradient coils, and radiofrequency (RF) coils (see figure 1.3).

MR signal MRI is based on the phenomenon of nuclear magnetic resonance (NMR). NMR relies on the concepts of resonance and excitation of the nuclear spin of a charged particle. Certain atomic nuclei can absorb and re-emit radiofrequency energy when placed in a magnetic field. The hydrogen atom ^1H nuclei are abundant in the human body and are sensitive to the magnetic field. The spin of the hydrogen nucleus

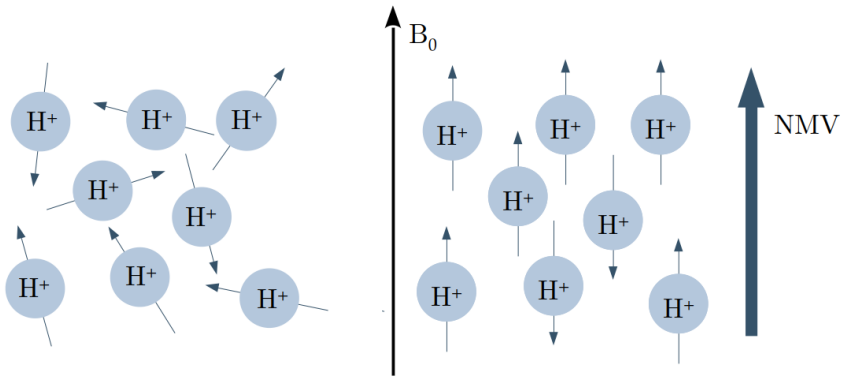
[3]https://www.ofsep.org/images/IMAGERIE/POS-COH-01f_V20_Protocole_IRM_OFSEP.pdf



(a) (b)

Figure 1.3: (a) 3T MRI scanner at IRIS platform in ICube laboratory, (b) Head antenna (RF coils)

corresponds to its rotation around an axis. As depicted in Figure 1.4, in a normal state, the rotation axes of hydrogen nuclei are oriented in random directions. In a MRI scanner, the magnetic field B_0 forces the rotation axes of hydrogen nuclei to align in the direction of the field in a parallel or anti-parallel fashion (in the direction of B_0 or in the opposite direction). The cumulative effect of the magnetic moments of the nuclei is the net magnetization vector (NMV).



(a) (b)

Figure 1.4: (a) Hydrogen nuclei spins oriented randomly, (b) Hydrogen nuclei subject to the B_0 magnetic field. The sum of all magnetic moments vectors is the Net Magnetization vector (NMV)

This causes the hydrogen atom to adopt a precession movement around the axis of B_0 Bitar et al. [2006]. The precession occurs at a frequency corresponding to the Larmor equation, $\omega_0 = \gamma \mathbf{B}_0$, with $\frac{\gamma}{2\pi} = 42.57 MHz.T^{-1}$ being the gyromagnetic ratio specific to the hydrogen nucleus. A radiofrequency signal B_1 matching the precession frequency of the hydrogen nuclei is emitted by the RF coils, causing the nuclei to flip at a specific angle α, resulting in the NMV having both a transversal and a longitudinal component, as shown in figure 1.5. As the RF signal is turned off, the spins realign with the B_0 field through a process called T1 recovery, where the longitudinal magnetization increases in magnitude (or recovers). The transverse magnetization decreases through

T2 and T2* decay and is due to the dephasing of the precession movement of the nuclei. The signal that is detected by the MRI scanner is the electrical current induced by the precession of the transverse magnetization around the receiver coil. Depending on the tissue in which the hydrogen atoms are encapsulated, the T1, T2 and T2* values vary, as schematized in Figure 1.6. T2* is dependent on the inhomogeneities of the RF field and B_0 field. Fat has a shorter T1 (*i.e.,* , recovers faster) and a shorter T2 (*i.e.,* , decays faster) than water, which has a relatively long T1 and T2. T2*decay occurs very quickly in both fat and water but its duration is dependent on the inhomogeneities.

Image encoding To encode the spatial location of the signals, the gradient coils are used along the three dimensions of space, x, y and z directions. Three gradient coils are used respectively for slice selection, frequency encoding and phase encoding ; these last two gradients are used to populate a 2D matrix, the K space, which is in the Fourier plan. The modulus of the inverse Fourier transform is applied to the K space to form the image of the selected slice. Note that isotropic 3D volume can also be created by using two phases and one frequency encoding gradients.

MRI modalities To exploit properties that are specific to tissue types and create images with tissue contrast, two key acquisition parameters can be tuned: Echo Time (TE) and Repetition Time (TR), depicted in Figure 1.6. TR is the time (in milliseconds) between the application of an RF excitation pulse and the start of the next RF pulse. TE (in milliseconds) is the time between the application of the RF pulse and the moment when the signal is measured. Adjusting the TR and TE parameters in MRI allows to highlight specific tissue contrasts. (*i.e.*, T1, T2, and proton density). Different combinations of TR, TE, TI, gradient configuration and acquisition time parameters can depict different tissue contrast and affect the final quality of the image in terms of spatial resolution.

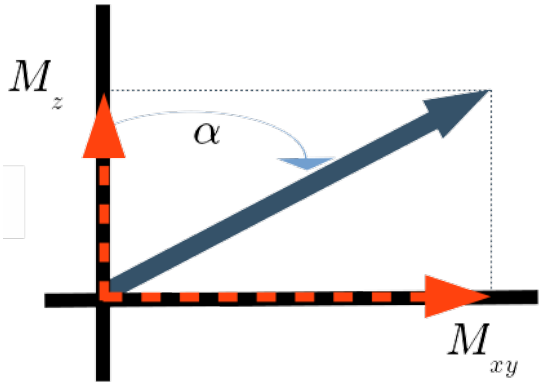
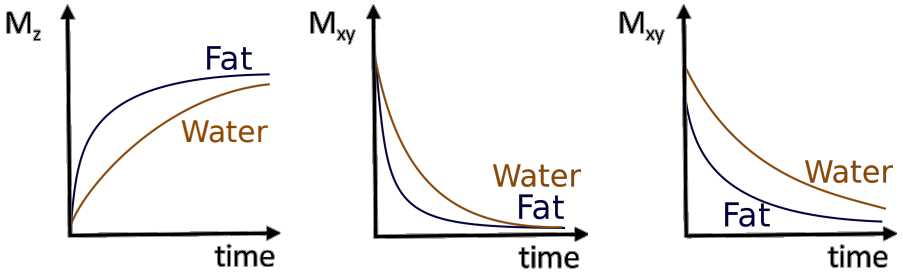


Figure 1.5: Upon excitation by the RF pulse, the NMV is flipped at an angle α and has a longitudinal M_z and a transverse M_{xy} component (source: [Bitar et al., 2006])



T1 recovery T2* decay T2 decay

Figure 1.6: Differences in T1 recovery and T2 and T2* decay for water and fat (source: Bitar et al. [2006])

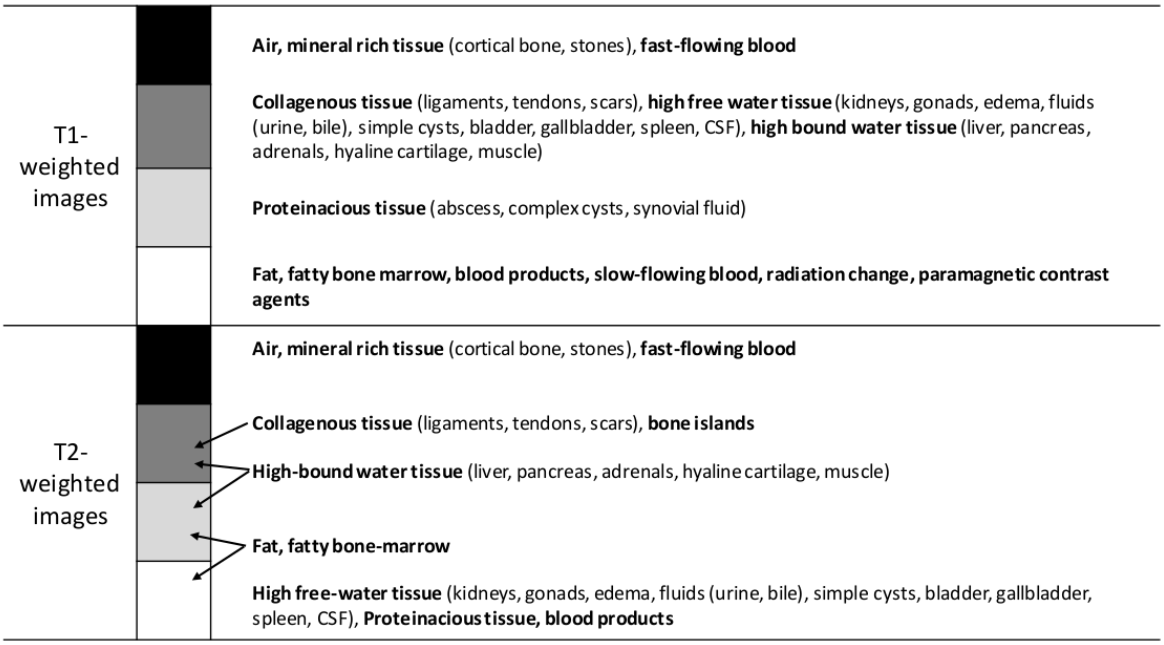
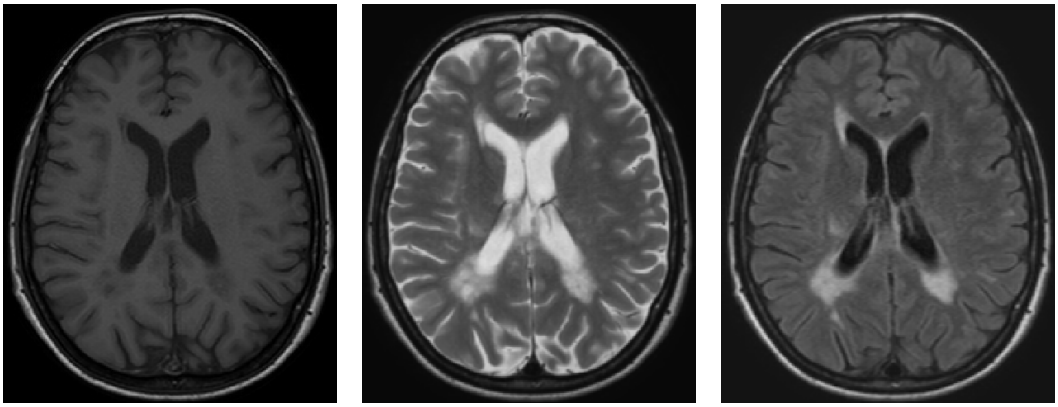


Figure 1.7: Diagram showing signal intensities of various tissues at T1- and T2-weighted imaging (source: [Bitar et al., 2006])

Diagram 1.7 illustrates that T1-weighted images are effective in displaying the anatomy and differentiating between different types of tissue. When contrast material is administered, they can also highlight pathological entities. T2-weighted images, on the other hand, are particularly useful in identifying disease as they make affected areas appear brighter due to the increased water content in pathological tissue.

In multiple sclerosis brain MRIs, the brain lesions often have poor tissue-to-lesion contrast in T1-weighted images, whereas they appear as bright as cerebrospinal fluid (CSF) in T2-weighted images, as depicted in Figure 1.8. Another MRI acquisition modality called FLAIR (FLuid Attenuated Inversion Recovery) can be used to cancel free water signal from the CSF tissue. This sequence can be carried out by applying a 180° pulse to flip the NMV and vanish the signal from a particular entity. Figure 1.8 illustrates the differences in inter-tissue contrast and lesion to tissue contrast for three MRI modalities.



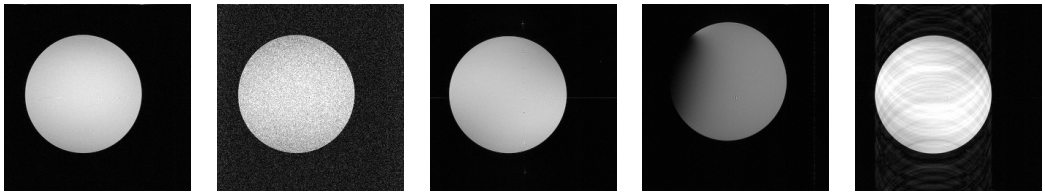
(a) (b) (c)

Figure 1.8: (a) Brain MRI from a MS patient in T1-w modality, (b) Brain MRI from a MS patient in T2-w modality, (c) Brain MRI from a MS patient in T1-w modality. (source: patient18 from Lesjak's public dataset Lesjak et al. [2016])

Noise and MR artifacts In Figure 1.9, basic examples of noise and MR image artifacts are illustrated. Noise 1.9 (b) is present in MR images and is modeled by a Rayleigh distribution in regions without signal, and a Rician distribution in areas with signal. Nevertheless, noise is commonly modeled as Gaussian noise, which is a reasonable assumption if the signal to noise ratio is adequate.

Various artifacts can appear in MR images, which stem from different phenomena that happen during the image acquisition process. The sources of these artifacts include non-uniformities in the magnetic field (B_0) 1.9 (c), the radiofrequency (RF) field 1.9 (d), patient motion (such as breathing, blood flow, and heartbeats) 1.9 (e), or aliasing. To process images that show such artifacts, several methods exist. In automated MR image processing, these artifacts must be rectified to utilize the images effectively.

In summary, there are multiple modalities of MRI protocols which highlight different inter-tissue tissue contrasts. With T1-w, T2-w or FLAIR modalities, the intensity



(a) (b) (c) (d) (e)

Figure 1.9: Noise and different artifacts in MR images (source: http://chickscope.beckman.illinois.edu/roosts/carl/artifacts.html) (a) Reference image, (b) Random noise in a low signal-to-noise ratio acquisition (c) Geometric distortion due to B_0 field inhomogeneity, (d) Geometric distortion due to RF field inhomogeneity, (e) Motion artefact

values within the images do not correspond to a quantitative measure, although it is possible to perform relaxometry to get quantitative information with longer acquisition time. Therefore, normalizing intensity values is necessary to automatically process these images. The nature and quantity of noise, spatial resolution and partial volume effects (when more than one type of tissue is contained in a voxel), inhomogeneities in the B_0 and B_1 fields and geometric distortions due to the encoding of the images are also sources of changes that need to be addressed.

The main difficulty in designing a reliable automatic change detection algorithm is that many factors can influence the detection of focal MS pathology, including the strength of the magnetic field, the pulse sequences, the spatial resolution, the coil technology used, and the dose of contrast agent.

1.3 MS biomarkers in MRI

Detecting pathological changes during the course of the disease is a crucial problem for practitioners. T2-FLAIR MRI and sometimes T1-weighted with gadolinium contrast enhancement protocols are favored because they provide contrast between lesions and healthy tissue. Therefore, in addition to clinical observations, these MRI protocols are used for diagnosis, prognosis and monitoring of MS patients. The most relevant biomarkers to predict disease evolution and disability include white matter lesions, gray matter lesions, white matter damage, gray matter atrophy, gray matter damage and spinal cord lesions, [Brownlee et al., 2019; Lazeron et al., 2005; Wattjes et al., 2015]

In order to diagnose multiple sclerosis (MS), clinicians typically look for the presence of lesions in specific locations within the central nervous system, such as the brain, spinal cord, and optic nerves. They also consider the dissemination in time of the occurrence of these lesions, and rule out any other potential explanations for their presence. Over the course of the disease, MRI scans are used to track the progression of MS over time, as they are more sensitive in detecting disease activity compared to clinical metrics. Factors such as the location and number of lesions, as well as brain volume loss, can provide important insights into the stage and progression of the disease, and are used to monitor remission, relapse, response to treatment, and the development

of disabilities.

The visual comparison of two MRI scans to find new lesions is quite time consuming and error-prone. The practitioner's sensibility to new or evolving lesions can be improved by automated tools. Automated tools are sometimes provided on MRI console for MRI comparison [Altay et al., 2013] and commercially available products exist [van Leeuwen et al., 2021]. For example, Philips MRI scanners provide a tool that highlight change areas to facilitate visual inspection for practitioners.

While T2-FLAIR MRI modality provides good tissue-to-white matter contrast, it can still be difficult to detect lesions in the gray matter properly. Other MRI modalities such as Double Inversion Recovery (DIR), where fatty tissue and water signals are canceled, might be a promising option for detecting gray matter lesions, however it is not routinely implemented clinical settings.

In summary, multiple sclerosis (MS) is a chronic disease that progresses over time, and MRI is an essential diagnostic and monitoring tool for patients. However, analyzing the evolution of lesions can prove difficult for healthcare practitioners, and automated processing of longitudinal acquisitions is challenging due to the presence of various non-interest sources of change that need to be corrected. This thesis concentrates on the detection of emerging, vanishing, expanding, or contracting lesions in both white-matter and gray-matter tissue. The following chapter addresses how the different sources of changes in longitudinal MRI are handled in the state-of-the-art to achieve automatic lesion change detection.

Chapter 2

State-of-the-art related to change detection in brain MRI

Changes in longitudinal MRI acquisitions can arise from both pathological and non-pathological causes. The pathological changes observed in MRI of MS patients are lesions load evolution and brain tissue atrophy. However, detecting these changes can be challenging, even for experienced practitioners, and is subject to some degree of subjectivity. This is where automated tools can be helpful. Nevertheless, automatically detecting these changes is still a challenging task, as non-interest changes can hinder performance. For instance, variations due to patient positioning, image aspect, and acquisition conditions need to be ruled out to achieve satisfactory automatic detection. These non-pathological factors can sometimes mask or mimic the pathological changes of interest, making it difficult to differentiate between them. As a result, developing robust and accurate automated tools for detecting pathological changes in longitudinal MRI acquisitions remains an active area of research. Such tools could potentially improve the accuracy and reliability of disease diagnosis, follow-up, and treatment monitoring, leading to better outcomes for patients with MS and other neurological disorders.

The objective of this chapter is to present the challenges associated with detecting changes in longitudinal MRI, with a particular emphasis on evolving MS lesions, as well as the available solutions for addressing these challenges. The first part of the chapter provides a brief description of the classical processing pipeline for detecting changes in MS lesions. The second part presents non-interest sources of change that can occur in MRI acquisitions and how they are typically corrected. The following section introduces approaches for compensating for non-interest sources of change. Methods for registration in longitudinal analysis are presented in the subsequent part. Finally, the last section of the chapter focuses on the state-of-the-art approaches for identifying changes in MS lesions.

2.1 Classical workflow for change detection in brain MRI

To isolate relevant changes in longitudinal brain MRI, it is necessary to correct for multiple sources of change such as differences in MRI acquisition conditions, patient

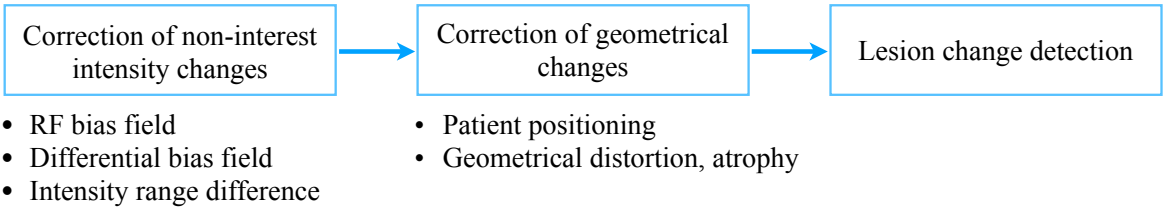


Figure 2.1: Workflow for lesion change detection.

positioning, and pathological changes. The standard workflow for analyzing these changes is a sequential approach that consists in estimating and correcting successively each source of change. The steps of this workflow can be grouped in three main categories outlined in Figure 2.1. First, global intensity changes that are not relevant to the analysis, such as RF bias field, differential bias field and difference of gray-level distribution between both MRI acquisitions, are corrected. Next, the images are spatially registered to correct patient positioning and geometric distortion, and finally focal intensity changes due to the evolution of lesions are detected. In the following sections of this chapter, we describe the state-of-the-art techniques at each step of this workflow.

In the context of multiple sclerosis, the two primary changes of interest are the evolution of lesion load and brain atrophy. However, many automatic change detection algorithms are designed to exclusively detect lesion changes, thus making the presence of atrophy only an obstacle to the performance of the algoriths and not an interpretable phenomenon that has to be estimated.

2.2 Non-interest sources of change

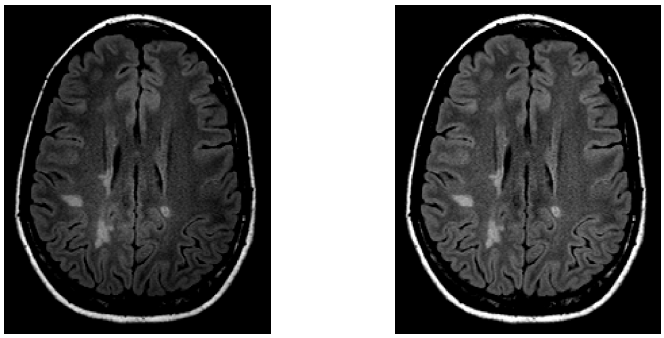
In this section, we introduce the standard image processing techniques used to reduce the impact of the non-interest sources of change in longitudinal MRI. As described in Section 2.1, they include intensity changes related to the MRI acquisition parameters and the displacements due to different patient positioning.

2.2.1 Changes due to MRI acquisition conditions

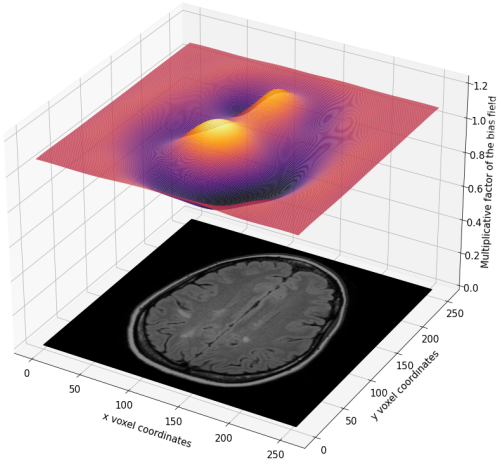
Inhomogeneity of the magnetic field The magnetic field applied during MRI acquisitions is inhomogeneous, which results in low-frequency intensity variations in the images. This type of artifact may be visible to the human eye, but it does not significantly impact the ability of a medical professional to interpret the image. However, it can alter the performance of automated analysis algorithms. As a result, it is necessary to use correction algorithms, such as the standard N4 bias field method [Tustison et al., 2010]. The N4 algorithm iteratively alternates deconvolution of the histogram of the image and estimation of the bias field represented with B-spline basis functions [Larsen et al., 2014]. The estimated bias is then used to correct for inhomogeneities in the magnetic field. An example obtained with the implementation in ANTs software suite is shown

in Figure 2.2.

Despite the individual correction of the bias field in each image, a residual bias might persist in the difference image of the same patient. To address this issue, the approach outlined in [Lewis and Fox, 2004] can be employed to correct the differential bias field, provided that images are registered.



(a) MRI T2-FLAIR slice (b) Corrected MRI T2-FLAIR slice



(c) Bias field as calculated by the N4 Algorithm

Figure 2.2: Illustration of N4 bias field inhomogeneity correction algorithm

Intensity discrepancies In an MRI image, each voxel is assigned a specific intensity that reflects the tissue density or composition. As explained in Section 1.2, the intensity of a voxel does not always represent an interpretable physical quantity. One can perform T1 relaxometry (in this case, a physical value is measured) or T1-weighted imaging, which is non-quantitative but allows highlighting contrast differences between tissues representing different T1 values. The contrast levels depend on the settings used for TE, TR and TI. As a result, the specific gray-level intensity values and the amount of contrast between tissues in an MRI image may vary from one scan to another. Although it does not affect the visual interpretation of the images, it can have a strong impact on the performance of an automatic change detection method. To be able to compare MRI images from the same patient over time, it is necessary to standardize the intensities.

One common approach for doing this is the Nyul histogram standardization [Nyúl et al., 2000]. It involves fitting a piecewise linear function on the histogram of an image using a division of the intensity range in quantiles. This method is effective at addressing both intensity range and contrast differences, and it is not significantly affected by the presence of lesions. However, it is not always used in automated change detection pipelines for detecting MS lesions. Some pipelines use simple linear fitting of intensities in the joint histogram of two images, but this type of approach is more sensitive to the presence of lesions. Histogram matching does not require prior registration of images, unlike techniques that employ joint histogram analysis. Other pipelines scale the intensities within a defined range or around an arbitrary median value.

Geometrical discrepancies The voxel size determines the spatial resolution of MRI acquisitions, and it can vary from one scan to another. In order to compare two images automatically, they need to be resized to have the same voxel size. The downside of the resizing operation is that it requires interpolation, which can degrade the quality of the images. To mitigate this effect, all the necessary transformations (image deformation or resizing) can be applied to each image at once with bi-cubic spline interpolation. Inhomogeneities in the RF field can create geometric distortions that can also be corrected with deformable registration.

2.2.2 Patient-related changes

Correction of patient positioning Longitudinal MRI acquisitions of a single patient are usually misaligned due to differences in patient positioning, as illustrated in Fig. 2.3. In order to achieve voxel-wise comparision of the images, it is necessary to perform a registration step that aims at aligning anatomical regions using a geometrical transformation. Simple displacements due to global positioning can be captured by rigid or affine models, which are the most commonly used in practice. However, these transformations do not capture other more complex and subtle deformations such as atrophy or geometric distortions due to MRI acquisition.

Removing changes outside areas of interest with brain and tissue segmentation MRI of the brain used for monitoring multiple sclerosis (MS) are acquired several months to a year apart. During this time, there can be changes in the head unrelated to the disease, such as weight loss or gain, or changes in the position of the neck, which can affect the appearance of the brain and spinal cord in the images. If these types of changes are not taken into account, they can lead to false detections of changes in MS lesions when analyzing the images. To address this issue, brain extraction algorithms are commonly used to create masks of the brain and restrict analyses of changes to this region of interest. This can help to minimize the impact of obvious non-interest anatomical changes on the automated detection of changes in MS lesions.

Some of the widely used implementations are FSL-Brain Extraction Tool [Jenkinson et al., 2012; Smith et al., 2004], Freesurfer[1], SPM, Samseg [Cerri et al., 2021], ROBEX

[1]https://surfer.nmr.mgh.harvard.edu/

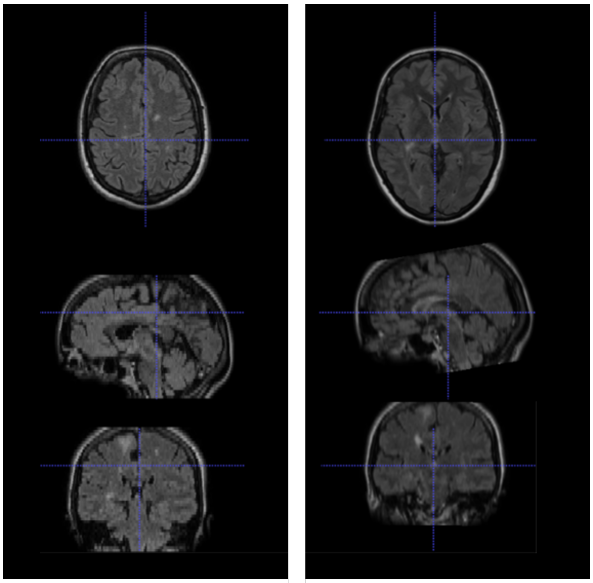


Figure 2.3: Two brain MRI from the same patient

[Iglesias et al., 2011] (used by [Ganiler et al., 2014], [Cabezas et al., 2016], [Salem et al., 2020].

Automated change detection approaches for Multiple Sclerosis (MS) lesion analysis can be improved by using a tissue segmentation software to extract and differentiate the types of tissue in the brain, including white matter, gray matter, and cerebrospinal fluid. This can help to rule out detections that are unlikely to exist, as brain tissue may be impacted over time by factors such as patient hydration or brain atrophy. Examples of brain parenchyma classification software include FSL-FAST, Freesurfer, SAMSEG [2], and the SPM software suite on Matlab. Some automated change detection approaches [Cabezas et al., 2014; Ganiler et al., 2014; Salem et al., 2018] focus the search on white-matter, as it is the primary location of MS lesions. However, as stated in 1.1 they can also be found in gray matter. Therefore, it may be beneficial for automated change detection approaches to consider these tissue types as well.

2.3 Registration of longitudinal brain MRI

In this section, we focus on the registration step of the sequential pipeline outlined in Section 2.1. The estimation of the deformation field between successive images can provide a quantitative measure of brain atrophy, which is essential for monitoring the disease, along with the assessment of the lesion load evolution.

The primary objective of registration in medical imaging is to determine a geometric transformation that maps corresponding anatomical or functional areas between two images. This process is crucial for combining information from various medical image modalities to enhance the accuracy of diagnosis and treatment planning. It is also

[2]https://surfer.nmr.mgh.harvard.edu/fswiki/Samseg

useful in longitudinal studies, where images taken at different times are compared to identify tissue changes or abnormalities. Additionally, registration can aid in analyzing a set of images to comprehend the anatomical variations within a population.

Registration models can be distinguished by four key aspects, which are: the transformation models, the similarity criteria that guide the mapping, the regularization that promotes certain desirable properties of the estimated deformation field, and the optimization algorithms implemented to find the optimal transformation. A survey of registration in medical imaging regarding the aforementioned points is proposed in [Sotiras et al., 2013] and an experimental comparison of their performances was carried out by [Klein et al., 2009]. In this section, we mainly focus on two key aspects of registration in MRI in the case of MS lesion change detection, namely transformation models and the similarity criteria.

2.3.1 Transformation models

Parametric transformations Simple parametric models such as rigid or affine are able to represent coarse image deformations. Rigid registration allows for global translation and rotation of the image, while affine registration allows shearing and scaling as well. These transformation models can be used in same-patient studies, either as initialization for more complex deformable registration models or as a preliminary step of MS lesion change detection.

Deformable registration Parametric transformation do not capture local changes in images and are not suitable for representing complex deformations. In MRI registration, especially in longitudinal brain MRI analysis, it is often relevant to capture local changes such as brain atrophy, local tissue changes, or non-linear deformations due to physiological changes or disease progression. Dense deformation fields can model arbitrarily complex local changes in images. They are represented as dense vector fields that map each point in the image to its corresponding point in the other image. In order to impose the vector field to respect physically plausible constraints, several strategies have been designed:diffeomorphic transformations [Noblet et al., 2005; Vercauteren et al., 2009] ensure that the topology of the image is preserved, symmetric transformations [Noblet et al., 2012] impose that the same deformation is applied to both images, and differomorphism (one-to-one mapping) constraints preserve certain properties of the image during registration such as number of connected components.

In case of brain atrophy, as shown in Figure 2.4, deformable models are required to obtain a deformation field able to invert the atrophy effect and make the registered brain match the reference brain. The counterpart of this perfect matching is the altetation of the apsect of appearing and evolving lesion in the deformed image, which is a hindrance in lesion change detection. This phenomena is illustrated in Figure 2.5, where the follow-up acquisition displays lesion growth, and the registered follow-up to baseline image shows a significant alteration of the lesion appearance. Table 2.1 sums up the effect of transformation models on lesion change detection. It should be noted that certain methods for detecting changes in multiple sclerosis lesions rely on

exploiting properties of the deformation field in longitudinal MRI. In that particular case, overcompensation of lesion appearance is a desirable feature.

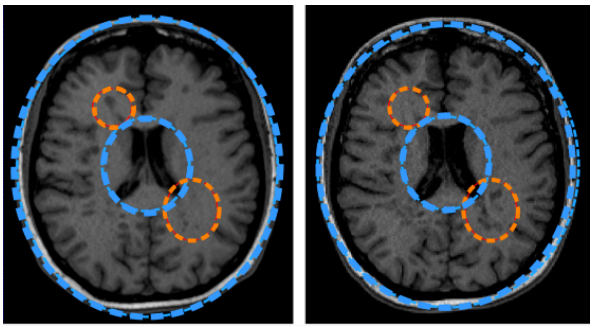
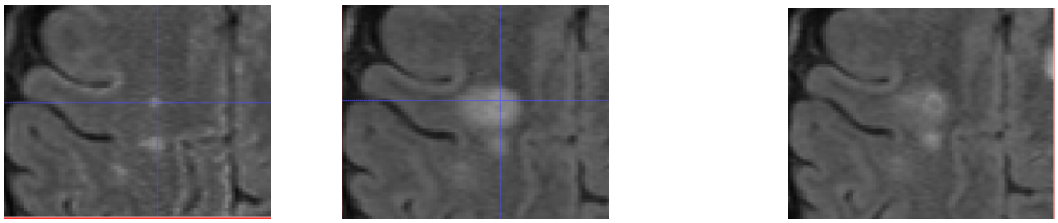


Figure 2.4: Two successive T1-w brain MRI with atrophy: the enlargment of the ventrcles, in the smaller blue circle is visible (blue) and lesions (orange)



(a) Lesion in baseline acquisition (b) Lesion growth in follow-up acquisition (c) Follow-up registered onto baseline with deformable registration

Figure 2.5: Lesion overcompensation phenomenon with deformable registration

Type	**Rigid or affine**	**Deformable**
State of the art usage	Standard	Rarely used
Atrophy compensation	No	Yes
Consequences	False detections due to atrophy: Lack of specificity	Compensation of changes of interest: Lack of sensitivity

Table 2.1: Registration types and their consequences on change detection

2.3.2 Similarity criteria

The registration process is guided by a measure of the similarity between the source image and the deformed target image. This similarity criterion can be intensity-based, comparing directly the intensity values of the images, or landmark-based, using correspondences of anatomical features or landmarks. Its choice is important because it can affect the accuracy and outcome of the registration process. We will primarily focus on intensity-based methods, as they are the most relevant in the context of longitudinal MRI studies (for a survey, see [Diez et al., 2014]).

In longitudinal brain MRI analysis, there are two possibilities: multi-modal registration, where images are acquired with different MRI protocols, and mono-modal registration, where images are acquired with the same MRI protocol. In multi-modal registration, the images have different intensity distributions and contrasts, while in mono-modal registration, the variations in intensity are caused by misalignment or tissue changes. While direct comparison of longitudinal acquisitions from different modalities is not feasible for change detection, information from these acquisitions can be integrated via data fusion methods.

Mono-modal registration In the mono-modal case, the matching criteria can be devised with the assumption that anatomical structures have similar intensity values across acquisitions. Therefore, sum of squared differences (SSD) or sum of absolute differences (SAD) can be considered as suitable choices. The SSD penalizes intensity differences between the two images at each voxel with a ℓ_2 norm. The SAD is similar to SSD, but it replaces the ℓ_2 norm by ℓ_1. Both SSD and SAD are sensitive to intensity non-uniformity, noise, and they are not robust against large intensity variations. Another possibility is to use cross-correlation if there is a linear relationship between signal intensities in the same anatomical structures over time. It is based on the correlation between the intensity values of the images and they can be sensitive to the signal-to-noise ratio (SNR) of the images.

Multi-modal registration Since the intensity distributions of each anatomical structure can be different across modalities, finding an appropriate similarity critieria is more challenging. Two main approaches exist to solve this problem: information theoretic measures, based on the statistical dependence between the intensity values of the images, and reduction to a mono-modal problem by simulating one modality from another or mapping both modalities to a common domain through intensity normalization. The most frequently used methods are information theoretic approaches, as they are relatively robust to image noise. Mutual Information (MI) is very often used in brain MRI registration. An intuitive way to understand this criteria is to see it as measure of the ability of one image to provide information about the other image. The mutual information between two images, I_{ref} and I_{mov}, is defined as the amount of information that image I_{ref} provides about image I_{mov}, and vice versa. It can be written as:

$$MI(I_{ref}, I_{mov}) = H(I_{ref}) + H(I_{mov}) - H(I_{ref}, I_{mov}) \tag{2.1}$$

where $H(I_{ref})$ and $H(I_{mov})$ are respectively the entropies of the reference and moving images, and $H(I_{ref}, I_{mov})$ is the joint entropy of the two images.

2.3.3 Registration in the presence of temporal inconsistencies

Similarity criteria discussed in Section 2.3.2 are based on the assumption that every point in the moving image has a corresponding point in the reference image. However, this assumption is not always true, especially when healthy brain regions at a given timepoint become filled with a pathological lesion in the follow-up image. Non correspondences

can also occur when brain tumors deform surrounding tissue or in post-resection brain images. To handle registration in such cases, some studies have addressed the issue of missing correspondences by separating the modifications induced by the lesion from other sources of changes. One approach is to segment the lesions and exclude them from the similarity criteria as in [Brett et al., 2001; Stefanescu et al., 2004], but accurate lesion segmentation is difficult to achieve in practice. To address this problem, some studies have investigated joint registration and segmentation models that simultaneously register the images and segment regions of inconsistent intensity matching. These approaches are closely related to our current work.

These joint models can be guided by biomechanic tumor growth models, as in [Gooya et al., 2011; Kwon et al., 2014]. For pre-operative and post-resection MRI registration, level sets have been employed to model the regions of interest, as shown in [Yezzi et al., 2003]. In [Risholm et al., 2009], registration was performed using a variant of the Demons algorithm that is bijective and incorporates anisotropic diffusion to restrict diffusion across the boundaries of segmented areas. Other studies have included a segmentation objective function in the registration objective to estimate a discrete segmentation map, using a fast primal-dual graph cut strategy, as in [Chen et al., 2015; Parisot et al.]. In [Hutchison et al., 2010; Lu et al., 2011], the missing correspondence regions, valid regions, and background were treated as marginalized latent variables in an EM algorithm for estimating the transformation.

2.4 Change detection

In this section, we review the different approaches for change detection. There are roughly three categories of MS lesion change detection methods [Lladó et al., 2012]: intensity-based, deformation-based and joint segmentation. Intensity-based methods generally rely on image subtraction to detect changes in lesion intensity between timepoints. Deformation-based methods exploit deformation-fields computed through deformable image registration to detect changes in lesion shape. When a lesion appears or disappears, the deformation field shows specific patterns. By calculating the Jacobian at each point in the image, contractions and dilations can be identified. In the work of [Rey et al., 2002], they exploit this phenomena by thresholding the contractions and dilations in the deformation field to detect significant change. Pieperhoff et al. [2008] analyze the local volume variation (ratio between deformed voxels in a source image and non-deformed voxels in a target image). Joint segmentation combines information from both timepoints to classify lesion evolution. These methods can be either supervised, where the model is trained on labeled data, or unsupervised, where the model does not require training with labeled data. Methods that are purely deformation-based methods are not pursued anymore because their sensitivity is insufficient, as pointed out in [Lladó et al., 2012]. Therefore, we focus in this section on reviewing intensity-based and longitudinal segmentation approaches.

2.4.1 Intensity-based methods

The starting point of most intensity-based methods is a subtraction between the two registered images. The different methods can be distinguished by the rules applied to find relevant detections from this subtraction image. The most intuitive and straightforward rule is to apply a threshold to the subtraction image. However, this simple approach is only valid if all changes are of interest. This section describes the different techniques designed to achieve more robust detection.

2.4.1.1 Integration of spatial information and rules

A voxel-wise approach assumes that each voxel is independent from its neighbors. However, lesions in medical images are clusters of voxels, thus, it is relevant to take into account other contextual information. To achieve this, some approaches opt to consider patches of a certain size, and others use image filtering to include context information such as contrast or texture. The choice of neighborhood size comes with a trade-off between sensitivity and robustness to false detections. Large neighborhoods bring robustness to noise, but can also alter the accuracy of the detection.

Bosc et al. [2003] proposed a method for change detection that utilizes a statistical model in a 3×3 voxel window in the difference image, and use a Generalized Likelihood Ratio Test (GRLT) to determine if a change is of interest. The detections are presented to the user in descending order of likelihood and require validation through a yes/no interface.

Ganiler et al. [2014] conceived a method for detecting appearing white-matter lesions in brain MRI images by applying a threshold to a subtraction image and then applying rules to the resulting binary image. These rules are based on neighboring voxel intensity features and are applied to remove irrelevant detections. Some false-positive regions are caused by inaccuracies in skull and white-matter extraction masks. They remove these areas by calculating the mean and standard deviation of the baseline image in the detected regions and removing regions that have intensities lower than mean - 2*standard deviation. They also take local neighborhood information into account by removing regions where the ratio between the intensities in the detected regions and surrounding tissue does not meet certain thresholds in baseline and follow-up images. The work of [Cabezas et al., 2016] uses similar rules and adds others based on deformation-field features in the detected areas.

Battaglini et al. [2014] purposefully over-detect lesion change on subtraction PD modality images, and then use a set of rules similar to those of [Ganiler et al., 2014], in addition to other features such as shape, orientation, extent of detected regions and contrast in the surrounding areas.

The work of Cheng et al. [2018] introduces a model that utilizes multiple scales of information such as voxel-wise, patch-based, and local texture based information. It calculates change maps from multiple MRI modalities (T1, T2 and FLAIR) by subtracting the z-scores of each scan, thus taking into account the variation in gray level intensities.

Elliott et al. [2013] propose a supervised method for detecting new lesions. The

MRI images are jointly segmented using a two-step classification process. Firstly, brain tissue at both timepoints is classified at the voxel-level using a Bayesian framework that considers image intensities and difference images. To incorporate spatial information and refine the classification, they employ a Markov random field framework and calculate a posteriori distributions using a soft ICM. After this classification, some voxels may be misclassified as new lesions. These voxels are then passed through an additional Random Forest classification that uses features based on intensity, context, size, registration quality, shape descriptors.

2.4.1.2 Use of multiple MRI channels

Using multiple MRI sequences can help to take advantage of the complementary information they provide and compensate for their limitations. For example, FLAIR imaging can provide high tissue-to-lesion contrast but has relatively low signal-to-noise ratio and low image quality. T1-w imaging can provide good inter-tissue contrast but lower tissue-to-lesion contrast, it can also help to reduce the number of false detections. Combining multiple protocols increases the amount of information and makes the approach more robust to noise as it is not correlated across modalities, thus reducing the number of false-positive results.

The Generalized Likelihood Ratio Test (GRLT) developed by Bosc et al. [2003] can be extended to incorporate multiple MRI sequences. Specifically, the subtraction images of each modality combination are weighted using the noise covariance matrix obtained from that particular modality combination.

The SuBLIME (Subtraction-Based Logistic inference for Modeling and Estimation) described in [Sweeney et al., 2013] is a logistic regression model derived from FLAIR, PD, T1 and T2 MRI modalities and their subtraction images. They found that the use of multiple MRI modalities improves new and enlarging lesion detection.

The approach proposed by [Cabezas et al., 2016] and its extension by [Salem et al., 2020] differentiate between true and false positives by using the intersection of masks computed using PD, T2, and FLAIR sequences.

2.4.1.3 Use of deformation field features

The approaches that use analysis of deformation fields take advantage of the behavior of deformable registration algorithms. While purely deformation-based approaches lack sensitivity, the deformation field properties can be interesting information to add, as in [Salem et al., 2018] and [Cabezas et al., 2016] where they are used to refine the results of purely subtraction-based approaches.

Cabezas et al. [Cabezas et al., 2016] builds upon the approach of Ganiler et al. [2014] by incorporating a post-processing step that utilizes features computed from the deformation field between consecutive scans. In regions of lesion evolution, the deformation vectors tend to converge towards the center of a lesion. This behavior can be observed by computing the Jacobian determinant matrix, divergence, and concentricity of the deformation field. Additional rules based on these characteristics further reduce the number of false detections.

Salem et al. [Salem et al., 2018] improves upon the approach of Sweeney et al. [2013] by incorporating deformation field characteristics such as Jacobian and divergence in the logistic regression model. The effectiveness of adding these deformation field characteristics is evaluated by comparing the detection results obtained with models that use different combinations of these characteristics.

2.4.1.4 Supervised learning approaches

Supervised approaches for detecting changes in lesions rely on training a model that describes lesion evolution using real data, feature images, and manual expert annotations. The availability of a large number of expert annotations is essential to obtain accurate models. The goal of these methods is to calculate the probability that each voxel belongs to a lesion change region. The differences between the methods include the size of the training sets, the protocols used to obtain the annotations, and the choice of descriptors. Examples of such approaches can be found in [Elliott et al., 2013; Salem et al., 2020; Sweeney et al., 2013].

2.4.2 Longitudinal segmentation

Consecutive timepoint lesion segmentations from the same patient often lack consistency, which means that the difference between the two segmentations cannot be considered a reliable change map. Longitudinal segmentation aims at improving lesion segmentation and model lesion growth by incorporating temporal information. To address this limitation, the methods described below show different approaches to combine segmentations of consecutive patient MRI.

MSMetrix-long, proposed by Jain et al. [2016], consists of three steps: first, brain tissue is segmented into white matter, gray matter and cerebrospinal fluid in the FLAIR modality using the MSMetrix-cross tool. Lesions are identified as outliers in the white matter intensities. Next, an initial change map is computed by subtracting the segmentations from the two time points. The change map is then refined using an Expectation-Maximization algorithm, and changes are classified into three categories: shrinking, growing, and stable. The effectiveness of this algorithm depends on the quality of individual timepoint segmentations. If false detections are present in both acquisitions, they will persist after optimization.

Schmidt et al. [Schmidt et al., 2019] also classified lesion evolution into growing and shrinking. They initially created a change map using a logical disjunction of independent lesion segmentations from two acquisitions. A common mask was established by dividing the white matter region into voxels that belonged to a lesion in either time point and voxels that belonged to no lesion in any time point. The change maps were updated when the intensity differences in FLAIR reached a threshold δ. This threshold was determined by approximating the distribution of intensities in white matter without lesions and depended on another threshold α, which characterized a range in this distribution.

In [Bernardis et al., 2013], joint segmentation of two brain MRIs was formulated as a 4D binary graph-cut (3D+t). The images were registered, and a 3D graph was

constructed for each image, where the weights between each vertex were determined by computing the "affinity" between voxels (attraction and repulsion). The correspondences between the two images were then established, and the corresponding voxel at time t was searched for at time $t-1$ in a neighborhood r. The affinities were aligned, and the graph was segmented with spatial grouping criteria. It is worth noting that this framework can be adapted to work with more than two time points.

2.4.3 Deep learning

Recent studies have demonstrated the efficacy of CNNs, particularly U-Net [une, 2015] architectures, in segmenting biomedical images. However, while there are several techniques to detect lesion loads in MS scans from a single timepoint, there are fewer methods addressing the detection of MS lesion activity through longitudinal analysis.

[Birenbaum and Greenspan, 2017] utilize a CNN to reduce false positive detections obtained through simple thresholding of the MRI scans. [Salem et al., 2020] proposed an approach that learns simultaneously the deformation field between baseline and follow-up images and the segmentation mask. [Denner et al., 2020] also proposes a multitask learning approach that comprises both deformable registration and change detection. One of the main limitations to the implementation of such approaches is the access to multiple timepoints data; the proposed approaches were trained and tested on in-house annotated datasets. In 2021, [Commowick et al., 2017] proposed a longitudinal new MS lesion segmentation challenge in MICCAI 2021 , providing access to a public dataset of 100 patient from the OFSEP HD cohort, and thus generating publications of new approaches. 29 approaches were proposed during the challenge, and a research topic in the Frontiers in Neuroscience and Frontiers in Neuroimaging was created[3] where 10 articles were accepted.

Most approaches are based on U-Net architectures with variations in either pre-processing techniques, data augmentation schemes and exploitation of image features. For example, Hitziger et al. [2022] process their images slice-wise along coronal, axial and sagittal orientations and fuse the predictions using unanimous voting. Data augmentation schemes include classical image transformations, lesion-aware schemes [Andresen et al., 2022; Basaran et al., 2022], image synthesis [Kamraoui et al., 2022; Plassard et al., 2018; Valencia et al., 2022], or in-house datasets. Interestingly, some chose to design schemes around nnU-Net [Isensee et al., 2018], an "out-of-the-box" self-adapting framework for U-Net-based image segmentation, and found that a crucial step for successful results is proper pre-processing of MRI images [Basaran et al., 2022] that also includes image registration.

It is worth noting that three methods actually consider the usage of deformation fields to improve detection results and reduce false-positive rates.

[3]https://www.frontiersin.org/research-topics/25869/automatic-methods-for-multiple-sclerosis-new-lesions-detection-and-segmentation#articles

2.4.4 Limitation of the sequential approach

The sequential processing of registration and change detection in subtraction-based and deformation-based methods has some limitations, as illustrated in Fig. 2.6 showing the baseline (Figure 2.6 (a)) and follow-up (Figure 2.6 (d)) MRI acquisitions of a patient suffering from MS. One can observe in the follow-up scan the apparition of a new lesion and a slight enlargement of the ventricle reflecting the brain atrophy process. In the case of affine registration (Fig. 2.6 (b) and (e)), we can see on the subtraction image that the lesion is well detected, but that spurious detection occur around the ventricles and in the cortical regions due to brain tissue atrophy. Using a deformable registration (Fig. 2.6 (c) and (f)) helps to remove these spurious detection by compensating the ventricles enlargement and the cortical atrophy. However, it also tends to make the new lesion disappear (Fig. 2.6 (c)), thus altering the shape of the corresponding detection (Fig. 2.6 (f)).

Despite the interdependence between registration and change detection, current approaches handle these problems separately. Without proper registration, accurate change detection cannot be achieved, but the presence of changes in an image can hinder registration. We hypothesize that decoupling these problems is the cause of these limitations. The contribution of this thesis is to propose a solution that performs atrophy correction accurately while preserving the shape of appearing and evolving lesions. In the following chapter, we present our joint registration and change detection approach to achieve this.

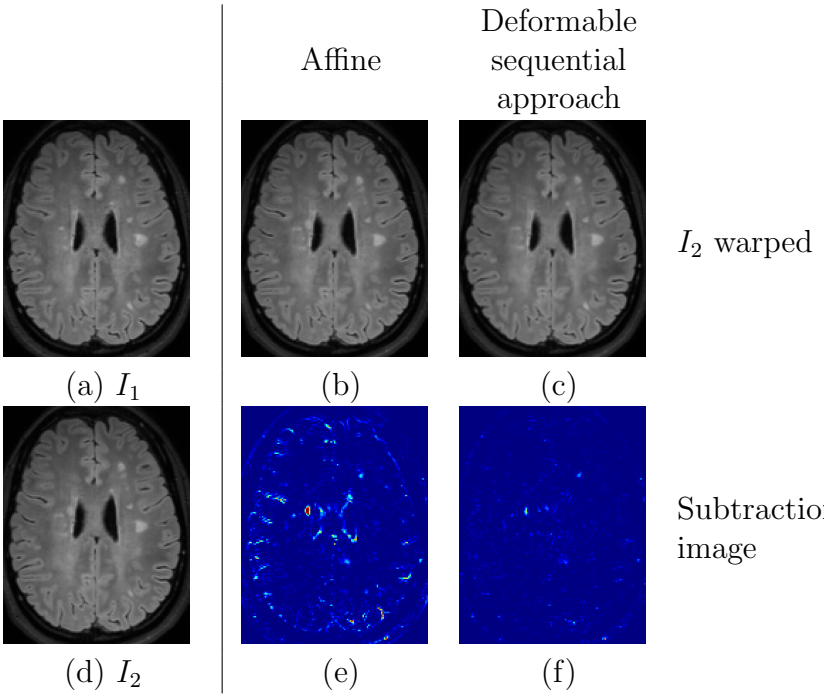


Figure 2.6: Warped follow-up images and subtraction images obtained with the sequential affine and deformable registration pipelines and the proposed joint deformable approach, on an example of the dataset described in [Commowick et al., 2021].

Chapter 3

Unified framework for registration and change detection

In the previous chapter, we identified the limitations of addressing registration and change detection sequentially as two distinct problems. In this chapter, we propose to model the relation between these two tasks in a unique cost function. Registration and change detection are performed jointly by minimizing this cost function, which allows us to overcome the flaws of the sequential approach. In the first section of this chapter, we formulate our joint model, and we detail the optimization procedure in the second section.

3.1 Formulation of the joint model

3.1.1 General formulation

In this section, we first provide a general formulation of registration and change detection as minimization problems, and then we show how to combine them in a single cost function. In what follows, we denote $I_1, I_2 : \Omega \to \mathbb{R}$ the baseline and follow-up MRI images, where $\Omega \subset \mathbb{R}^3$ is the image domain.

3.1.1.1 Registration

The registration of I_1 and I_2 can be formulated as the minimization problem

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_1 \Psi(\mathbf{w}), \tag{3.1}$$

where $\mathbf{w} : \Omega \to \mathbb{R}^3$ represents the deformation field, $\rho(\cdot)$ is a data similarity term, and $\Psi(\cdot)$ is a regularizer weighted by a scalar $\lambda_1 > 0$. The choices of the transformation model of $\mathbf{w}$ and the data term $\rho(\cdot)$ are discussed in Sections 2.3.1 and 2.3.2, respectively.

As mentioned in Section 2.3.1, in the context of MS it is important to consider a non-parametric dense deformation model (one displacement vector is estimated at each pixel) in order to capture complex deformations due to brain atrophy. If the cost

function is defined only by the data term ($\lambda_1 = 0$ in (3.1)), assuming a dense deformation makes the problem (3.1) underdetermined because of the high number of unknowns (three times the number of voxels). Therefore, the solution has to be constrained with an additional regularization term $\Psi(\mathbf{w})$ that enforces desirable properties on the deformation field, in order to make the problem well-posed. Standard regularization promote smooth solutions by penalizing the ℓ_2 or ℓ_1 norm of the gradient of $\mathbf{w}$.

3.1.1.2 Change detection

As detailed in Section 2.4, the change detection step generally consists in thresholding a map of feature differences between registered baseline and follow-up images. Thus, this threshold can be considered as the detection of the regions of imperfect matching between I_1 and I_2 after registration. Since the registration error is quantified by the data similarity term $\rho(I_1, I_2, \mathbf{w}, \mathbf{x})$ in (3.1), we can also use it for change detection to account for the intertwining of the two tasks. Accordingly, the binary change map $c : \Omega \to \{0, 1\}$ can be defined at each voxel $\mathbf{x}$ as follow:

$$c(\mathbf{x}) = \begin{cases} 0 & \text{if } \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) \leq \lambda_2 \\ 1 & \text{otherwise,} \end{cases} \tag{3.2}$$

where $\lambda_2 \in \mathbb{R}^+$ is the detection threshold. The thresholding scheme (3.2) can be reformulated as the following optimization problem:

$$\widehat{c}(\mathbf{x}) = \underset{c:\Omega\to\{0,1\}}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} (1 - c(\mathbf{x}))\ \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_2 c(\mathbf{x}). \tag{3.3}$$

Since simple thresholding can yield noisy results, most MS lesion change detection methods also integrate a denoising step in post-processing to obtain the final change map. The denoising can be realized jointly with the change detection by integrating a regularization term $\Phi(\cdot)$ in (3.3):

$$\widehat{c}(\mathbf{x}) = \underset{c:\Omega\to\{0,1\}}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} (1 - c(\mathbf{x}))\ \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_2 c(\mathbf{x}) + \lambda_3 \Phi(c), \tag{3.4}$$

where $\lambda_3 \in \mathbb{R}^+$ weights the regularization term.

3.1.1.3 Joint model

To overcome the limitations of the sequential approach, we propose a joint modeling of registration and change detection. The two steps are fundamentally intertwined, since registration aims at finding correspondences between images, while change detection determines regions that does not admit correspondences. Therefore, both tasks should be defined with the same objective function to work in synergy. We formulate the following joint minimization problem that achieves this goal by unifying the principles described previously:

$$\widehat{\mathbf{w}}, \widehat{c} = \underset{\mathbf{w},c}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} \left[(1 - c(\mathbf{x}))\ \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_2 c(\mathbf{x})\right] + \lambda_1 \Psi(\mathbf{w}) + \lambda_3 \Phi(c). \tag{3.5}$$

With this model, the data term is cancelled in change regions (where $c(\mathbf{x}) = 1$), so that the estimation of the transformation is only driven by the regularization term $\Psi(\mathbf{w})$, thus producing smoothed deformation field in these areas.

3.1.2 Specification of the cost function terms

The formulation (3.5) is generic and could be instantiated with a variety of data and regularization terms. In this chapter, we aim at demonstrating the superiority of the joint formulation over the sequential approach under standard modeling choices.

Firstly, we assume that after intensity normalization performed in preprocessing, intensities of both images are comparable. Thus, we consider a standard data term that penalizes deviations from intensity constancy with a ℓ_2 norm:

$$\rho(I_1, I_2, \mathbf{w}, \mathbf{x}) = \frac{1}{\sigma^2} \|I_2(\mathbf{x} - \mathbf{w}(\mathbf{x})) - I_1(\mathbf{x})\|_2^2, \tag{3.6}$$

where σ is a normalization constant defined by the median absolute deviation of the intensity differences between I_1 and I_2. This data term is representative of intensity-based features commonly used in change detection methods.

Secondly, we assume that the deformations induced by brain tissue atrophy are complex but still locally smooth. This is why we consider a first order Tikhonov regularization term that penalizes the gradient of the deformation field:

$$\Psi(\mathbf{w}) = \sum_{\mathbf{x} \in \Omega} \|\nabla \mathbf{w}(\mathbf{x})\|_2^2, \tag{3.7}$$

where $\nabla \cdot$ is the gradient operator.

Finally, to ensure spatially coherent lesion detection, we regularize the change map c with a standard binary Potts model:

$$\Phi(c) = \sum_{\mathbf{x} \in \Omega} \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} (1 - \delta(c(\mathbf{x}), c(\mathbf{y}))), \tag{3.8}$$

where δ is the Kronecker function equal to 1 if its argument is true and 0 otherwise, and $\mathcal{N}(\mathbf{x})$ is the 6-neighborhood of $\mathbf{x}$.

3.2 Optimization

When the objective function in (3.5) is minimized w.r.t. a single variable $\mathbf{w}$ or c, the problem amounts to registration or change detection, as described in Sections 3.1.1.1 and 3.1.1.2, for which it is possible to design efficient optimization methods. However, the joint problem is more difficult to handle. Since $\mathbf{w}$ takes continues values in $\mathbb{R}^3$, and c is a binary variable, they should be estimated with continuous and discrete optimization methods, respectively, which prevents from a common approach. Therefore, we separate the joint optimization in two subproblems associated to $\mathbf{w}$ and c in an alternating

scheme. At each iteration the optimization problem (3.5) is solved w.r.t. each variable while keeping the other fixed. Since the joint problem is non-convex, convergence to a global minimum cannot be guaranteed and can be influenced by initialization. We observe experimentally that assuming no changes at the first iteration ($c(\mathbf{x}) = 0$, $\forall \mathbf{x} \in \Omega$) provided an acceptable initialization that converged to satisfying solutions.

3.2.1 Minimization w.r.t. w

When c is fixed in (3.5), the deformation field is estimated by solving the problem (3.1). The data term defined in (3.6) is non-convex because of the non-linearity induced by the term $I_2(\mathbf{x} + \mathbf{w}(\mathbf{x}))$, which leads to particularly difficult optimization problems. To make the problem more tractable, we follow the standard practice and apply a Taylor development to $I_2(\mathbf{x} + \mathbf{w}(\mathbf{x}))$ in (3.6) to obtain the following convex data term:

$$\rho_l(I_1, I_2, \mathbf{w}, \mathbf{x}) = \frac{1}{\sigma^2} \left\| \boldsymbol{\nabla}^\top I_2(\mathbf{x})\, \mathbf{w}(\mathbf{x}) + I_t(\mathbf{x}) \right\|_2^2, \tag{3.9}$$

where $I_t(\mathbf{x}) = I_2(\mathbf{x}) - I_1(\mathbf{x})$ is the temporal derivative. By substituting ρ by ρ_l in (3.1), the optimisation problem becomes convex and can be addressed with a variety of efficient optimization methods.

We chose to consider the alternated direction method of multipliers (ADMM) [Boyd, 2010]. To this end, we introduce a splitting variable $\mathbf{z}$ that decouples the two terms of (3.1) that depend on $\mathbf{w}$, and we formulate the problem in the constrained form:

$$\min_{\mathbf{w}} \sum_{x \in \Omega} (1 - c(\mathbf{x}))\ \rho_l(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_1 \Psi(\mathbf{z}) \quad \text{s.t} \quad \mathbf{w} = \mathbf{z}. \tag{3.10}$$

The Augmented Lagrangian associated with (3.10) can be written in its scaled form

$$\mathcal{L}(\mathbf{w}, \mathbf{z}, \boldsymbol{\alpha}) = \sum_{x \in \Omega} (1 - c(\mathbf{x}))\ \rho_l(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_1 \Psi(\mathbf{z}) + \frac{\mu}{2} \left\| \mathbf{w} - \mathbf{z} + \frac{\boldsymbol{\alpha}}{\mu} \right\|_2^2 + \frac{1}{2\mu} \|\boldsymbol{\alpha}\|_2^2, \tag{3.11}$$

where $\boldsymbol{\alpha}$ is a Lagrange multiplier and $\mu > 0$. The ADMM algorithm is based on alternating the minimization of the Augmented Lagrangian w.r.t. $\mathbf{w}$ and $\mathbf{z}$, and a gradient ascent on the dual variable [Boyd, 2010]. It leads to the following iterative updates of $\mathbf{w}$ and $\mathbf{z}$ (see [Fortun et al., 2018] for a similar derivation with different data and regularization terms):

$$\mathbf{w}^{k+1} = \operatorname*{argmin}_{\mathbf{w}} \sum_{\mathbf{x} \in \Omega} (1 - c(\mathbf{x}))\ \rho_l(I_1, I_2, \mathbf{w}, \mathbf{x}) + \frac{\mu}{2} \left\| \mathbf{w} - \mathbf{z}^k + \frac{\boldsymbol{\alpha}^k}{\mu} \right\|_2^2 \tag{3.12}$$

$$\mathbf{z}^{k+1} = \operatorname*{argmin}_{\mathbf{z}} \lambda_1 \Psi(\mathbf{z}) + \frac{\mu}{2} \left\| \mathbf{w}^{k+1} - \mathbf{z} + \frac{\boldsymbol{\alpha}^k}{\mu} \right\|_2^2 \tag{3.13}$$

$$\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k + \mu(\mathbf{w}^{k+1} - \mathbf{z}^{k+1}) \tag{3.14}$$

In what follows, we show how we compute the updates of $\mathbf{w}$ (subproblem (3.12)) and $\mathbf{z}$ (subproblem (3.13)).

Update of w The problem (3.12) is voxel-wise and can be written for each $\mathbf{x} \in \Omega$:

$$\mathbf{w}^{k+1}(\mathbf{x}) = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1-c(\mathbf{x})}{\gamma^2}\|\boldsymbol{\nabla}^\top I_2(\mathbf{x})\,\mathbf{w}(\mathbf{x}) + I_t(\mathbf{x})\|^2 + \frac{\mu}{2}\left\|\mathbf{w}(\mathbf{x}) - \mathbf{z}^k(\mathbf{x}) + \frac{\boldsymbol{\alpha}^k(\mathbf{x})}{\mu}\right\|_2^2. \tag{3.15}$$

In regions where lesions are detected ($c(\mathbf{x}) = 1$), we have $\mathbf{w}^{k+1}(\mathbf{x}) = \mathbf{z}^k(\mathbf{x}) - \frac{\boldsymbol{\alpha}^k(\mathbf{x})}{\mu}$. When $c(\mathbf{x}) = 0$, the problem is convex and quadratic, and we have the following straightforward closed-form solution:

$$\mathbf{w}^{k+1}(\mathbf{x}) = \left(\frac{2}{\gamma^2}\boldsymbol{\nabla}^\top I_2(\mathbf{x})\boldsymbol{\nabla} I_2(\mathbf{x}) + \mu\mathbf{I}\right)^{-1}\left(\mu\left(\mathbf{z}^k(\mathbf{x}) - \frac{\boldsymbol{\alpha}^k(\mathbf{x})}{\mu}\right) - \frac{2}{\gamma^2}\boldsymbol{\nabla}^\top I_2(\mathbf{x})I_t(\mathbf{x})\right). \tag{3.16}$$

The matrix $(\boldsymbol{\nabla}^\top I_2(\mathbf{x})\boldsymbol{\nabla} I_2(\mathbf{x}) + \mu\mathbf{I})$ is 3×3 and is easily invertible in closed-form.

Update of z The problem (3.13) can be written in matrix form

$$\mathbf{z}^{k+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \frac{\lambda}{2}\|\mathbf{D}\mathbf{z}\|_2^2 + \frac{\mu}{2}\left\|\mathbf{w}^{k+1} - \mathbf{z} + \frac{\boldsymbol{\alpha}^k}{\mu}\right\|_2^2 \tag{3.17}$$

where $\mathbf{D} = (\mathbf{D}_x, \mathbf{D}_y, \mathbf{D}_z)^\top$ is a discrete differential operator concatenating first-order forward finite-difference matrices in the three spatial directions. Similarly with (3.16), the solution of (3.17) writes

$$\mathbf{z}^{k+1} = (\mathbf{D}^\top\mathbf{D} + \mu\mathbf{I})^{-1}(\mu\mathbf{w}^{k+1} + \boldsymbol{\alpha}^k) \tag{3.18}$$

The matrix $(\mathbf{D}^\top\mathbf{D} + \mu\mathbf{I})$ is too large to be inverted, but the update (3.2.1) can be be efficiently computed computed in the Fourier domain. Since $\mathbf{D}$ is a block-circulant matrix, it is diagonalizable by the discrete Fourier transform, which gives $\mathbf{D} = \mathbf{F}^\top\Lambda\mathbf{F}$, where $\mathbf{F}$ is the Fourier Transform matrix and Λ is a diagonal matrix with the Fourier coefficients of $\mathbf{D}$ in its diagonal entries. Using this property, we obtain after simple manipulations

$$\mathbf{z}^{k+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mu\mathbf{w}^{k+1} + \boldsymbol{\alpha}^k)}{\lambda|\mathcal{F}(\mathbf{D}^\top\mathbf{D})|^2 + \mu\mathbf{I})}\right) \tag{3.19}$$

The term $|\mathcal{F}(\mathbf{D}^\top\mathbf{D})|^2$ is constant and can be calculated offline, such that the computational cost of the update (3.19) essentially consists in the computation of two Fourier transforms.

The main advantage of the ADMM framework for the problem at hand is to be flexible enough to cope with different registration models with low computational cost. The requirement is to be able to design a splitting of the cost function such that the updates of $\mathbf{w}$ and c have computationally efficient solutions. Examples of admissible models comprise data terms based on the ℓ_1 penalty function or cross-correlation [Vogel et al., 2013], and regularizations by total variation or Nuclear norm of the Jacobian [Bostan et al., 2014].

The Taylor expansion applied to obtain the data term (3.9) is valid only for small deformations. To cope with large displacements, we repeat the ADMM optimization described above at each level of a coarse-to-fine pyramid, such that large displacements are capture at coarser levels, and they are iteratively refined at finer levels of the pyramid. This coarse-to-fine estimation is a common practice in registration and motion estimation [Fortun et al., 2016; Hill et al., 2001].

3.2.2 Minimization w.r.t. c

When $\mathbf{w}$ is fixed, the estimation of c amounts to a binary segmentation problem with Potts regularization:

$$\widehat{c} = \underset{c}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} \left[\lambda_2 - \ \rho(I_1, I_2, \mathbf{w}, \mathbf{x})\right] c(\mathbf{x}) + \lambda_3 \Phi(c). \tag{3.20}$$

This amounts to a Markov random field segmentation with binary labels, pairwise interactions and a submodular cost function. Under these conditions, the optimal solution can be computed efficiently with the graph cut optimization technique described in [Boykov et al., 2001].

Chapter 4

Evaluation of the proposed framework

This chapter outlines the results of the experiments conducted to assess the performances of the joint model for longitudinal lesion change detection. The first section describes the evaluation framework, including the datasets and the evaluation metrics. The second part discusses the convergence and stopping criteria of the model. Given that the joint model is governed by three hyperparameters, the procedure for selecting them is described in the third segment. The final section presents the results obtained with the joint model and compares them to those of the sequential with deformable registration and the sequential with affine registration schemes.

4.1 Evaluation framework

We evaluate our method on one synthetic dataset and two publicly available real patients datasets. The synthetic dataset offers the advantage to have an unambiguously defined ground truth change detection map, while controlling the amount of noise, bias field inhomogeneity and brain atrophy that corrupt the images. The real datasets are used to evaluate the proposed approach in conditions that are closer to the clinical routine, with different acquisition conditions and various pathological evolution. The first real patients dataset, denoted as *LesjakDB* [Lesjak et al., 2016], is dedicated to assess the ability of methods to detect every kinds of MS lesion evolutions (shrinkage, growth, new and disappearing), whereas the second dataset, denoted in the sequel as *MSSEG-2* [Commowick et al., 2021], only focuses on the ability to detect new appearing lesions.

4.1.1 Synthetic dataset

We use T2-weighted synthetic volumes generated with the Brainweb simulator [Cocosco et al., 1997]. We consider the *normal* anatomical model (*i.e.,* without lesion) and two multiple sclerosis anatomical models with *moderate* and *severe* lesion load. The images are simulated at a $1mm^3$ isotropic resolution (image size: 181x217x181) with bias field inhomogeneity (20%). To simulate realistic brain atrophy for the follow-up image, we applied a deformation field that has been estimated using a deformable registration [Avants et al., 2008] from two T1-weighted MRI scans acquired at four years apart of a

Table 4.1: Simulated longitudinal acquisitions

scenario	baseline	follow-up	simulated atrophy
lesion appearance without atrophy	normal	moderate	no
lesion growth without atrophy	moderate	severe	no
lesion appearance with atrophy	normal	moderate	yes
lesion growth with atrophy	moderate	severe	yes

patient suffering from MS that exhibits a significant brain atrophy evolution (in-house dataset). It should be noted that these real data images were first registered with an affine model onto the brainweb image to ensure the estimated deformable registration to be consistent with the underlying anatomy. The visual inspection confirms that the simulated image exhibits a realistic atrophy pattern. Gaussian additive noise was added with a standard deviation fixed at 5% of the mean intensity in the brightest tissue (cerebrospinal fluid in the T2-weighted simulation). We consider several scenarios of simulated longitudinal acquisition that are summarized in Table 4.1.

4.1.2 Real dataset LesjakDB: all kinds of lesion evolution

The LesjakDB dataset [Lesjak et al., 2016] is composed of 20 longitudinal MRI acquisitions of MS patients with two timepoints. The median time between the baseline and follow-up studies was 311 days, ranging from 81 to 723 days. Each MRI acquisition consists in a 2D T1-weighted, a 2D T2-weighted and 2D-FLAIR sequences. Change detection was conducted on the FLAIR images only. The FLAIR image size is $256 \times 256 \times 49$ with an anisotropic spatial resolution of $0.9 \times 0.9 \times 3$ mm. Ground truth change detection maps are also provided, which were obtained from manual annotations done by two expert raters. We adjusted some of the ground truth annotations that did not match the real lesion changes. The annotated changes include appearing, growing, shrinking and disappearing lesions. Ground-truth detection maps are compared to binary detection maps that include both positive and negative intensity changes.

4.1.3 Real dataset MSSEG-2: only appearing lesions

MSSEG-2 dataset [Commowick et al., 2021] is composed of 100 pairs of FLAIR MRI scans from MS patients acquired on various MR scanners. The provided ground-truth is limited to new appearing lesions, and was build from the consensus of manual annotations delineated by four experts. The dataset is separated into training (40 patients) and testing (60 patients) sets. Since the proposed approach does not require any training step, we consider the whole dataset for testing. However, we distinguish two subgroups of data, namely *MSSEG-2-Change*, which regroups the 61 subjects that exhibit at least one new appearing lesion and *MSSEG-2-NoChange*, which regroups the 39 subjects that do not exhibit any new appearing lesion. Since the provided ground-truth is limited to new appearing lesions, they are compared only to the positive binary change detection maps obtained with the different methods. Note, that the

proposed method framework does not discriminate appearing from evolving lesion. Consequently, lesion evolutions, which are not labeled in the ground truth detection maps, are erroneously considered as false positive detection, thus introducing a bias in some of the evaluation metrics.

4.1.4 Metrics

We report six metrics (four voxel-wise and two lesion-wise metrics) to evaluate the performance of the methods to detect changes, namely the Dice Similarity Coefficient (DSC), the Positive Predictive Value (PPV), the True Positive Ratio (TPR) and the *local* DSC. Let TP, TN, FP, FN be the number of voxels from estimated change detection map that correspond to *True Positive*, *True Negative*, *False Positive* and *False Negative*, respectively.
The DSC is defined as:

$$DSC = 2TP/(2TP + FP + FN)$$

and reflects the overall good overlap between the detection map and the ground truth.
The PPV is defined as:

$$PPV = TP/(TP + FP)$$

and reflects the proportion of relevant detections among all the detected changes.
The TPR is defined as:

$$TPR = TP/(TP + FN)$$

and reflects the proportion of the ground-truth changes that have been detected.
The *local* DSC correspond the DSC computed on a restricted area defined as the dilation with a 4-voxel radius spherical structuring element to the ground truth. This metric allows us to focus the evaluation on the local spatial accuracy of the detection method
In addition to the voxel-wise metrics, we also report lesion-wise metrics, namely the Lesion True Positive Ratio (L-TPR) and the Lesion Positive Predictive Value (L-PPV). These metrics have been evaluated thanks to the *animaSegPerfAnalyzer* validation tool while considering the same hyperparameters as in [Commowick et al., 2018]:
Since all these metrics are not relevant for data that do not exhibit any changes, we consider in that specific case the number of detected connected components as well as the volume of detected changes to characterize the false positive detections.

4.2 Convergence and stopping conditions of the optimization

The optimization method is composed of two nested iterative schemes: (i) the alternating updates of the deformation field $\mathbf{w}$ (3.9) and the change map c (3.20) and (ii) the iterative ADMM solution of the minimization problem (3.9) at each iteration of the alternating scheme. In this section, we analyze the convergence of these two parts of the algorithm.

Convergence of registration

As we detailed in Section 3.2.1, the minimization of the deformation field is calculated with the ADMM algorithm. The convergence of the algorithm is mainly determined by the parameter μ which is associated with the quadratic penalty in the Augmented Lagrangian. The parameter μ determines the rate at which the constraint $\mathbf{w} = \mathbf{z}$ is enforced during the optimization process. While not technically a step size, it has a similar effect on the convergence speed. We report here experiments to select illustrate the impact of μ. This experiment was conducted using the synthetic dataset described in Section 4.1.1. For a fixed value of the regularization parameter λ_1 that provides satisfactory registration, we provide the evolution of the following quantities:

- The energy as defined in (3.5)
- The registration error, defined as $\text{Err} = \|\mathbf{w}_{\text{ref}} - \mathbf{w}\|_2^2$, where $\mathbf{w}_{\text{ref}}$ is the reference deformation field that was used to transform the synthetic images.

As illustrated in Figure 4.1, higher values of μ lead to faster convergence. To evaluate the impact of μ on the registration quality, we conducted experiments using a synthetic BrainWeb image and the same image deformed with a known deformation field we will denote $\mathbf{w}_{ref}$. We measured the error between the estimated and known deformation fields for each iteration of ADMM. As shown in Figure 4.2, higher values of μ increases the error between the estimated and true deformation fields. We also visually inspected the deformation fields and observed that even for $\mu = 100$, the quality of registration remained satisfactory, as depicted in Figure 4.3. Therefore, we selected this value for the remainder of our experiments as it strikes a good balance between convergence speed and registration quality.

Stopping criteria of change detection The iterations of the alternated minimization of (3.5) and of the ADMM algorithm (3.12)-(3.14) are stopped when a stopping criterion is verified or when a maximum number of iterations is reached. The stopping criterion is a threshold on the norm of the relative changes between two consecutive iterations, and is set to 10^{-3} for the alternated minimization and 2.10^{-3} for ADMM. The maximum number of iterations is set to 5 for the alternated minimization and 300 for ADMM.

4.3 Setting the hyperparameters of the joint model

The proposed joint formulation in equation (3.5) requires setting of three hyperparameters:

- λ_1 controls the spatial regularization of the deformation field,
- λ_2 serves as a threshold for the intensity difference map,
- λ_3 governs the spatial regularization of the change map.

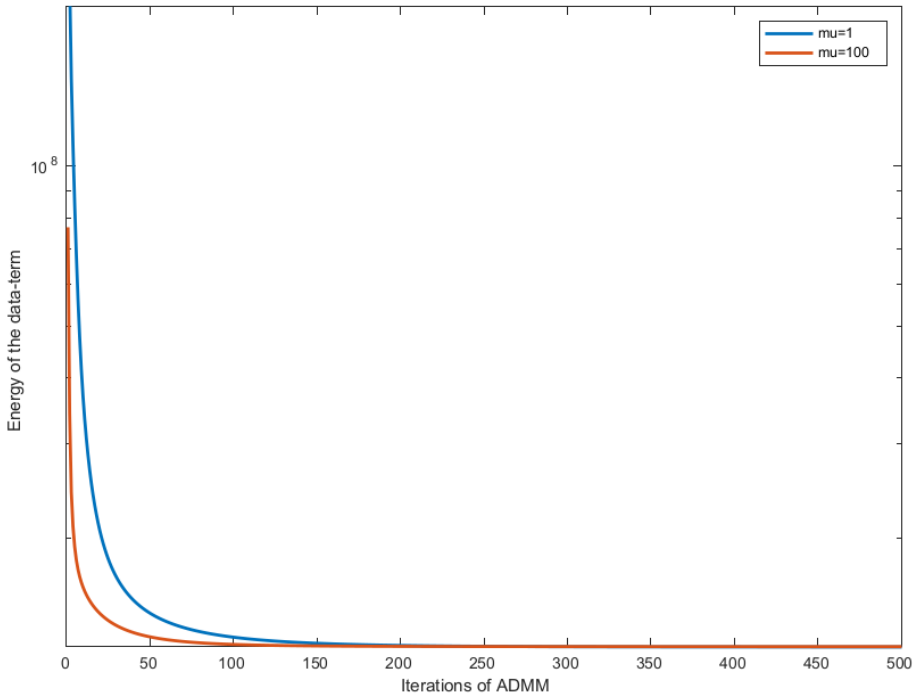


Figure 4.1: Energy of the minimization wrt. $\mathbf{w}$ along iterations of the ADMM algorithm

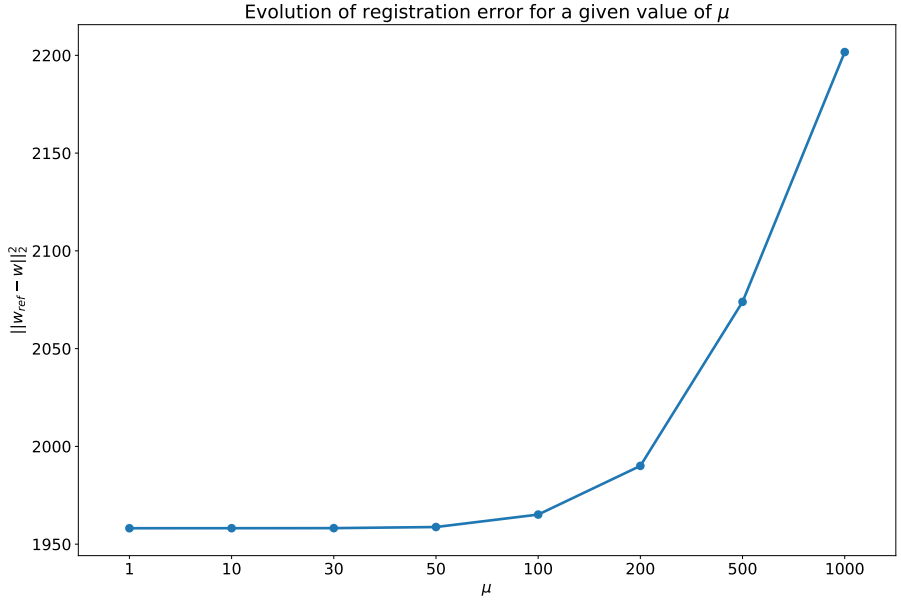


Figure 4.2: Evolution of registration error for different values of μ

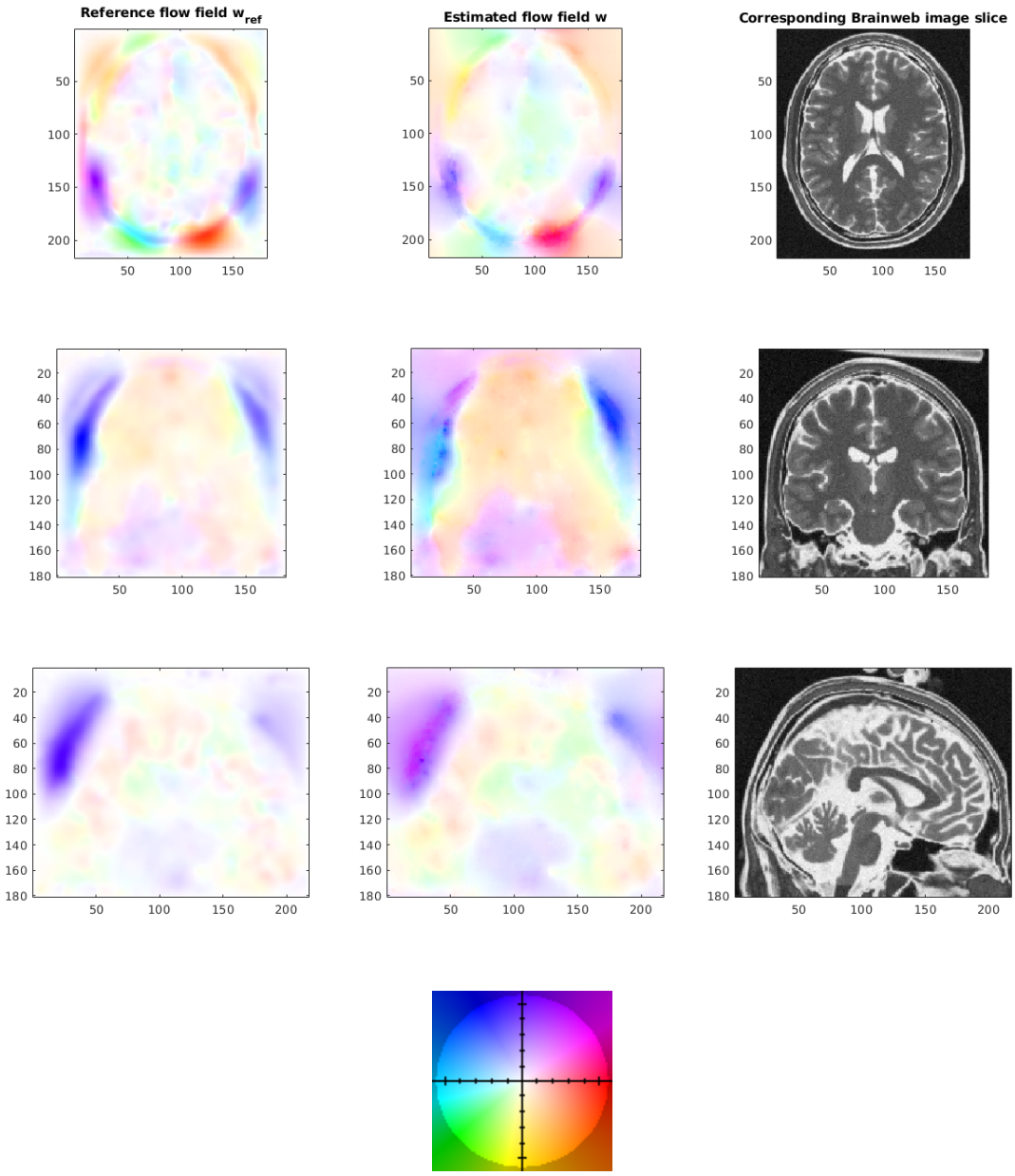


Figure 4.3: Visualization of the estimated deformation field alongside the actual deformation field

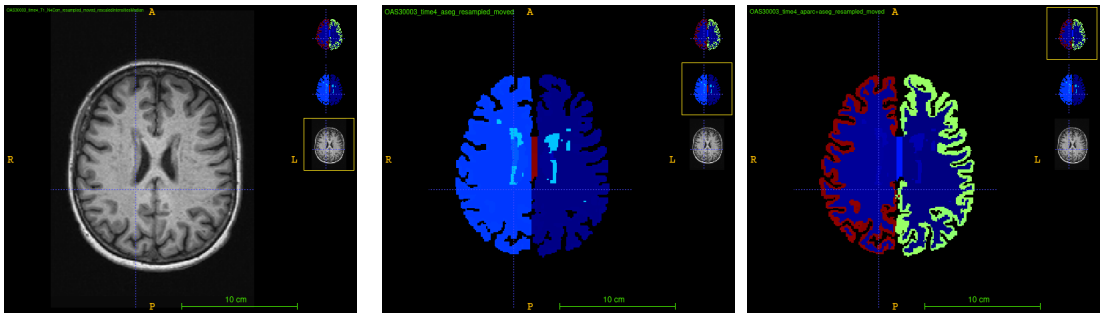
Tuning these hyperparameters in the absence of ground truth can be challenging, and letting the user setting empirically these hyperparameters can introduce a significant bias in the output results. To make the algorithm more useable in a clinical scenario, we have to find out a set of values that would work for most pairs of longitudinal MRI. We also propose a procedure to automatically set each of these hyperparameters in order to optimise performance for a given acquisition protocol, with the prerequisite of having a set of images with an associated ground truth. In Section 4.3.1, we demonstrate how the weight of the registration regularization, λ_1, can be determined independently from λ_2 and λ_3. Section 4.3.2 describes how the hyperparameters λ_2 and λ_3, related to change detection, can be selected to achieve a specified false positive rate.

4.3.1 Regularization of the registration term: λ_1

The role of λ_1 is to weigh the regularization term of the deformation field. In our approach, the goal of registration is to capture brain atrophy. Therefore, we designed an experiment to find a value of λ_1 that compensates typical amounts of atrophy that arise in MS without taking the presence of lesions into account. There is no available ground-truth to assess the registration quality. To circumvent this problem, we utilize a dataset of brain MRI scans collected over time that display changes in brain tissue volume similar to those seen in Multiple Sclerosis (MS), but without any lesions. The OASIS3 LaMontagne et al. [2019] dataset meets these criteria and includes scans of both healthy aging patients and patients with Alzheimer's disease. The scans are accompanied by expert-corrected automatic segmentation files with 44 subcortical or 112 subcortical and cortical labels (denoted as *aseg* and *aparc+aseg* respectively) depicted in Figure 4.4. We chose to use T1 images because of the poor quality of FLAIR volumes. The MRI images in this dataset vary in terms of visual quality and slice thickness, even within a single patient. To limit the anatomical changes in the brains of the selected patients to natural atrophy, we chose a subset of 21 healthy aging patients with 4 to 7 timepoints and no brain lesions.

We apply pre-processing composed of N4 bias field correction, Nyul's histogram matching and rigid registration to all the images. The estimated rigid transforms were used to warp segmentations onto the registered MRI scans, using nearest neighbor interpolation to maintain the label values. The registration part of the joint method (3.6) was applied to the selected pairs of longitudinal MRI scans, with varying values of λ_1. To evaluate the registration quality for each tested λ_1, we applied the deformation field to the Freesurfer segmentations and computed the mean Dice similarity coefficients (denoted as DSC_{mean}) between each labeled region of interest.

The results are shown in Figure 4.5, which plots the median DSC_{mean} for each tested value of λ_1. The DSC_{mean} obtained with the rigidly registered segmentations is also displayed (dashed lines). It is evident that higher values of λ_1 produce results that are similar to those obtained with rigid registration. Based on this metric, the optimal value for λ_1 is found to be between 40 and 70, depending on the considered brain parcellation type. We chose to use $\lambda_1 = 70$, as a higher value of λ_1 will lead to less lesion overcompensation in practice.



(a) (b) (c)

Figure 4.4: (a) MRI image of OASIS3 patient, (b) Subcortical parcellation segmentation (aseg), (c) Cortical and Subcortical parcellation segmentation (aparc+aseg)

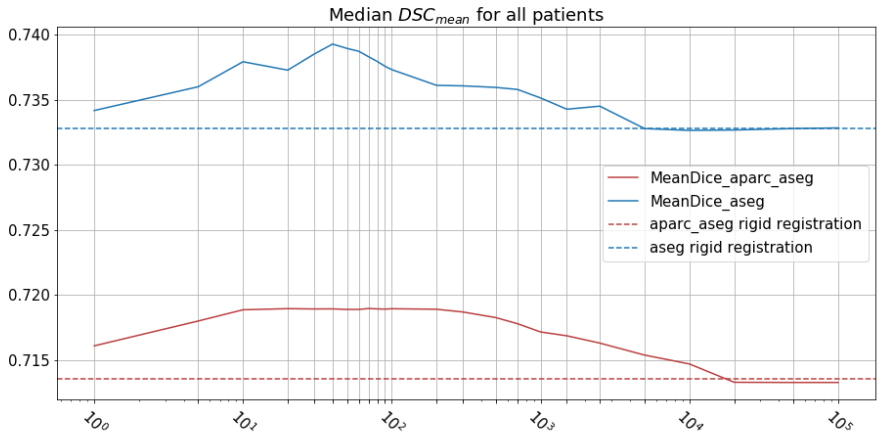


Figure 4.5: Median DSC_{mean} for the whole dataset. The dashed line represents the median DSC_{mean} of the rigidly registered Freesurfer segmentations.

4.3.2 Threshold λ_2 and spatial regularization of the lesion map λ_3

Once λ_1 is chosen, the values of λ_2 and λ_3 have to be set to find the best compromise regarding: (i) the expected intensity difference, (ii) the noise level that corrupts the images, and (iii) the spatial extent of the changes. Here, we suggest approaches to find out a range of valid settings that may work on most datasets.

Valid values for λ_2 The λ_2 value is directly linked to the false positive rate. Practitioners find the false positive rate to be a meaningful measure, so our approach is to allow the user to select a desired false positive rate and determine the corresponding λ_2 value. Our scheme involves testing the λ_2 threshold under the null hypothesis using longitudinal brain MRI with no changes, and displaying the associated false positive ratios (FPR). The Kirby21 dataset, which comprises 18 healthy patients with longitudinal acquisitions separated by a few hours, is used to achieve this. The dataset is prepared with N4 bias field correction, Nyul's histogram matching and rigid registration to all the images. The results are shown in figures 4.6 and 4.7.

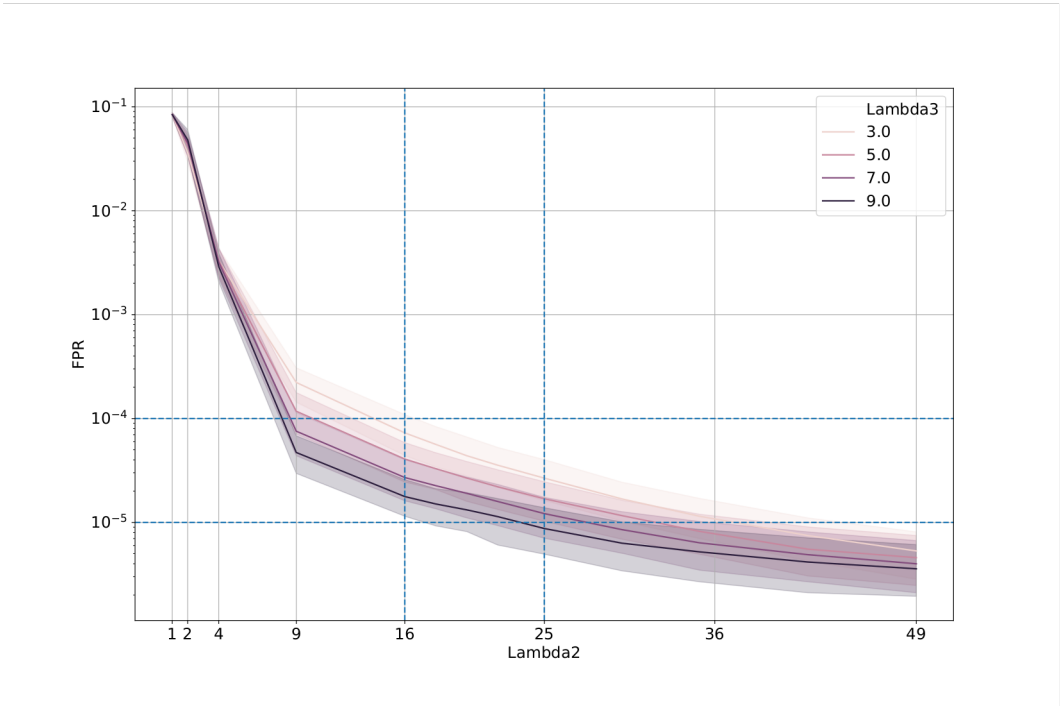
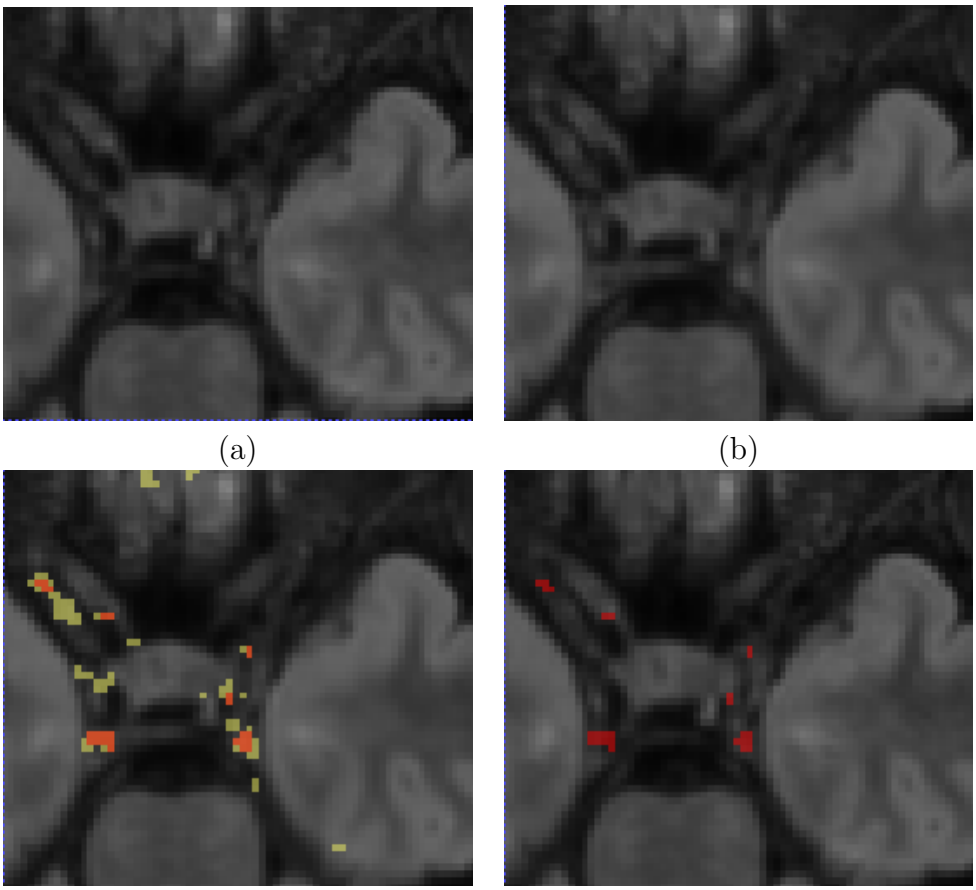


Figure 4.6: FPR vs. λ_2 for various values of λ_3

Figure 4.6 displays the false positive detection rate for various λ_2 values and horizontal lines are used to indicate false positive rates of FPR $= 10^{-4}$ and FPR $= 10^{-5}$. Vertical lines on the values $\lambda_2 = 16$ and $\lambda_2 = 25$ have been drawn to show that for any λ_3 considered, the range of false-positive ratios remain between FPR $= 10^{-4}$ and FPR $= 10^{-5}$. Figure 4.7 shows change maps for a particular patient at these two FPR values, where false positives were observed in a region of the brain. Although



(a) (b)

(c) (d)

Figure 4.7: Example change maps, λ_2 corresponding to $FPR = 10^{-4}$ et $FPR = 10^{-5}$. (a) First timepoint, (b) Registered second timepoint, (c) Change map, $FPR = 10^{-4}$(yellow), $\lambda_2 = 10$, (d) Change map, $FPR = 10^{-5}$(red), $\lambda_2 = 23$

FPR $= 10^{-5}$ can detect subtle changes, FPR $= 10^{-4}$ seems overly sensitive. The user may have to decide on an appropriate threshold.

Choosing λ_2 and λ_3 based on lesion coverage We made the final choices for λ_2 and λ_3 by taking their interdependence into account. If a low threshold is chosen, a higher value of λ_3 may be necessary to remove noisy detections by smoothing the change map. Conversely, if a higher threshold is chosen, lower values of λ_3 could be acceptable. Our goal was to select a combination of these two hyperparameters, with λ_1 fixed, that would maximize lesion coverage. To achieve this, we conducted a grid search experiment to determine the pairing of λ_2 and λ_3 that maximized the performance of the *affine sequential* approach in terms of local Dice Similarity Coefficient (local DSC, see Section 4.1.4) for each dataset. Considering the local DSC ensures that we focus on the ability of the detection scheme (Eq. 3.20) to recover the detected changes without being affected by false positive detections that may occur in other brain regions. By using the affine transformation model, we ensure that the registration step does not alter the geometry of evolving regions. We used two datasets of real longitudinal acquisitions with manual annotations denoted LesjakDB and MSSEG-2. (see 4.1.2 and sec:MSSEGdb for dataset descriptions.

This procedure leads us to find $\lambda_2 = 16$ and $\lambda_3 = 5$ as optimal setting for LesjakDB dataset and $\lambda_2 = 25$ and $\lambda_3 = 3$ for MSSEG-2 dataset.

4.4 Results

4.4.1 Variants used for comparison

To demonstrate the benefits of the proposed joint modeling, we consider three variants of the change detection framework:

- *joint*: the proposed joint change detection and registration method described in Section 3.1.1.3.

- *sequential*: The sequential counterpart of the proposed method, which successively performs deformable registration and change detection. For the two steps, we use the same model and optimization algorithms as in substeps of the *joint* approach described in Section 3.2.

- *affine*: The *sequential* approach where the deformable registration has been replaced by affine registration, which corresponds to the most common case. The affine registration was estimated using ANTs library [Avants et al., 2011][1] with default parameters and mutual information metric. Then, the thresholding and smoothing of the change map routine follows model (3.20) like in the sequential deformable model.

[1] https://github.com/ANTsX/ANTs

4.4.2 Synthetic dataset

First, a qualitative visual comparison of the three methods is provided in Fig.4.8 for the lesion appearance with atrophy scenario. The *affine* method succeeds to detect almost all the lesion areas, but it suffers from false positive detection around the ventricles due to brain atrophy. Both the *sequential* and the *joint* methods compensated for brain atrophy deformation since none of them exhibit false detections around the ventricles. However, the *sequential* method failed to detect the whole lesion areas due to the over-compensation of lesion changes. This limitation is overcome by the *joint* approach that succeeds to detect the entire lesion areas.

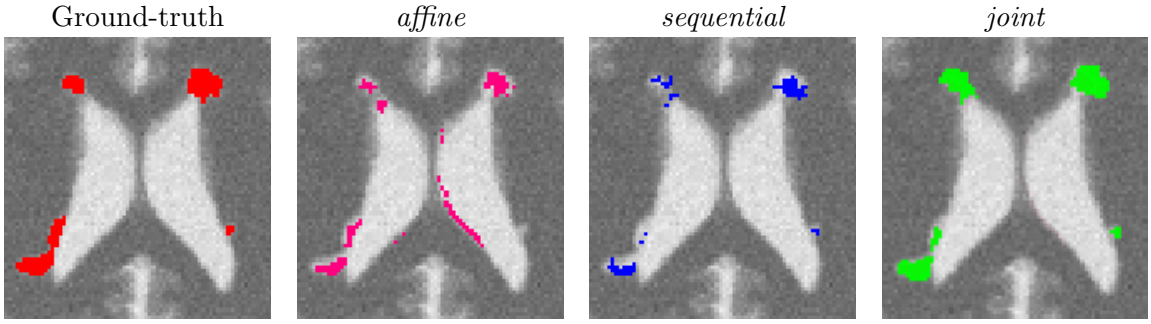


Figure 4.8: Qualitative comparison of the binary change detection maps obtained with the three methods on the synthetic dataset (lesion appearance with atrophy scenario)

A quantitative comparison of the three methods under four scenarios is provided in Table 4.2. First, we consider the appearance of lesion without atrophy. Unsurprisingly, this scenario is the most favorable for the *affine* method since there is no geometric difference to compensate. The *sequential* approach yields to significantly lower values of DSC and local DSC. This is due to the lesion over-compensation effect, as supported by the observed low TPR value (*i.e.,* lack of sensitivity) and high PPV value (*i.e* high specificity). Finally, the *joint* approach overcomes the shortcoming of the *sequential* approach and have performances similar to the *affine* method, with a slight tendency to underestimate the detected area. Similar observations can be made about the second scenario involving lesion growth without atrophy.

The conclusions are drastically different for the two scenarios involving simulated atrophy. The performance of the *affine* method is significantly hampered by the numerous false detections due to the atrophy. This is illustrated by the significant decrease of the DSC and PPV values compared to the cases without atrophy, while the TPR and local DSC values are less modified. The *sequential* approach succeeds to compensate for the simulated brain atrophy, as highlighted by the high PPV value, but still underestimates the changes to detect, as indicated by the low TPR value. The *joint* approach clearly outperforms the two previous approaches in terms of detection accuracy, as objectified by the significantly higher DSC value.

The behavior of the *joint* approach can be monitored through the iterations of the alternating optimization scheme and is illustrated in Fig. 4.9. We can see that the DSC increases across the iterations, and the convergence is reached in a few iterations.

Table 4.2: Results computed on the synthetic dataset.

Scenario	Method	DSC	PPV	TPR	local DSC
Lesion appearance no atrophy	*affine*	**0.830**	0.782	**0.885**	**0.830**
	sequential	0.684	**0.921**	0.544	0.684
	joint	0.814	0.887	0.751	0.814
Lesion growth no atrophy	*affine*	0.766	0.635	**0.964**	0.770
	sequential	0.685	**0.734**	0.641	0.688
	joint	**0.806**	0.726	0.902	**0.808**
Lesion appearance simulated atrophy	*affine*	0.460	0.329	**0.767**	**0.810**
	sequential	0.626	**0.960**	0.465	0.627
	joint	**0.743**	0.925	0.621	0.744
Lesion growth simulated atrophy	*affine*	0.652	0.505	**0.919**	0.827
	sequential	0.753	**0.869**	0.664	0.754
	joint	**0.847**	0.833	0.861	**0.848**

Concerning the computational cost of the joint approach, it is about 24min on one single core (Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz) for an experiment on the synthetic dataset (image size: 181x217x181).

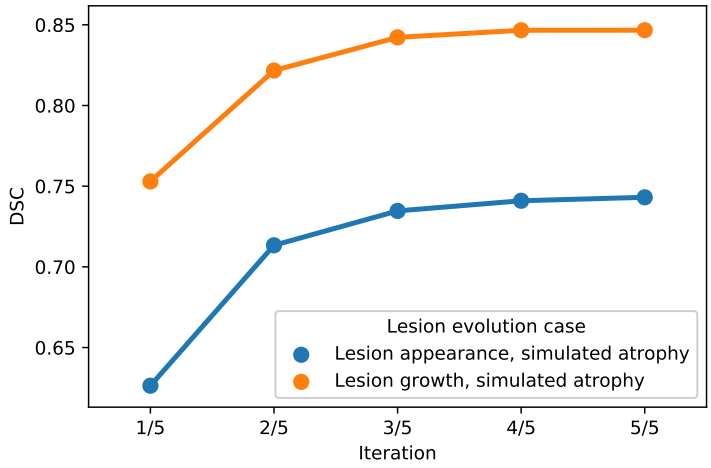


Figure 4.9: Evolution of DSC across iterations of the alternating optimization scheme of the *joint* approach on the synthetic dataset (blue: Lesion appearance, simulated atrophy, orange: Lesion growth, simulated atrophy).

4.4.3 LesjakDB

First, a qualitative visual comparison of the three methods is provided in Fig.4.10. Fig. 4.8 shows that we can draw similar conclusions as for the synthetic dataset. The *affine*

method demonstrates a high sensitivity (*i.e.,* the lesion evolution is well detected) but a lack of specificity (*i.e.,* numerous false positive detections are detected around the ventricles and in the posterior part of the cortex). Conversely, the *sequential* method has high specificity but lacks sensibility. The *joint* approach provides the best visual results, which illustrates its ability to achieve both high sensitivity and high specificity. Figures 4.10(c) and 4.10(d) show the Jacobian of the deformation fields obtained by the *sequential* and *joint* methods, respectively. The specific pattern characterized by the alternance of both high and low values of the Jacobian (see areas highlighted by the red squares in Fig. 4.10(c)) reflects the high local contraction and dilation induced by the deformation field to to make the lesion disappear, thus explaining the lack of sensitivity of the detection results.

The quantitative evaluation shown in Table 4.3 and 4.4 (first rows) and in Fig. 4.11 confirms the conclusions of the visual analysis. The high sensitivity of the *affine* method is objectified at the voxel level by a statistically significantly higher TPR than the two other methods. At the lesion level, all the three methods exhibit similar L-TPR values, thus emphasizing their ability to detect the same amount of changing areas. Both the *sequential* and *joint* methods yield significantly higher PPV and L-PPV compared to the *affine* method, which illustrates their ability to reduce the number of false detections induced by brain atrophy at both voxel and lesion levels. This result highlights the benefit of using deformable registration in the context of MS lesion monitoring. The significantly lower TPR achieved by the *sequential* method compared to the *joint* method is the consequence of the lesion over-compensation effect. Finally, the *joint* approach significantly outperforms the two other approaches in term of voxel-wise global accuracy (see DSC).

Fig. 4.12 highlights the variability of the performance of the methods across the subjects. It is interesting to notice that, although the performance of the detection methods greatly varies from one subject to the other, the ranking among the three methods appears to be highly consistent across the subjects. When investigating for the factors that may explain the observed variability, it appears that the volume of the ground-truth seems to play a prominent role: the larger is the volume to detect, the better is the performance of the change detection algorithm, as displayed in Fig. 4.13.

4.4.4 MSSEG-2

Similarly as for the synthetic and LesjakDB datasets, the qualitative visual comparison of the two approaches based on deformable registration in Fig. 4.14 highlights the lack of sensitivity of the *sequential* method due the lesion over-compensation effect. The resulting change detection map (purple) is too small compared to the ground-truth (underlying transparent red) due to the deformable registration that significantly shrinks the lesion. With the *joint* approach, the shape of the lesion is almost preserved in the warped follow-up image and the change detection map (green) matches almost perfectly the ground-truth.

The quantitative evaluation on the subset MSSEG-2-Change is reported in the second row of Tables 4.3 and 4.4 and in the upper part of Fig. 4.15.

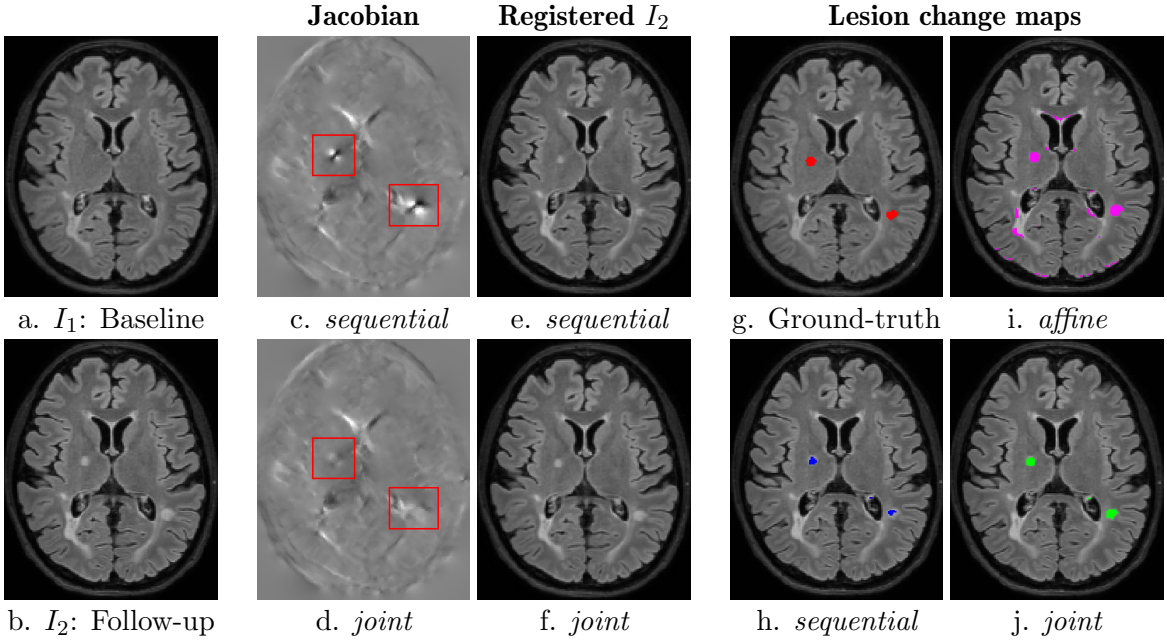


Figure 4.10: Qualitative comparison of the binary change detection maps obtained with the three methods and of the jacobian of the deformation field estimated with the *sequential* and *joint* approaches on one selected subject from the MSSEG-2 dataset. Hyperparameters: $\lambda_1 = 70$ (for *sequential* and *joint* methods), $\lambda_2 = 25, \lambda_3 = 3$

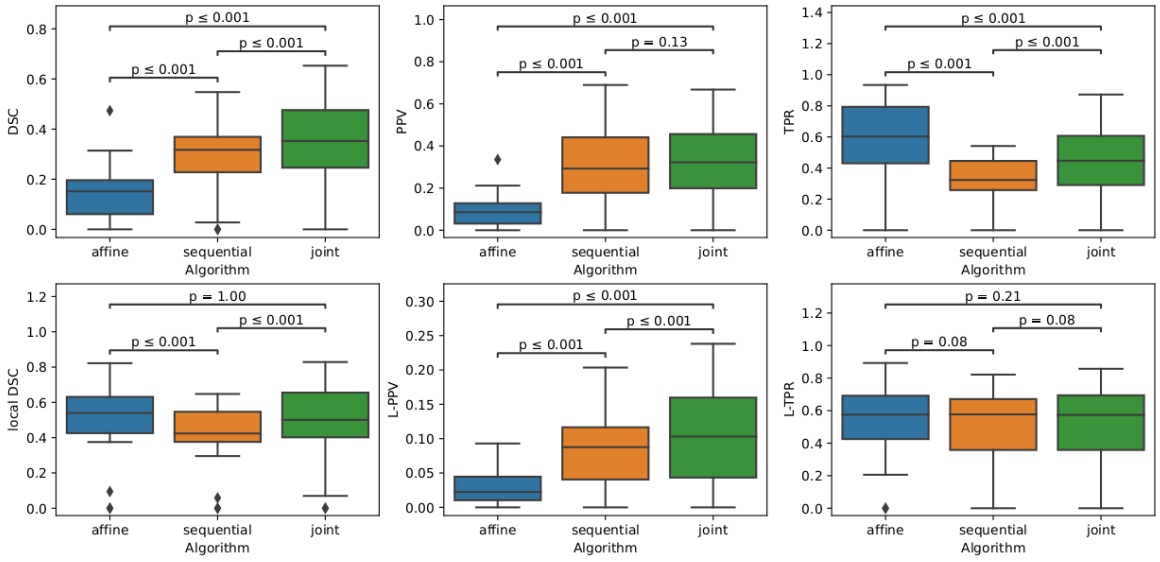


Figure 4.11: Boxplots corresponding to the results summarized in Tables 4.3 and 4.4 for LesjakDB ($N_{Subject} = 20$). Statistical significancy is evaluated thanks to the Wilcoxon signed-rank test between each pair of methods while applying Benjamini/Hochberg FDR correction.

Table 4.3: Results computed on LesjakDB and MSSEG-2-Change datasets (Voxel-wise metrics). The median ± the median absolute deviation (MAD) computed over all subjects are reported for each metric. The MSSEG-2-Change inverse experiment consist in swapping the baseline and follow-up images to evaluate the ability of the methods to detect disappearing lesions.

DataSet	Method	local DSC	DSC	PPV	TPR
LesjakDB	*affine*	$\mathbf{0.539 \pm 0.174}$	0.152 ± 0.087	0.086 ± 0.060	$\mathbf{0.603 \pm 0.223}$
	sequential	0.424 ± 0.139	0.317 ± 0.117	0.293 ± 0.152	0.323 ± 0.125
	joint	0.501 ± 0.179	$\mathbf{0.353 \pm 0.144}$	$\mathbf{0.323 \pm 0.148}$	0.447 ± 0.207
MSSEG-2-Change	*affine*	$\mathbf{0.626 \pm 0.224}$	0.142 ± 0.165	0.081 ± 0.139	$\mathbf{0.633 \pm 0.269}$
	sequential	0.520 ± 0.196	0.310 ± 0.178	0.298 ± 0.264	0.379 ± 0.176
	joint	0.579 ± 0.219	$\mathbf{0.356 \pm 0.208}$	$\mathbf{0.336 \pm 0.254}$	0.474 ± 0.222
MSSEG-2-Change inverse	*affine*	0.626 ± 0.224	0.142 ± 0.165	0.081 ± 0.139	$\mathbf{0.633 \pm 0.269}$
	sequential	0.625 ± 0.217	0.348 ± 0.216	0.290 ± 0.244	0.550 ± 0.243
	joint	$\mathbf{0.655 \pm 0.237}$	$\mathbf{0.378 \pm 0.233}$	$\mathbf{0.312 \pm 0.250}$	0.619 ± 0.266

Table 4.4: Results computed on LesjakDB and MSSEG-2-Change datasets (Lesion-wise metrics) The median ± the median absolute deviation (MAD) computed over all subjects are reported for each metric. The MSSEG-2-Change inverse experiment consist in swapping the baseline and follow-up images to evaluate the ability of the methods to detect disappearing lesions.

DataSet	Method	L-PPV	L-TPR
LesjakDB	*affine*	0.022 ± 0.022	0.576 ± 0.195
	sequential	0.088 ± 0.051	$\mathbf{0.577 \pm 0.197}$
	joint	$\mathbf{0.103 \pm 0.061}$	0.574 ± 0.208
MSSEG-2-Change	*affine*	0.015 ± 0.042	0.840 ± 0.271
	sequential	0.095 ± 0.150	$\mathbf{0.872 \pm 0.307}$
	joint	$\mathbf{0.111 \pm 0.155}$	$\mathbf{0.872 \pm 0.304}$
MSSEG-2-Change inverse	*affine*	0.015 ± 0.042	0.840 ± 0.271
	sequential	$\mathbf{0.094 \pm 0.151}$	$\mathbf{0.977 \pm 0.292}$
	joint	0.091 ± 0.164	0.947 ± 0.304

The fact that both the *sequential* and *joint* approaches lead to significantly higher PPV values as compared to the *affine* approach advocates the use of deformable registration to reduce the number of false detections. The benefit of considering the *joint* over the *sequential* approach to overcome the lesion overcompensation effect is clearly demonstrated by the significantly higher TPR and local DSC values obtained with the *joint* method.

It is also interesting to note that the lesion over-compensation effect does not affect the special case of disappearing lesion. Indeed, when registering an image without lesion on a image with a lesion, the dissimilarity in the area of the disappearing lesion cannot be corrected by the transport of intensity of the registration (this is in fact only the case for non symmetric image registration method, see [Noblet et al., 2004] for further explanations). To illustrate this phenomenon, we consider the MSSEG-2-Change inverse experiment (see the third row of both Tables 4.3 and 4.4 and the bottom part

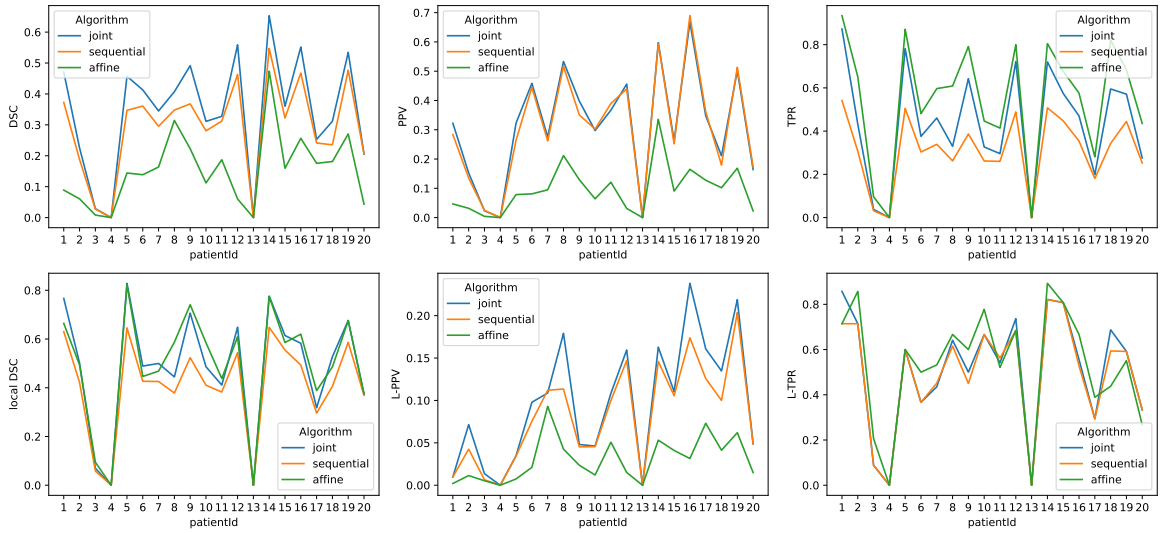


Figure 4.12: Metrics reporting the performance of the three methods for each subject of LesjakDB.

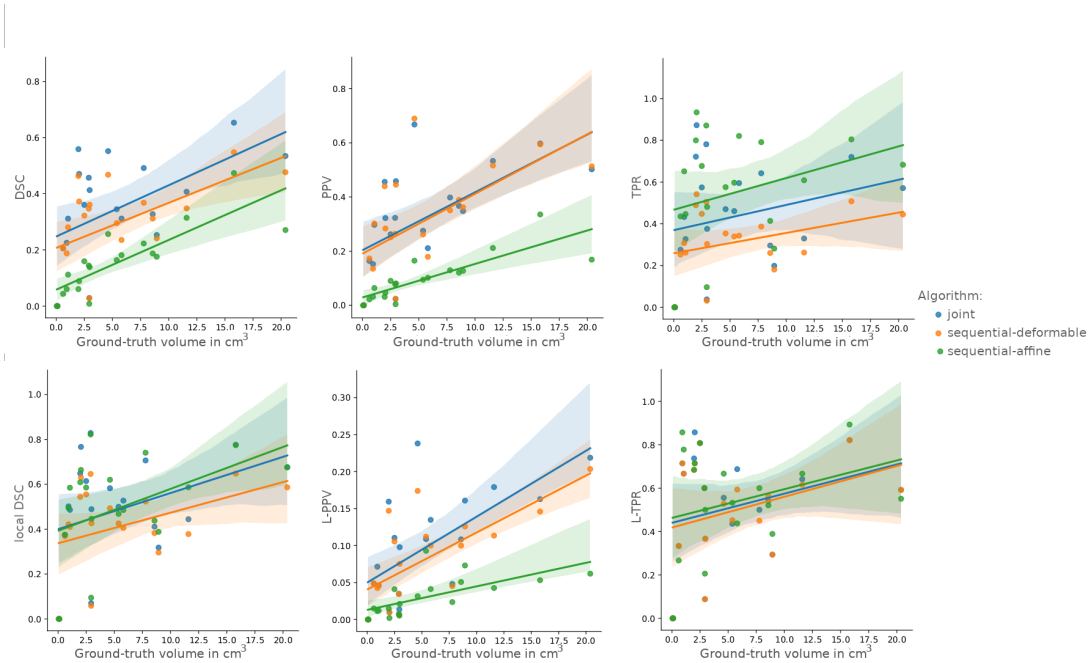


Figure 4.13: Correlation between the performance of the three methods and the ground-truth evolving lesion load (cm^3) of the subjects of LesjakDB.

of Fig. 4.15) that consists in swapping the baseline and the follow-up image, so that the ground-truth now corresponds to disappearing lesions. The same conclusion can be drawn from the DSC, PPV and TPR as compared to the MSSEG-2-Change experiment. The most interesting point concerns the local DSC that focuses the evaluation on the disappearing lesion. In that case, there is no significant difference any more between

sequential and *joint* approaches contrary to the MSSEG-2-Change experiment, showing the absence of lesion over-compensation effect in the specific scenario of detecting disappearing lesions.

Note that all the results presented above in this section are evaluated on the 61 subjects of MSSEG-2-Change (*i.e.,* subject presenting at least one new appearing lesion). Indeed, the presented metrics cannot be computed anymore for the 39 subjects of MSSEG-2-NoChange since the ground-truth change detection map is empty. This is why we only report the volume of detected changes for this subset of MSSEG-2 in Fig. 4.16. We can notice that both *sequential* and *joint* approaches lead to significantly lower volume of detected changes as compared to the *affine*, which appears in line with previous findings that support the use of deformable registration to reduce the number of false detections. Also note that the *joint* method yields consistently to slightly higher volume of detected changes as compared to the *sequential* method. This is also the consequence of the lesion over-compensation effect that affects the *sequential* appproach.

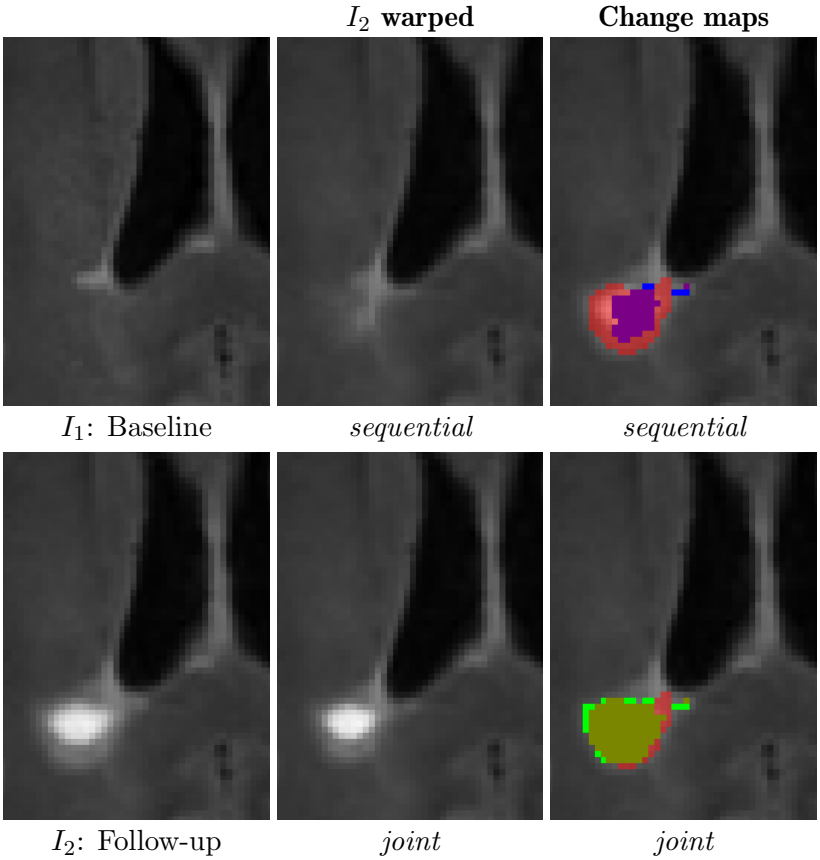
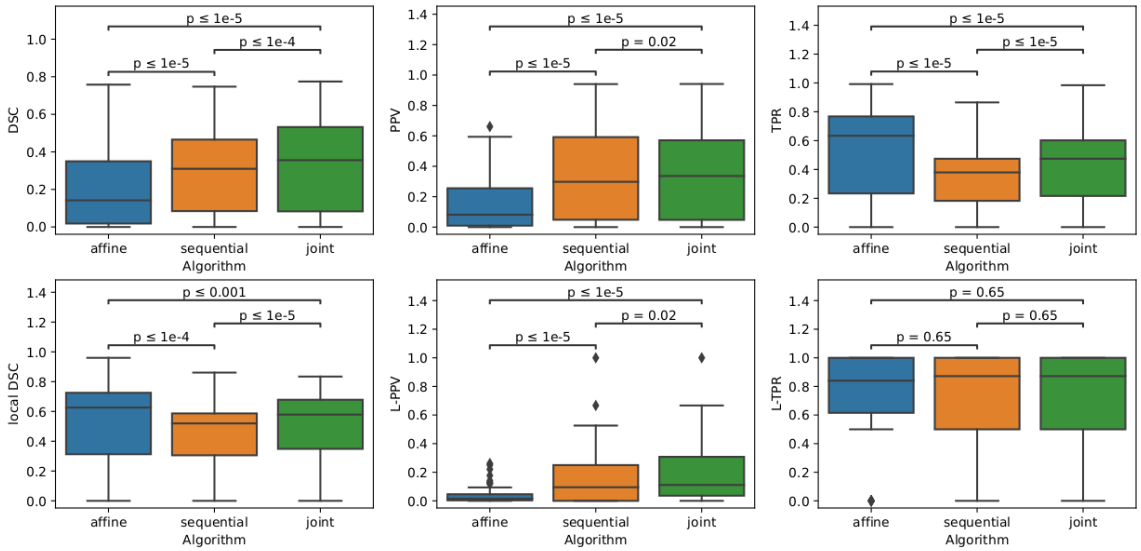
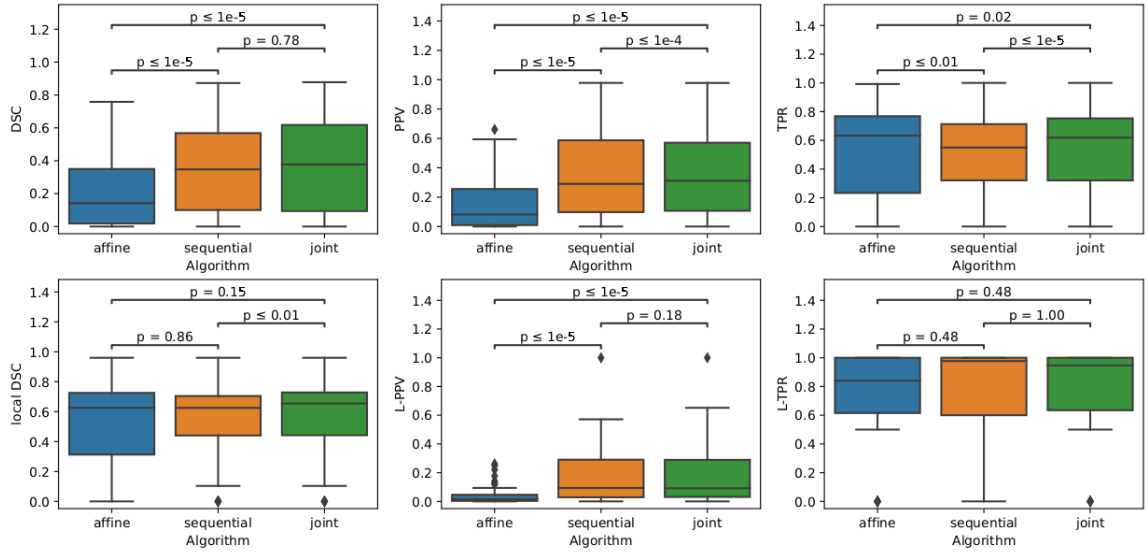


Figure 4.14: Qualitative comparison of the binary change detection maps obtained with the *sequential* (purple) and *joint* (green) approaches as compared to the ground truth (underlying transparent red) on one selected subject from the MSSEG-2 dataset.



MSSEG-2-Change



MSSEG-2-Change inverse

Figure 4.15: Boxplots corresponding to the results summarized in Tables 4.3 and and 4.4 for both MSSEG-2-Change and MSSEG-2-Change inverse ($N_{Subject} = 61$). Statistical significancy is evaluated thanks to the Wilcoxon signed-rank test between each pair of methods while applying Benjamini/Hochberg FDR correction.

4.5 Discussion

There is a high inter-subject variability in the metrics which is a reflection of the high variability of the cases considered, especially in terms of lesion loads and amount of atrophy. In MSSEG-2, 50% of the patients with changes have a total lesion load below

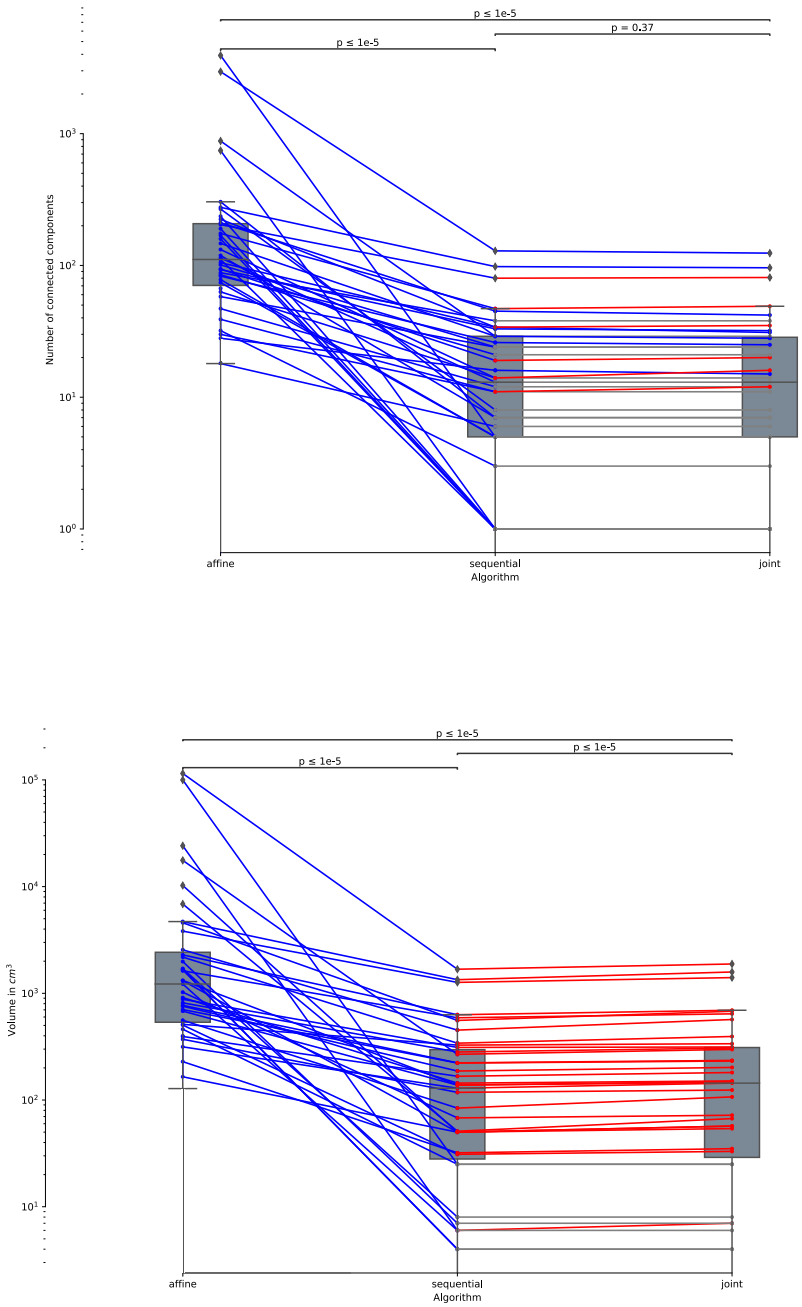


Figure 4.16: Number of connected components (left) and volume in cm^3 (right) of the detected changes computed for the 39 subjects of MSSEG-2-NoChange datasets (*i.e.*, subjects without new appearing lesion). Statistical significancy is evaluated thanks to the Wilcoxon signed-rank test between each pair of methods while appling Benjamini/Hochberg FDR correction.

200mm^3, and 25% of them below 100mm^3. These small lesion loads can be challenging to detect and metrics such as the DSC are sensitive to the ratio of detections overlapping and outside the ground-truth, which is more likely to occur if the lesions are small. In figure 4.13, we show that for small lesion loads, the amount of error is higher. The variability of the results can also be explained by the different acquisition conditions; inter-tissue contrast and noise levels are highly dependent on the acquisition parameters.

Influence of the hyperparameters on the detection results In the results, we affirm that the choice of hyperparameters, provided that they are in a certain range of validity, does not affect the ranking of the results or the qualitative properties of the change maps. Figure 4.17 illustrates this statement.

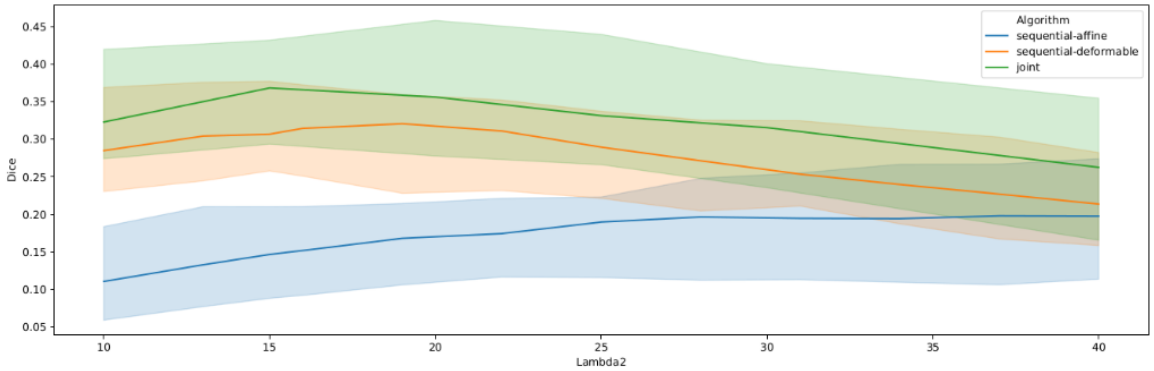


Figure 4.17: Influence of the threshold on the Dice metric on the Lesjak dataset

Availability and diversity of data and ground-truth Comparison with other approaches can be challenging because all research groups use datasets of different size and quality, and the manual annotation protocols is likely not normalized. Therefore, it is difficult to assess the generalizability of each approach. Deep-learning methods are only starting to emerge in the MS lesion change detection field because of this issue, but the recent MICCAI2021 chellenge for new lesion detection has improved the situation by proposing a publicly available dataset of 100 patients. Moreover, obtaining annotated data is very costly in time and skill and the datasets are often relatively small. The fact that the ground-truths are binary images is also a limitation since the delineation of lesion is not always clear.

Chapter 5

Variants of the model and perspectives

In Chapter 3, we described a generic joint estimation framework based on the minimization of a unified cost function (3.5). In Chapter 4, we presented results obtained with simple modeling assumptions detailed in Section 3.1.2. The versatility of our optimization approach allows us to consider more sophisticated models. In this chapter, we explore several variations from the basic form of our method presented in Chapter 3. All the models introduced in this chapter have been implemented during the thesis, and preliminary experiments have been conducted. However, further validation is required to conclude on the potential added value of these presented approaches.

5.1 Variants of the data-term

In this section, we present alternative versions of the basic data term (3.6), which are designed to address specific limitations detailed in the following subsections.

5.1.1 ℓ_1 data term

The data term in (3.6) penalizes deviations from intensity conservation with a ℓ_2 norm, based on the assumption that corresponding points in the two images belong to the same tissue, and consequently have the same intensity value. However, this assumption may be violated in the presence of outliers produced by acquisition artefacts that could not be properly corrected during preprocessing. More importantly, evolving lesions also create situations of corresponding points representing different types of tissue. The ℓ_2 norm is very sensitive to these violations of the intensity conservation assumption, and this leads to the overcompensation of lesions during registration described in Section 2.4.4. A possible way to prevent this effect is to replace the ℓ_2 norm with a more robust function that is able to tolerate deviations from the intensity conservation constraint locally. We choose to use a ℓ_1 norm, which leads to the following data term after linearization:

$$\rho_l(I_1, I_2, \mathbf{w}, \mathbf{x}) = \frac{1}{\sigma^2} |\boldsymbol{\nabla}^\top I_2(\mathbf{x})\, \mathbf{w}(\mathbf{x}) + I_t(\mathbf{x})|. \tag{5.1}$$

With this new data term, it is still possible to find a closed-form solution for the subproblem dedicated to the data term in the ADMM updates (3.12), with the same

computational cost as for the ℓ_2 case. Since the ℓ_1 norm is not differentiable, the three cases $p(\mathbf{w}) > 0$, $p(\mathbf{w}) < 0$ and $p(\mathbf{w}) = 0$, where $p(\mathbf{w}) = \boldsymbol{\nabla}^\top I_2(\mathbf{x})\,\mathbf{w}(\mathbf{x}) + I_t(\mathbf{x})$, should be distinguished, and the final solution writes (see Zach et al. [2007] for a similar case)

$$\mathbf{w}^{k+1} = \mathbf{r} + \begin{cases} \frac{\nabla I_2}{\mu} & p(\mathbf{r}) < -\frac{|\nabla I_2|^2}{\mu} \\ -\frac{\nabla I_2}{\mu} & p(\mathbf{r}) > \frac{|\nabla I_2|^2}{\mu} \\ -p(\mathbf{r})\frac{\nabla I_2}{|\nabla I_2|^2} & |p(\mathbf{r})| \leq \frac{|\nabla I_2|^2}{\mu} \end{cases} \tag{5.2}$$

where $\mathbf{r} = \mathbf{z} - \frac{\boldsymbol{\alpha}}{\mu}$, and we dropped the dependency on $\mathbf{x}$ for the sake of clarity.

5.1.2 Patch-based data-term

The data term (3.6) is evaluated voxel-wise, which makes it is sensitive to noise and is susceptible to generate false positive detections. To circumvent this limitation, spatial context can be introduced. We achieve this by applying a Gaussian kernel, following Bruhn et al. [2005]:

$$\rho(I_1, I_2, \mathbf{w}, \mathbf{x}) = \frac{1}{\sigma^2} \sum_{y \in \Omega} h_\sigma(\mathbf{x} - \mathbf{y})\ \|I_2(\mathbf{y} - \mathbf{w}(\mathbf{x})) - I_1(\mathbf{y})\|_2^2 \tag{5.3}$$

where $h_\sigma(\cdot)$ is a Gaussian kernel with standard-deviation σ. The filtering of the data term by $h_\sigma(\cdot)$ improves the robustness to noise, both for registration and thresholding of the subtraction image.

5.1.3 Symmetric data-term

The data term defined in (3.6) gives asymetric roles to the two images I_1 and I_2. The image I_2 is arbitrarily chosen as the moving image that is warped on the fixed reference I_1 with the estimated transformation field. The consequence of this asymetry on the registration of longitudinal MS images is that the overcompensation problem discussed in Section 2.4.4 occurs only in one direction and not the other. As illustrated in 5.1, when a lesion appears in I_2, the estimated deformation tends to eliminate it to favor similarity between I_1 and the registered I_2. The asymetry also has an impact on our joint estimation method described in Chapter 3: if a lesion is completely removed by the registration at the first iteration of the alternating minimization scheme described in Section 3.2, the deformation fields estimated at the following iterations will not be able to retrieve the lesion.

To overcome this limitation, we have investigated a symmetric registration approach. We follow the idea of [Noblet et al., 2012] and consider a deformation field that simultaneously warps both images in opposite directions to create a "halfway" solution. This is achieved with the following data term:

$$\rho(I_1, I_2, \mathbf{w}, \mathbf{x}) = \sum_{\mathbf{y} \in \Omega} h_\sigma(\mathbf{x} - \mathbf{y})\ \|I_2(\mathbf{y} + \mathbf{w}(\mathbf{x})) - I_1(\mathbf{y} - \mathbf{w}(\mathbf{x}))\|_2^2. \tag{5.4}$$

While in (3.6) the deformation field $\mathbf{w}$ represents the warping of I_2 on I_1, in (5.4) it represents the warping of I_1 and I_2 on an intermediate timepoint halfway between the two images. If a lesion appears in I_2, it is also present (with half of its size) in this intermediate timepoint. Therefore, the symmetric data term (5.4) reduces the overcompensation phenomenon to prevent from lesion elimination. Moreover, the behavior of the change detection method becomes the same for appearing and disappearing lesions.

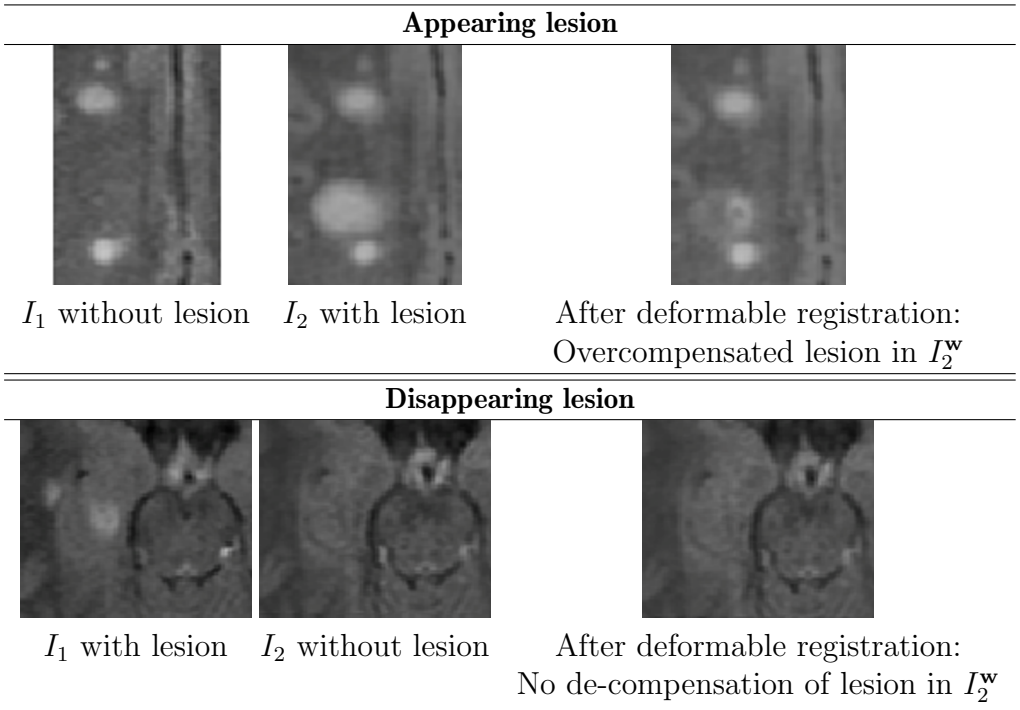


Figure 5.1: Illustration of different behaviors with the asymetric registration scheme.

Figure 5.2 illustrates the difference of registration results obtained with the symmetric and asymmetric approaches. The symmetric approach outperforms the asymmetric method in terms of lesion recovery in case of appearing lesion. In the asymmetric approach, the lesion is initially overcompensated, and although it is recovered by the end of the iterations, it remains smaller than the original lesion. Conversely, the overcompensation phenomenon is limited in the symmetric approach due to the smaller overall displacement of the second image. However, the symmetric model as we implemented it has limitations in detecting lesion evolution when lesions are present in both images since they register onto each other, leading to complete compensation of the lesion in the first iteration. Moreover, the symmetric approach incurs significantly higher processing time and memory costs compared to the asymmetric method. Nevertheless, this variant of registration should be further investigated with different data-terms or hyperparameter combinations that could alleviate both the processing time (μ for convergence speed) and the tradeoff between lesion overcompensation and estimation of atrophy (λ_1).

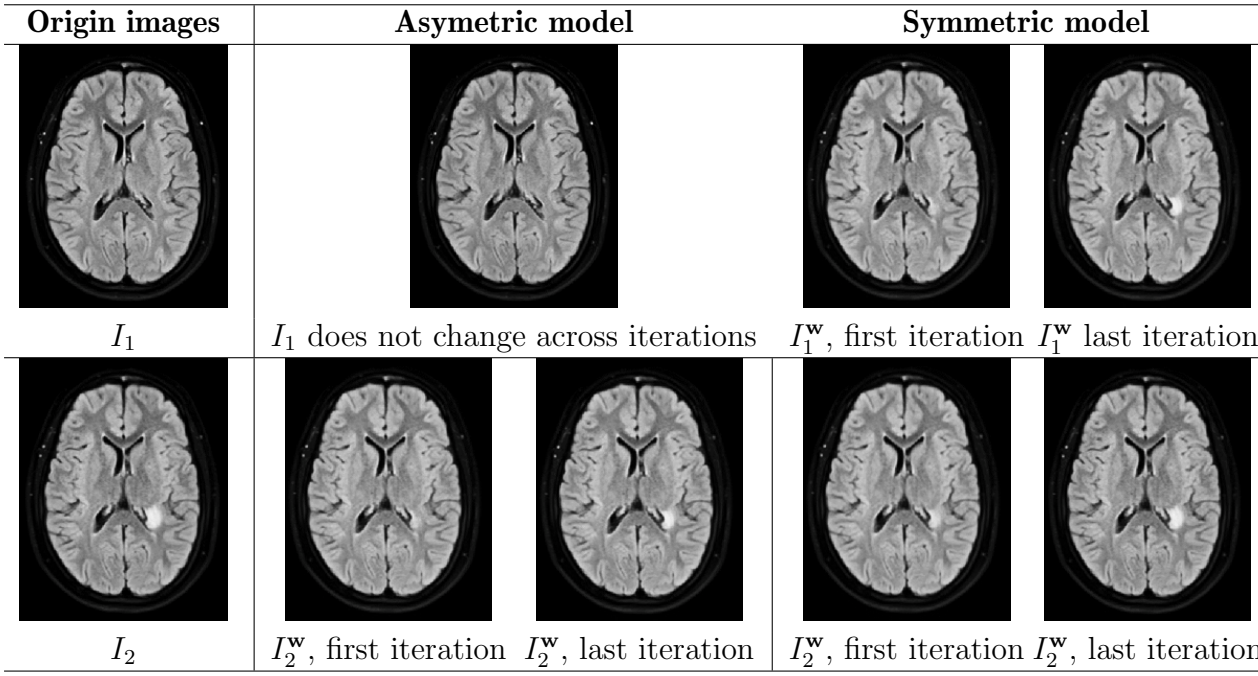


Figure 5.2: Illustration of the symmetric approach.

5.1.4 Multimodal version of the data-term

In MS studies, the FLAIR modality is typically regarded as providing the best tissue-to-lesion contrast, despite its inter-tissue contrast being slightly lower than in T1-w sequences. To take advantage of all MRI sequences, MS lesion change detection methods often rely on information from multiple sequences, which has been shown to be beneficial, as discussed in Section 2.4.1.2. FLAIR-based detections are more sensitive than T1-w-based detections but also have higher false detection rates. Therefore, T1-w sequences are frequently utilized to eliminate false detections. Our model can be extended to account for the availability of multimodal data at a specific timepoint. Let us denote:

- m, the number of modalities,
- $I^M(\mathbf{x}) = \begin{bmatrix} I^1(\mathbf{x}) \\ \vdots \\ I^m(\mathbf{x}) \end{bmatrix}$, the multimodal measurement in the m modalities at a given voxel $\mathbf{x}$,
- $\mathcal{C} = \begin{pmatrix} \gamma_{1,1}^2 & \cdots & \gamma_{1,m}^2 \\ \vdots & \ddots & \vdots \\ \gamma_{m,1}^2 & \cdots & \gamma_{m,m}^2 \end{pmatrix}$, the covariance matrix between the m image modalities.

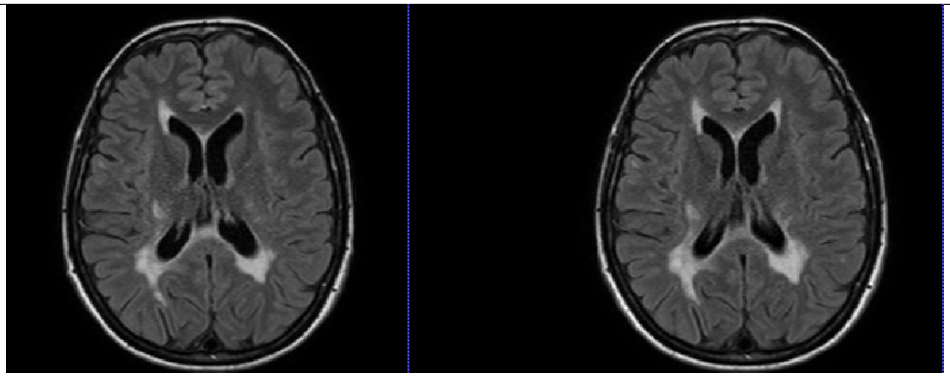
The data term can be written

$$\rho(I_1^M, I_2^M, \mathbf{w}, \mathbf{x}) = (I_2^M(\mathbf{x} - \mathbf{w}(\mathbf{x})) - I_1^M(\mathbf{x}))^\top \mathcal{C}^{-1} (I_2^M(\mathbf{x} - \mathbf{w}(\mathbf{x})) - I_1^M(\mathbf{x})). \quad (5.5)$$

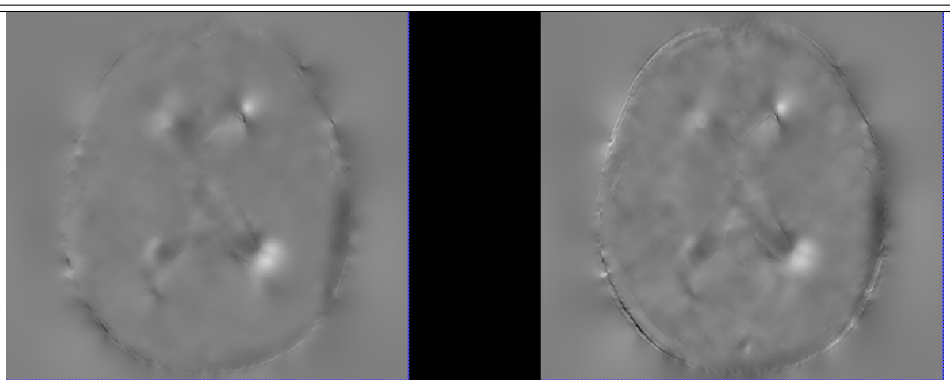
We have implemented our change detection method with the data term (5.5), considering T1-w and FLAIR modalities. In practice, we set the inter-modality covariances $\gamma_{i,j}, i \neq j$ to 0, making $\mathcal{C}$ a diagonal matrix, and therefore assumed independence between modalities. Following our model, and under the independence hypothesis, the estimated deformation field amounts to the mean deformation field weighted by the covariance matrix of the Median Absolute Distance of subtraction images for each modality combination. Figure 5.3 illustrates the difference between mono- and multi-modal deformation fields, where the jacobian norm images show that the mono-modal deformation field appears smoother than the multimodal one.

We observed that this implementation did not improve the performance compared to the unimodal model applied to the FLAIR sequence. The aim of data fusion is to exploit the strengths of both modalities while minimizing the risk of obtaining suboptimal results due to improper fusion. However, the preliminary results were not very promising as the discordance between the information from modalities was observed in some cases. A way to mitigate this problem would be to estimate the covariance matrix locally, which can help reduce the impact of areas with discordances across modalities while increasing the contribution of concordant regions.

n



I_1 and I_2



Jacobian norm of $\mathbf{w}$ in the mono-modal (left) and multi-modal (right) cases

Figure 5.3: Illustration of the deformation field of the multimodal scheme.

5.2 Normalization

As explained in section 2.2, intensity normalization is a crucial pre-processing task that involves aligning the histograms of the images. One common approach is Nyul's histogram standardization, which matches the quantiles of the histograms using a piecewise-linear function. In individual studies, changes in lesion load can impact the appearance of the normalized images, and cause decreased tissue-to-lesion contrast in case of appearing lesion, or increased contrast in case of disappearing lesion, as shown in Figure 5.4. Thus, similarly to the registration step, the normalization is also intertwined with the change detection step and should be performed jointly.

The normalization can be formulated as a function $f_\theta : \mathbb{R} \to \mathbb{R}$ such that the normalized image $f_\theta(I_2(\cdot))$ has a similar intensity distribution to I_1 within each class of tissue. Examples of the parametric form of the normalization function f_θ can be linear ($f_\theta(I_2) = aI_1 + b$, where $\theta = (a, b)$), piecewise-linear (θ would be the slopes and origins of each linear function that would be estimated), quadratic, or higher-order polynomial.

We formulate the joint estimation of the normalization parameters θ together with the registration and change detection map as the minimization of the same energy as (3.5):

$$\widehat{\mathbf{w}}, \widehat{c}, \widehat{\theta} = \underset{\mathbf{w},c,\theta}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} \left[(1 - c(\mathbf{x}))\ \rho(I_1, f_\theta(I_2), \mathbf{w}, \mathbf{x}) + \lambda_2 c(\mathbf{x})\right] + \lambda_1 \Psi(\mathbf{w}) + \lambda_3 \Phi(c). \tag{5.6}$$

We keep an alternating minimization strategy: in addition to the resolution of the two subproblems related to the registration and the change map, detailed in Section 3.2, we have to minimize (5.6) w.r.t. θ in a third step. This update of θ takes the form:

$$\widehat{\theta} = \arg\min_{\theta} \sum_{x\in\Omega} (1 - c(\mathbf{x}))\| f_\theta(I_2(\mathbf{x} + \mathbf{w}(\mathbf{x}))) - I_1(\mathbf{x})\|_2^2. \tag{5.7}$$

The problem (5.7) is equivalent to fitting the parametric curve f_θ to the joint histogram of I_1 and I_2 after registration, where the joint histogram is computed only in regions with no changes (where $c(\mathbf{x}) = 0$). We carried out experimentations using a linear-fitting of the joint histogram. In the end, we found the linear model to be irrelevant: it is not sufficient to capture the different variations in corresponding tissue intensities. This is illustrated in Figure 5.5 that displays joint histograms of successive FLAIR acquisitions, which suggests that a piecewise-linear model would be a better fit.

The initial findings of the variants discussed in Section 5.1 did not indicate a considerable improvement in change detection, but instead suggested a likely improvement in registration accuracy. This is particularly relevant for estimating atrophy, which was not assessed in this thesis but could be a promising area of research in the future.

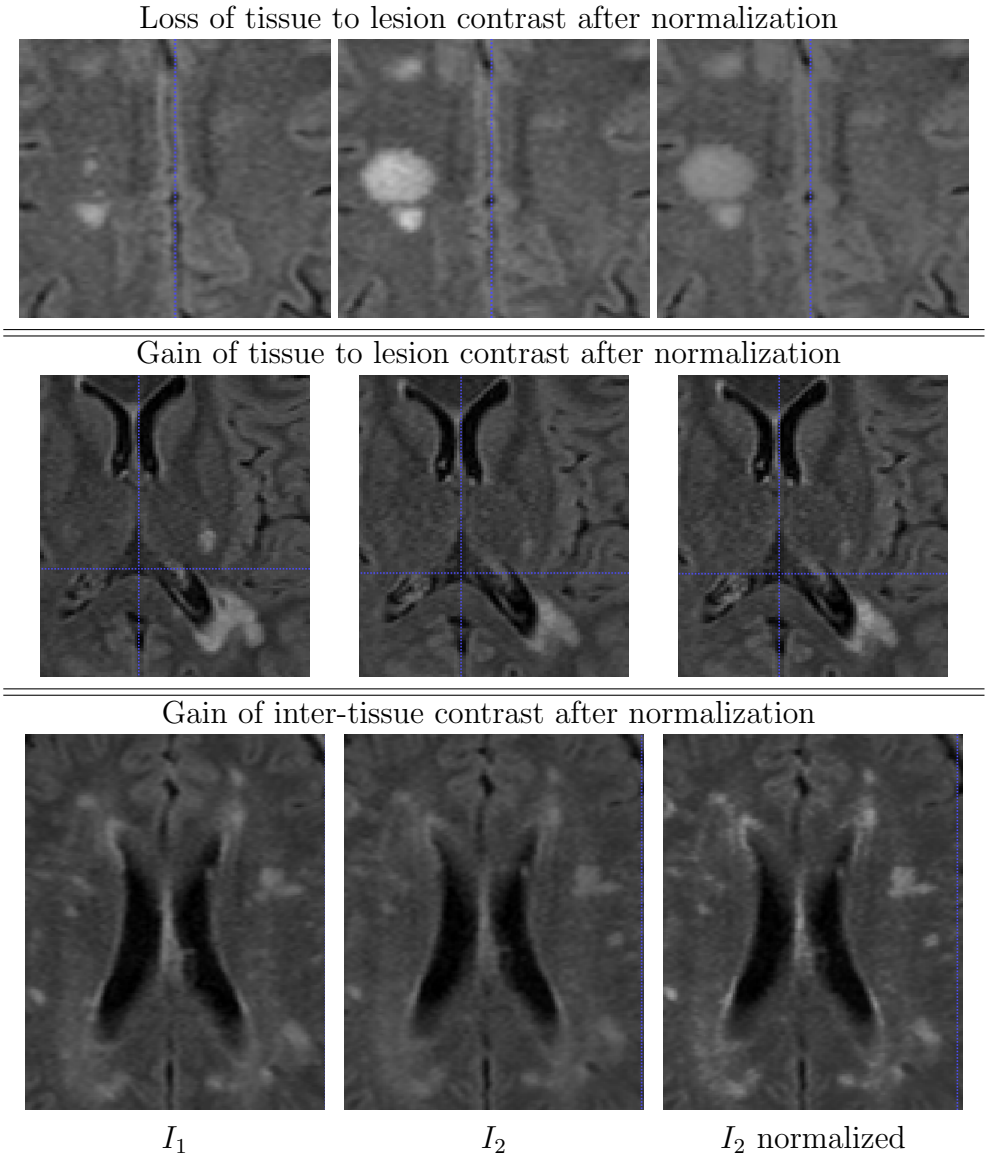


Figure 5.4: Illustration of different behaviors with Nyul's normalization.

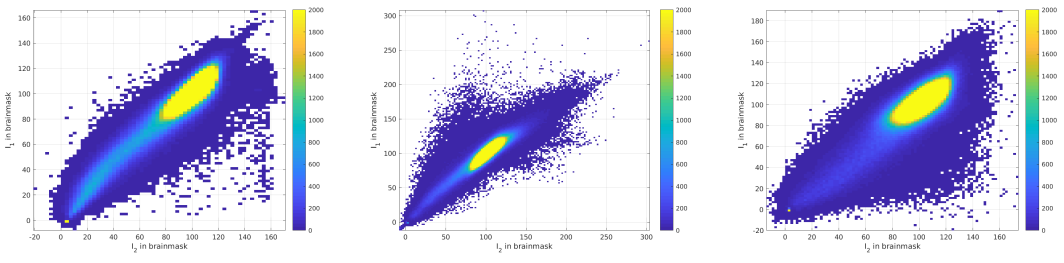


Figure 5.5: Examples of joint histograms of successive acquisitions in the brainmask

Conclusion

The aim of this thesis work was to propose a method for detecting emerging, vanishing, expanding, or contracting MS lesions in both white-matter and gray-matter tissue. The classical pipeline for detecting changes in MS lesions in subtraction-based and deformation-based methods processes registration and change detection sequentially. However, accurate change detection cannot be achieved without proper registration, but the presence of changes in an image can hinder registration. We hypothesize that the decoupling of these problems is the cause of these limitations.

To address this issue, in this thesis work, we have proposed a method for integrating both tasks in a unified model. Our approach involves jointly modeling these two tasks as a single objective function. To solve the optimization problem, the alternated direction method of multiplier (ADMM) framework was used for registration, and graph-cut was used for change detection. The ADMM framework is flexible enough to cope with different registration models with low computational cost, making it advantageous for the problem at hand.

We evaluated the proposed approach in the context of multiple sclerosis lesion follow-up, where deformable registration is required to capture brain atrophy, which can impair the performance of the change detection method. The presented results demonstrate the superior performance of the joint model in longitudinal change detection, compared to its sequential counterparts. While the affine method provided good lesion coverage, it was prone to false positive detections due to brain atrophy. Both the sequential and joint methods can correct for brain atrophy deformations, but the sequential method failed to accurately detect lesion changes due to over-compensation. This limitation was overcome by the joint approach, which successfully detected more extensive lesion areas. In longitudinal acquisitions without lesion load evolution, the joint approach produced significantly fewer detections than the affine method. Thus, our proposed joint approach combines the deformable registration's ability to correct brain atrophy and the preservation of lesions' shape to ensure accurate change detection. However, if the lesions are too small and overcompensated during the first iteration of the joint approach, their recovery through the iterations may be incomplete, leading to less sensitive detections.

While the current implementation is based on relatively simple modeling assumptions, the optimization framework can be extended to accommodate more sophisticated models, incorporating alternative data fidelity terms and regularization techniques. Possible extensions of the model include data-terms that exhibit increased noise resilience, incorporation of multi-channel information, symmetric registration, or integration of

intensity normalization tasks within the joint model. The implemented experiments suggested potential for improvement in registration accuracy, which is particularly relevant for estimating atrophy - an area not assessed in this thesis but with promising research opportunities in the future.

Chapter 6

Résumé en français

Le sujet de cette thèse est l'analyse longitudinale d'IRM cérébrales de patients atteints de sclérose en plaques (SEP). La sclérose en plaques est une maladie auto-immune dégénérative du système nerveux central. Elle se caractérise par la présence de lésions évolutives et une atrophie cérébrale survenant plus rapidement que la normale. En pratique clinique, l'imagerie par résonance magnétique (IRM) est privilégiée pour le diagnostic et le suivi de l'évolution de la pathologie. Cependant, l'analyse visuelle des acquisitions est une tâche laborieuse et coûteuse en temps. L'appréciation de l'évolution des lésions est subjective et sujette à une forte variabilité intra- et inter-opérateur.

La détection automatique des changements entre deux acquisitions d'IRM cérébrales est un outil important pour l'analyse longitudinale de la maladie. Dans les essais cliniques, cela pourrait servir à extraire les biomarqueurs dérivés de l'imagerie afin d'évaluer les effets d'un traitement. Dans les examens longitudinaux d'IRM d'un sujet donné, plusieurs facteurs peuvent entraîner des différences entre deux images, tels que les changements de charge lésionnelle, la déformation des structures cérébrales due à l'atrophie ou les artefacts d'acquisition. Le principal défi est d'identifier l'origine des changements afin de quantifier leur contribution individuelle aux changements globaux. La plupart des méthodes de détection des changements de lésions reposent sur un schéma séquentiel qui corrige successivement les différentes sources de changement. Cependant, corriger chaque source de changement séparément sans tenir compte de leur interdépendance peut entraîner des estimations biaisées.

Le recalage est une étape particulièrement cruciale qui peut avoir un impact majeur sur la performance de la détection des changements de lésions de la SEP. La pratique la plus courante consiste à employer des modèles de transformation rigides ou affines pour la correction globale du positionnement du patient sans altérer l'apparence des lésions. Cependant, de tels modèles linéaires ne sont pas capables de capturer les déformations complexes induites par l'atrophie cérébrale, qui se produit généralement dans la SEP. Ces déformations résiduelles peuvent entraîner des détections erronées dans les zones atrophiées, notamment dans le cortex et autour des ventricules. Le recalage déformable est une solution pour estimer des déformations plus complexes, mais il a tendance à faire disparaître les nouvelles lésions en minimisant la dissimilarité entre les deux images, ce qui résulte en une détection qui manque de sensibilité. Cette limitation des approches

séquentielles est due à la dissociation des problèmes de recalage et de détection de changements.

Contributions

Notre contribution vise à concilier le recalage déformable et la détection de changements en les estimant conjointement. Nous formulons ces deux tâches distinctes comme un problème d'optimisation unique impliquant une seule fonction de coût englobant ces deux problèmes. Les zones correspondant aux changements détectés sont ignorées dans le critère de similarité de recalage, ce qui empêche l'effet d'élimination des lésions décrit ci-dessus. Nous proposons un schéma d'optimisation alternée efficace pour résoudre ce problème d'optimisation unifié. Nous nous concentrons sur la démonstration des avantages de cette formulation conjointe en utilisant un critère strandard de similarité de données basé sur l'intensité sur des patients atteints de SEP synthétiques et réels. Nous comparons ces résultats à ceux obtenus avec le recalage déformable suivi de la détection des changements et avec le recalage affine suivi de la détection des changements. Pour améliorer encore la détection des changements de lésions, nous proposons des extensions possibles du modèle qui intègrent la normalisation de l'intensité, le recalage symétrique et l'intégration multimodalité dans la minimisation conjointe. Notre travail a été publié dans un article de conférence internationale pour ISBI 2020 , un article de conférence nationale pour GRETSI2022, et un article de revue.

Plan du manuscrit

- **Chapitre 1** : Ce chapitre pose le contexte scientifique de la thèse. Il présente les caractéristiques de la sclérose en plaques et de son évolution. Ensuite, les principes de base de l'IRM et sa pertinence pour le diagnostic et le suivi des patients en milieu clinique sont présentés. Le chapitre aborde également les difficultés rencontrées avec l'analyse automatisée de l'IRM. Enfin, le flux de travail clinique standard est expliqué, mettant en évidence l'importance du traitement automatisé pour les examens longitudinaux par IRM.

- **Chapitre 2 :** Ce chapitre expose le processus séquentiel classique de détection automatisée des changements de lésions dans l'IRM cérébrale. Il couvre les techniques utilisées pour éliminer les sources de changement non pertinentes causées par des facteurs tels que les paramètres d'acquisition de l'IRM, la position du patient et les fausses détections en dehors du cerveau. Le chapitre présente également les principes de base du recalage longitudinal par IRM du cerveau, passe en revue l'état de l'art des méthodes de détection des changements de lésions et identifie les inconvénients des approches séquentielles.

- **Chapitre 3 :** Notre hypothèse est que les limitations des approches séquentielles découlent du traitement du recalage et de la détection des changements comme des tâches distinctes. Dans ce chapitre, nous proposons un cadre unifié qui effectue ces tâches conjointement. Nous formulons d'abord le recalage et la détection des

changements comme des problèmes de minimisation et expliquons comment ils peuvent être combinés en un seul problème de minimisation. Nous résolvons ce problème avec une minimisation alternée par rapport au champ de déformation et à la carte de détection des changements. Nous utilisons une approche basée sur la méthode des multiplicateurs de directions alternées (ADMM) pour le sous-problème de recalage et le graph-cut pour le sous-problème de détection des changements.

- **Chapitre 4 :** Ce chapitre présente une comparaison entre les résultats du modèle proposé et les approches séquentielles conventionnelles. Le cadre d'évaluation est composé d'un ensemble de données synthétiques et de deux ensembles de données de patients atteints de sclérose en plaques disponibles publiquement et longitudinaux. L'évaluation de la performance du modèle conjoint est réalisée à l'aide de métriques basées voxel et basées lésions. Le chapitre décrit également le processus suivi pour sélectionner les valeurs appropriées des hyperparamètres du modèle. Les résultats quantitatifs pour les ensembles de données des patients synthétiques et réels sont présentés, indiquant que le modèle conjoint offre une meilleure sensibilité globale par rapport au modèle séquentiel avec recalage déformable et une meilleure spécificité globale par rapport au modèle séquentiel avec recalage affine. De plus, les limitations du modèle proposé sont discutées.

- **Chapitre 5 :** Ce chapitre illustre l'adaptabilité du cadre unifié proposé en introduisant des variantes du terme de données qui présentent une résilience accrue au bruit, intègrent des informations multi-canaux, permettent le recalage symétrique ou intègrent la tâche de normalisation de l'intensité dans le modèle conjoint. La mise en œuvre de ces variantes et leur potentiel pour résoudre certaines des limitations du modèle existant sont discutées.

Contexte médical

La sclérose en plaques (SEP) est une maladie neurodégénérative auto-immune qui touche plus de 2,3 millions de personnes dans le monde1 et 110 000 personnes en France. Elle est généralement diagnostiquée entre 25 et 35 ans avec une prévalence plus élevée chez les femmes (rapport hommes/femmes de 1 pour 3). Les mécanismes auto-immuns attaquent la gaine de myéline qui entoure les nerfs, entraînant l'apparition de lésions appelées plaques. Ces lésions correspondent à une inflammation, à une démyélinisation et souvent à une dégradation axonale. La gaine de myéline assure le transfert correct des impulsions nerveuses des neurones aux terminaisons nerveuses. Les dommages causés par l'inflammation peuvent perturber la transmission des signaux dans le corps, entraînant divers types d'incapacités, notamment motrices (comme des engourdissements ou une faiblesse dans les membres, des sensations de chocs électriques, des tremblements, un manque de coordination et une démarche instable), sensorielles (comme une perte partielle ou complète de la vision, une vision double prolongée et une vision floue) ou des troubles cognitifs.

La progression de la SEP et de ses symptômes varie considérablement d'une personne à l'autre. Le diagnostic est établi en observant les signes cliniques et les lésions cérébrales grâce à l'imagerie par résonance magnétique (IRM). Les symptômes peuvent évoluer avec le temps et varier d'une personne à l'autre en fonction des zones du cerveau ou de la moelle épinière touchées par les lésions. De plus, l'atrophie cérébrale progresse à un rythme plus rapide que le vieillissement normal.

Pour diagnostiquer la sclérose en plaques (SEP), les cliniciens recherchent généralement la présence de lésions dans des emplacements spécifiques du système nerveux central, tels que le cerveau, la moelle épinière et les nerfs optiques. Ils tiennent également compte de la dissémination dans le temps de l'apparition de ces lésions et excluent toute autre explication potentielle de leur présence. Au fil de la maladie, les IRM sont utilisées pour suivre la progression de la SEP car elles sont plus sensibles pour détecter l'activité de la maladie par rapport aux mesures cliniques. Des facteurs tels que l'emplacement et le nombre de lésions, ainsi que la perte de volume cérébral, peuvent fournir des informations importantes sur le stade et la progression de la maladie, et sont utilisés pour surveiller la rémission, les rechutes, la réponse au traitement et le développement de handicaps.

La comparaison visuelle de deux IRM pour détecter de nouvelles lésions est assez chronophage et sujette aux erreurs. La sensibilité du praticien aux lésions nouvelles ou évolutives peut être améliorée par des outils automatisés. Parfois, des outils automatisés sont fournis sur la console d'IRM pour la comparaison des IRM ; par exemple, les appareils d'IRM de Philips fournissent un outil qui met en évidence les zones de changement pour faciliter l'inspection visuelle par les praticiens.

Analyser l'évolution des lésions peut s'avérer difficile pour les professionnels de la santé, et le traitement automatisé des acquisitions longitudinales est difficile en raison de la présence de diverses sources de changement non pertinentes qui doivent être corrigées. Cette thèse se concentre sur la détection des lésions émergentes, disparues, expansives ou rétractiles dans les tissus de matière blanche et grise.

Modèle séquentiel classique

Notons $I_1, I_2 : \Omega \to \mathbb{R}$ deux acquisitions IRM 3D successives, où Ω est le domaine d'une image. L'approche séquentielle est composée de deux étapes : premièrement, recaler I_2 sur I_1, puis détecter les changements entre les images recalées.

Recalage La plupart des méthodes de recalage peuvent être formulées par :

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}:\Omega\to\mathbb{R}^3}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_1 \Psi(\mathbf{w}), \tag{6.1}$$

où $\widehat{\mathbf{w}}$ est le champ de déformation dense entre I_1 et I_2, $\lambda_1 > 0$ étant un paramètre de pondération. Le terme d'attache aux données est une mesure de similarité assurant la conservation de certaines caractéristiques de l'image après déformation et le terme de régularisation favorise les solutions lisses. Plusieurs choix pour $\rho(\cdot)$ et $\Psi(\cdot)$ sont

possibles ; l'article [Sotiras et al., 2013] donne une vue d'ensemble des méthodes de recalage déformable en imagerie médicale.

Détection de changements En général, l'étape de détection de changement consiste à seuiller une carte de différence de caractéristiques calculées sur les acquisitions recalées similaire au terme $\rho(I_1, I_2, \mathbf{w}, \mathbf{x})$. Pour chaque voxel $\mathbf{x}$, la carte binaire de changements $c : \Omega \to 0, 1$ vaut :

$$\widehat{c}(\mathbf{x}) = \begin{cases} 0 & \text{if } \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) \leq \lambda_2 \\ 1 & \text{otherwise,} \end{cases} \tag{6.2}$$

où λ_2 est un seuil de détection. Il est possible de reformuler (6.2) comme un problème de minimisation :

$$\widehat{c}(\mathbf{x}) = \underset{c:\Omega\to\{0,1\}}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} (1 - c(\mathbf{x}))\ \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_2 c(\mathbf{x}). \tag{6.3}$$

La plupart des caractéristiques seuillées sont basées sur l'intensité [Bosc et al., 2003; Cabezas et al., 2016; Ganiler et al., 2014; Salem et al., 2018; Sweeney et al., 2013] ou le champ de déformation [Cabezas et al., 2016; Rey et al., 2002; Salem et al., 2018] (vue d'ensemble dans [Lladó et al., 2012]). Cependant, un seuillage simple résulte souvent dans des cartes de détections bruitées. C'est pourquoi il est régulièrement précédé ou suivi d'une étape de lissage pour prendre en compte les informations spatiales du voisinage.

Approche conjointe proposée

Pour surmonter les limitations de l'approche séquentielle, nous proposons une modélisation conjointe du recalage et de la détection des changements. Les deux étapes sont fondamentalement liées, car le recalage vise à trouver des correspondances entre les images, tandis que la détection des changements détermine les régions qui n'admettent pas de correspondances. Par conséquent, les deux tâches devraient être définies avec la même fonction objective pour travailler en synergie. Nous formulons le problème de minimisation conjoint suivant qui atteint cet objectif en unifiant les principes décrits précédemment :

$$\widehat{\mathbf{w}}, \widehat{c} = \underset{\mathbf{w},c}{\operatorname{argmin}} \sum_{\mathbf{x}\in\Omega} \left[(1 - c(\mathbf{x}))\ \rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_2 c(\mathbf{x})\right] + \lambda_1 \Psi(\mathbf{w}) + \lambda_3 \Phi(c). \tag{6.4}$$

Avec ce modèle, le terme de données est annulé dans les régions de changement (où $c(\mathbf{x}) = 1$), de sorte que l'estimation de la transformation est uniquement dirigée par le terme de régularisation $\Psi(\mathbf{w})$, produisant ainsi un champ de déformation lissé dans ces régions. Un terme de régularisation sur c est également inclus via le terme $\Phi(c)$.

Spécification des termes de la fonction de coût

Tout d'abord, nous supposons qu'après la normalisation d'intensité effectuée lors du prétraitement, les intensités des deux images sont comparables. Ainsi, nous considérons un terme de données standard qui pénalise les déviations par rapport à la constance de l'intensité avec une norme ℓ_2 :

$$\rho(I_1, I_2, \mathbf{w}, \mathbf{x}) = \frac{1}{\sigma^2} \|I_2(\mathbf{x} - \mathbf{w}(\mathbf{x})) - I_1(\mathbf{x})\|_2^2, \tag{6.5}$$

où σ est une constante de normalisation définie par la déviation médiane absolue des différences d'intensité entre I_1 et I_2. Ce terme de données est représentatif des caractéristiques basées sur l'intensité couramment utilisées dans les méthodes de détection des changements.

Deuxièmement, nous supposons que les déformations induites par l'atrophie du tissu cérébral sont complexes mais toujours localement lisses. C'est pourquoi nous considérons un terme de régularisation de Tikhonov du premier ordre qui pénalise le gradient du champ de déformation :

$$\Psi(\mathbf{w}) = \sum_{\mathbf{x}\in\Omega} \|\nabla \mathbf{w}(\mathbf{x})\|_2^2, \tag{6.6}$$

où $\nabla\cdot$ est l'opérateur de gradient.

Enfin, pour garantir une détection cohérente spatialement des lésions, nous régularisons la carte de changement c avec un modèle de Potts binaire standard :

$$\Phi(c) = \sum_{\mathbf{x}\in\Omega} \sum_{\mathbf{y}\in\mathcal{N}(\mathbf{x})} (1 - \delta\left(c(\mathbf{x}), c(\mathbf{y})\right)), \tag{6.7}$$

où δ est la fonction de Kronecker égale à 1 si son argument est vrai et 0 sinon, et $\mathcal{N}(\mathbf{x})$ est le voisinage 6 de $\mathbf{x}$.

Optimisation

Une stratégie de minimisation alternée est adoptée pour résoudre le problème (6.4). A chaque itération, $\mathbf{w}$ est mis à jour avec c fixé, puis c est mis à jour avec $\mathbf{w}$ fixé. Nous détaillons dans cette section comment nous résolvons ces deux sous-problèmes.

Minimisation par rapport à w avec c fixé Une pratique courante en recalage et estimation du mouvement est de considérer la version linéarisée du terme de données (6.5) en remplaçant $I_2(\mathbf{y} - \mathbf{w}(\mathbf{x}))$ par son développement de Taylor en y [Hill et al., 2001]. Le champ de déformation w est estimé avec la méthode ADMM (Alternated Direction Method of Multipliers) [14]. Nous transformons le problème initial en un problème sous contrainte en introduisant la variable $\mathbf{z}$, qui sépare le terme de données et le terme de régularisation :

$$\underset{\mathbf{w}}{\operatorname{argmin}} \sum_x (1 - c(\mathbf{x}))\rho(I_1, I_2, \mathbf{w}, \mathbf{x}) + \lambda_1 \Psi(\mathbf{z}), \text{ s.t. } \mathbf{w} = \mathbf{z}. \tag{6.8}$$

L'algorithme ADMM consiste à minimiser le Lagrangien augmenté associé à cette forme contrainte par rapport à $\mathbf{w}$ puis à $\mathbf{z}$, et à mettre à jour la variable duale par montée de gradient [**?**]. Les équations de mise à jour de $\mathbf{w}$ et $\mathbf{z}$ sont :

$$\mathbf{w}^{k+1} = \text{prox}_{\sum_x (1-c(\mathbf{x}))\rho(I_1, I_2, \cdot, \mathbf{x})} \left(\mathbf{z} - \frac{\boldsymbol{\alpha}}{\mu} \right) \tag{6.9}$$

$$\mathbf{z}^{k+1} = \text{prox}_{\lambda_1 \Psi} \left(\mathbf{w} + \frac{\boldsymbol{\alpha}}{\mu} \right) \tag{6.10}$$

$$\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k + \mu(\mathbf{w}^{k+1} - \mathbf{z}^{k+1}) \tag{6.11}$$

$\boldsymbol{\alpha}$ est la variable duale, μ est un paramètre de pénalité, et $\text{prox}_f(\mathbf{x}) = \underset{\mathbf{y}}{\text{argmin}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + f(\mathbf{y})$ est l'opérateur proximal de f. Le sous-problème (6.9) est voxellique et quadratique et admet une solution analytique simple. Le sous-problème (6.10) est équivalent à une opération de débruitage régularisé par $\Psi(\cdot)$; elle admet également une solution analytique calculable efficacement dans le domaine de Fourier. La linéarisation du terme de données restreint l'amplitude des déformations. C'est pourquoi, afin de gérer d'importants niveaux d'atrophie, l'estimation est intégrée à une approche multi-échelle [Brox et al., 2004]. Le cadre ADMM est suffisamment flexible pour s'accommoder de différentes modélisations de ρ et ψ (voir [Fortun et al., 2018] pour d'autres combinaisons de termes de données et régularisation).

Minimisation par rapport à c avec w fixé Le problème (6.4) par rapport à c revient à une segmentation binaire avec un terme de régularisation. Il est résolu exactement par une approche de graph-cut [Boykov et al., 2001].

Résultats

Nous évaluons notre méthode sur un ensemble de données synthétiques et deux ensembles de données réelles de patients disponibles publiquement. L'ensemble de données synthétiques présente l'avantage d'avoir une carte de détection des changements de référence clairement définie, tout en contrôlant le niveau de bruit, les inhomogénéités de champs et l'atrophie cérébrale qui altèrent les images.

Les ensembles de données réelles sont utilisés pour évaluer l'approche proposée dans des conditions plus proches de la routine clinique, avec différentes conditions d'acquisition et diverses évolutions pathologiques. Le premier ensemble de données de patients réels, désigné sous le nom de *LesjakDB* [Lesjak et al., 2016], sert à l'évaluation de la capacité des méthodes à détecter tous types d'évolutions des lésions de la sclérose en plaques (réduction, croissance, nouvelles lésions et disparition), tandis que le deuxième ensemble de données, désigné sous le nom de *MSSEG-2* [Commowick et al., 2021], se concentre uniquement sur la capacité à détecter les nouvelles lésions.

La formulation conjointe proposée dans l'équation (6.4) nécessite le réglage de trois hyperparamètres :

- λ_1 contrôle la régularisation spatiale du champ de déformation,

- λ_2 sert de seuil pour la carte de différence d'intensité,
- λ_3 régit la régularisation spatiale de la carte de changement.

Ajuster ces hyperparamètres en l'absence de vérité terrain peut être difficile, et laisser l'utilisateur régler empiriquement ces hyperparamètres peut introduire un biais significatif dans les résultats de sortie. Pour rendre l'algorithme plus utilisable dans un contexte clinique, nous avons trouvé un ensemble de valeurs qui fonctionnerait pour la plupart des paires d'IRM longitudinales.

Nous présentons six métriques (quatre métriques basées sur les voxels et deux métriques basées sur les lésions) pour évaluer la performance des méthodes de détection des changements, à savoir le coefficient de similarité de Dice (DSC), la valeur prédictive positive (PPV), le taux de vrais positifs (TPR) et le DSC local. Soient TP, TN, FP et FN le nombre de voxels de la carte de détection des changements estimée correspondant respectivement à Vrais Positifs, Vrais Négatifs, Faux Positifs et Faux Négatifs.
Le DSC est défini comme suit :

$$DSC = 2TP/(2TP + FP + FN)$$

et reflète la bonne superposition globale entre la carte de détection et la vérité terrain.
Le PPV est défini comme suit :

$$PPV = TP/(TP + FP)$$

et reflète la proportion de détections pertinentes parmi tous les changements détectés.
Le TPR est défini comme suit :

$$TPR = TP/(TP + FN)$$

et reflète la proportion des changements réels qui ont été détectés.
Le DSC local correspond au DSC calculé sur une zone restreinte définie comme la dilatation avec un élément structurant sphérique de rayon 4 voxels par rapport à la vérité terrain. Cette métrique nous permet de focaliser l'évaluation sur la précision spatiale locale de la méthode de détection.
En plus des métriques basées sur les voxels, nous rapportons également des métriques basées sur les lésions, à savoir le taux de vrais positifs pour les lésions (L-TPR) et la valeur prédictive positive pour les lésions (L-PPV). Ces métriques ont été évaluées grâce à l'outil de validation *animaSegPerfAnalyzer* en considérant les mêmes hyperparamètres que dans [Commowick et al., 2018] :
Étant donné que toutes ces métriques ne sont pas pertinentes pour les données qui ne présentent aucun changement, nous considérons dans ce cas spécifique le nombre de composantes connexes détectés ainsi que le volume des changements détectés pour caractériser les détections de faux positifs.

Pour démontrer les avantages du modèle conjoint proposé, nous considérons trois variantes du cadre de détection des changements :

- *conjoint* : la méthode conjointe proposée pour la détection des changements et l'inscription décrite dans la Section 6.

- *séquentiel* : Le pendant séquentiel de la méthode proposée, qui effectue successivement l'inscription déformable et la détection des changements. Pour les deux étapes, nous utilisons le même modèle et les mêmes algorithmes d'optimisation que dans les sous-étapes de l'approche *conjointe* décrite dans la thèse.

- *affine* : L'approche *séquentielle* où l'inscription déformable a été remplacée par une inscription affine, ce qui correspond au cas le plus courant. L'inscription affine a été estimée à l'aide de la bibliothèque ANTs [Avants et al., 2011][1] avec les paramètres par défaut et la métrique d'information mutuelle. Ensuite, le seuillage et le lissage de la carte des changements suivent le modèle déformable séquentiel.

Ensemble de données synthétiques

La méthode *affine* parvient à détecter presque toutes les zones de lésion, mais elle souffre de détections de faux positifs autour des ventricules en raison de l'atrophie cérébrale. Les méthodes *séquentielle* et *conjointe* compensent la déformation due à l'atrophie cérébrale car aucune d'entre elles ne présente de fausses détections autour des ventricules. Cependant, la méthode *séquentielle* échoue à détecter l'ensemble des zones de lésion en raison de la surcompensation des changements de lésion. Cette limitation est surmontée par l'approche *conjointe* qui parvient à détecter l'intégralité des zones de lésion.

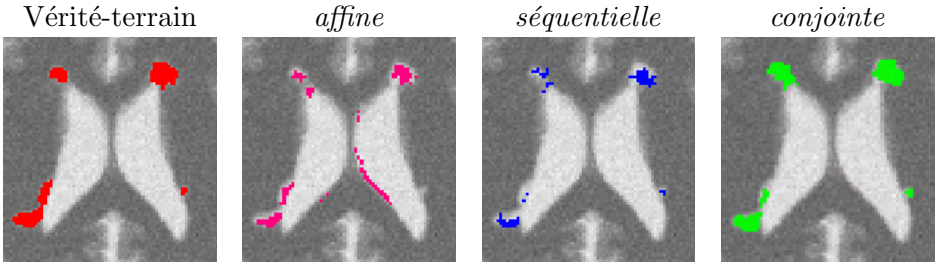


Figure 6.1: Comparaison qualitative des cartes de détection des changements binaires obtenues avec les trois méthodes sur l'ensemble de données synthétiques (scénario d'apparition de lésions avec atrophie)

Une comparaison quantitative des trois méthodes dans quatre scénarios est présentée dans le Tableau 6.1. Tout d'abord, nous examinons l'apparition de lésions sans atrophie. Sans surprise, ce scénario est le plus favorable pour la méthode *affine* puisqu'il n'y a aucune différence géométrique à compenser. L'approche *séquentielle* conduit à des valeurs significativement plus faibles de DSC et de DSC local. Cela est dû à l'effet de surcompensation des lésions, comme le confirme la faible valeur de TPR observée (*c'est-à-dire* le manque de sensibilité) et la haute valeur de PPV (*c'est-à-dire* la haute spécificité). Enfin, l'approche *conjointe* surmonte le défaut de l'approche *séquentielle* et présente des performances similaires à la méthode *affine*, avec une légère tendance à

[1] https://github.com/ANTsX/ANTs

Table 6.1: Résultats obtenus sur l'ensemble de données synthétiques.

Scénario	Méthode	DSC	PPV	TPR	DSC local
Apparition des lésions Sans atrophie	*affine*	**0,830**	0,782	**0,885**	**0,830**
	séquentielle	0,684	**0,921**	0,544	0,684
	conjointe	0,814	0,887	0,751	0,814
Croissance des lésions Sans atrophie	*affine*	0,766	0,635	**0,964**	0,770
	séquentielle	0,685	**0,734**	0,641	0,688
	conjointe	**0,806**	0,726	0,902	**0,808**
Apparition des lésions Atrophie simulée	*affine*	0,460	0,329	**0,767**	**0,810**
	séquentielle	0,626	**0,960**	0,465	0,627
	conjointe	**0,743**	0,925	0,621	0,744
Croissance des lésions Atrophie simulée	*affine*	0,652	0,505	**0,919**	0,827
	séquentielle	0,753	**0,869**	0,664	0,754
	conjointe	**0,847**	0,833	0,861	**0,848**

sous-estimer la zone détectée. Des observations similaires peuvent être faites pour le deuxième scénario impliquant une croissance de lésion sans atrophie.

Les conclusions sont radicalement différentes pour les deux scénarios impliquant une atrophie simulée. Les performances de la méthode *affine* sont significativement entravées par les nombreuses fausses détections dues à l'atrophie. Cela est illustré par la diminution significative des valeurs de DSC et de PPV par rapport aux cas sans atrophie, tandis que les valeurs de TPR et de DSC local sont moins modifiées. L'approche *séquentielle* parvient à compenser l'atrophie cérébrale simulée, comme le montre la valeur élevée de PPV, mais sous-estime toujours les changements à détecter, comme l'indique la faible valeur de TPR. L'approche *conjointe* surpasse clairement les deux approches précédentes en termes de précision de détection, comme en témoigne la valeur de DSC significativement plus élevée.

Le comportement de l'approche *conjointe* peut être suivi à travers les itérations du schéma d'optimisation alternée et est illustré dans la Fig. 6.2. On peut observer que le DSC augmente au fil des itérations et que la convergence est atteinte en quelques itérations. En ce qui concerne le coût computationnel de l'approche conjointe, il est d'environ 24 minutes sur un seul cœur (Intel(R) Xeon(R) Gold 6130 CPU @ 2,10 GHz) pour une expérience sur l'ensemble de données synthétiques (taille de l'image : 181x217x181).

LesjakDB

Tout d'abord, une comparaison visuelle qualitative des trois méthodes est présentée dans la Fig. 6.3. La Fig. 6.1 montre que nous pouvons tirer des conclusions similaires à celles de l'ensemble de données synthétiques. La méthode *affine* démontre une sensibilité élevée (c'est-à-dire que l'évolution de la lésion est bien détectée) mais un manque de spécificité (c'est-à-dire de nombreuses détections de faux positifs autour des ventricules et dans la partie postérieure du cortex). En revanche, la méthode *séquentielle* a une

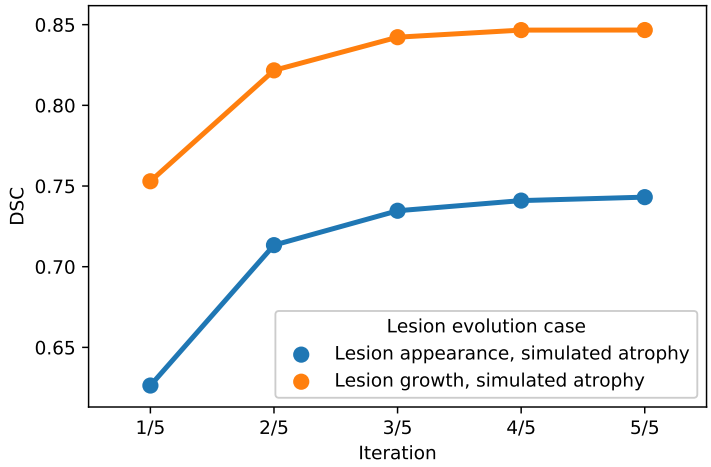


Figure 6.2: Évolution du DSC au fil des itérations du schéma d'optimisation alternée de l'approche *conjointe* sur l'ensemble de données synthétiques (en bleu : apparition des lésions, atrophie simulée, en orange : croissance des lésions, atrophie simulée).

spécificité élevée mais manque de sensibilité. L'approche *conjointe* offre les meilleurs résultats visuels, ce qui illustre sa capacité à atteindre à la fois une sensibilité élevée et une spécificité élevée. Les Figures 6.3(c) et 6.3(d) montrent le jacobien des champs de déformation obtenus par les méthodes *séquentielle* et *conjointe*, respectivement. Le motif spécifique caractérisé par l'alternance de valeurs élevées et faibles du jacobien (voir les zones mises en évidence par les carrés rouges dans la Fig. 6.3(c)) reflète la contraction et la dilatation locales importantes induites par le champ de déformation pour faire disparaître la lésion, expliquant ainsi le manque de sensibilité des résultats de détection.

L'évaluation quantitative présentée dans les Tableaux 4.3 et 4.4 (premières lignes) ainsi que dans la Fig. 6.4 confirme les conclusions de l'analyse visuelle. La sensibilité élevée de la méthode *affine* est objectivée au niveau du voxel par un TPR statistiquement significativement plus élevé que les deux autres méthodes. Au niveau de la lésion, les trois méthodes présentent des valeurs L-TPR similaires, mettant ainsi en évidence leur capacité à détecter la même quantité de zones changeantes. Tant les méthodes *séquentielle* que *conjointe* produisent des PPV et L-PPV significativement plus élevés par rapport à la méthode *affine*, ce qui illustre leur capacité à réduire le nombre de fausses détections induites par l'atrophie cérébrale tant au niveau du voxel que de la lésion. Ce résultat met en évidence l'avantage de l'utilisation du recalage déformable dans le contexte du suivi des lésions de la SEP. La sensibilité significativement plus faible obtenue par la méthode *séquentielle* par rapport à la méthode *conjointe* est la conséquence de l'effet de surcompensation de la lésion. Enfin, l'approche *conjointe* surpasse de manière significative les deux autres approches en termes de précision globale au niveau du voxel (voir DSC).

La Fig. 6.5 met en évidence la variabilité des performances des méthodes entre

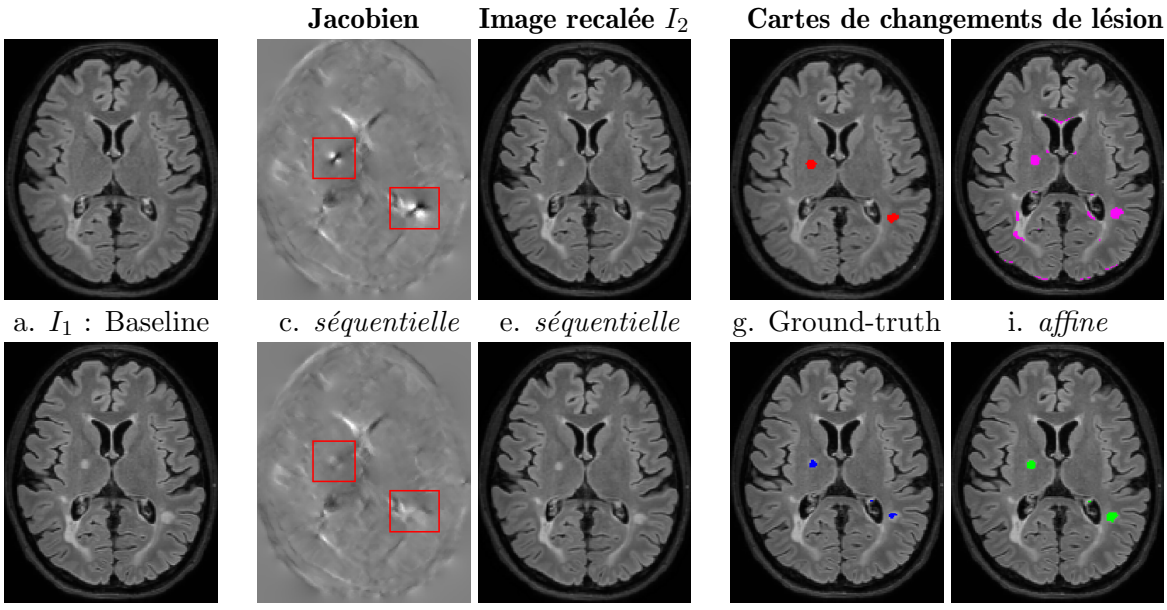


Figure 6.3: Comparaison qualitative des cartes binaires de détection de changements obtenues avec les trois méthodes et du jacobien du champ de déformation estimé avec les approches *séquentielle* et *conjointe* sur un sujet sélectionné de l'ensemble de données MSSEG-2. Hyperparamètres : $\lambda_1 = 70$ (pour les méthodes *séquentielle* et *conjointe*), $\lambda_2 = 25$, $\lambda_3 = 3$

les sujets. Il est intéressant de noter que, bien que les performances des méthodes de détection varient considérablement d'un sujet à l'autre, le classement parmi les trois méthodes semble être très cohérent d'un sujet à l'autre. Lors de l'investigation des facteurs pouvant expliquer cette variabilité observée, il semble que le volume de la ground-truth joue un rôle prépondérant : plus le volume à détecter est grand, meilleures sont les performances de l'algorithme de détection de changement, comme le montre la Fig. 6.6.

MSSEG-2

De manière similaire aux ensembles de données synthétiques et LesjakDB, la comparaison visuelle qualitative des deux approches basées sur le recalage déformable dans la Fig. 6.7 met en évidence le manque de sensibilité de la méthode *séquentielle* en raison de l'effet de surcompensation des lésions. La carte de détection du changement résultante (violet) est trop petite par rapport à la vérité terrain (en rouge transparent en dessous) en raison du recalage déformable qui réduit significativement la taille de la lésion. Avec l'approche *jointe*, la forme de la lésion est presque préservée dans l'image de suivi transformée et la carte de détection du changement (en vert) correspond presque parfaitement à la vérité terrain.

L'évaluation quantitative sur le sous-ensemble MSSEG-2-Change est rapportée dans la deuxième rangée des Tables 6.2 et 6.3 et dans la partie supérieure de la Fig. 6.8.

Table 6.2: Résultats calculés sur les ensembles de données LesjakDB et MSSEG-2-Change (mesures au niveau du voxel). La médiane ± l'écart absolu médian (MAD) calculée sur l'ensemble des sujets est rapportée pour chaque métrique. L'expérience inverse MSSEG-2-Change consiste à échanger les images de référence et de suivi pour évaluer la capacité des méthodes à détecter les lésions disparues.

Ensemble de données	Méthode	DSC local	DSC	PPV	TPR
LesjakDB	*affine*	$\mathbf{0.539 \pm 0.174}$	0.152 ± 0.087	0.086 ± 0.060	$\mathbf{0.603 \pm 0.223}$
	sequential	0.424 ± 0.139	0.317 ± 0.117	0.293 ± 0.152	0.323 ± 0.125
	joint	0.501 ± 0.179	$\mathbf{0.353 \pm 0.144}$	$\mathbf{0.323 \pm 0.148}$	0.447 ± 0.207
MSSEG-2-Change	*affine*	$\mathbf{0.626 \pm 0.224}$	0.142 ± 0.165	0.081 ± 0.139	$\mathbf{0.633 \pm 0.269}$
	sequential	0.520 ± 0.196	0.310 ± 0.178	0.298 ± 0.264	0.379 ± 0.176
	joint	0.579 ± 0.219	$\mathbf{0.356 \pm 0.208}$	$\mathbf{0.336 \pm 0.254}$	0.474 ± 0.222
MSSEG-2-Change inverse	*affine*	0.626 ± 0.224	0.142 ± 0.165	0.081 ± 0.139	$\mathbf{0.633 \pm 0.269}$
	sequential	0.625 ± 0.217	0.348 ± 0.216	0.290 ± 0.244	0.550 ± 0.243
	joint	$\mathbf{0.655 \pm 0.237}$	$\mathbf{0.378 \pm 0.233}$	$\mathbf{0.312 \pm 0.250}$	0.619 ± 0.266

Table 6.3: Résultats calculés sur les ensembles de données LesjakDB et MSSEG-2-Change (mesures au niveau des lésions). La médiane ± l'écart absolu médian (MAD) calculée sur l'ensemble des sujets est rapportée pour chaque métrique. L'expérience inverse MSSEG-2-Change consiste à échanger les images de référence et de suivi pour évaluer la capacité des méthodes à détecter les lésions disparues.

Ensemble de données	Méthode	L-PPV	L-TPR
LesjakDB	*affine*	0.022 ± 0.022	0.576 ± 0.195
	sequential	0.088 ± 0.051	$\mathbf{0.577 \pm 0.197}$
	joint	$\mathbf{0.103 \pm 0.061}$	0.574 ± 0.208
MSSEG-2-Change	*affine*	0.015 ± 0.042	0.840 ± 0.271
	sequential	0.095 ± 0.150	$\mathbf{0.872 \pm 0.307}$
	joint	$\mathbf{0.111 \pm 0.155}$	$\mathbf{0.872 \pm 0.304}$
MSSEG-2-Change inverse	*affine*	0.015 ± 0.042	0.840 ± 0.271
	sequential	$\mathbf{0.094 \pm 0.151}$	$\mathbf{0.977 \pm 0.292}$
	joint	0.091 ± 0.164	0.947 ± 0.304

Le fait que les approches *séquentielle* et *jointe* conduisent à des valeurs de PPV significativement plus élevées par rapport à l'approche *affine* plaide en faveur de l'utilisation du recalage déformable pour réduire le nombre de fausses détections. Le bénéfice de considérer l'approche *jointe* par rapport à l'approche *séquentielle* pour surmonter l'effet de surcompensation des lésions est clairement démontré par les valeurs de TPR et de local DSC significativement plus élevées obtenues avec la méthode *jointe*.

Il est également intéressant de noter que l'effet de surcompensation des lésions n'affecte pas le cas particulier des lésions disparues. En effet, lors du recalage d'une image sans lésion sur une image avec une lésion, la dissimilarité dans la zone de la lésion disparue ne peut pas être corrigée par le recalage (ce n'est le cas que pour les méthodes de recalage non symétriques, voir [Noblet et al., 2004] pour de plus amples explications). Pour illustrer ce phénomène, nous considérons l'expérience inverse MSSEG-2-Change (voir la troisième rangée des Tables 6.2 et 6.3 ainsi que la partie inférieure de la Fig.

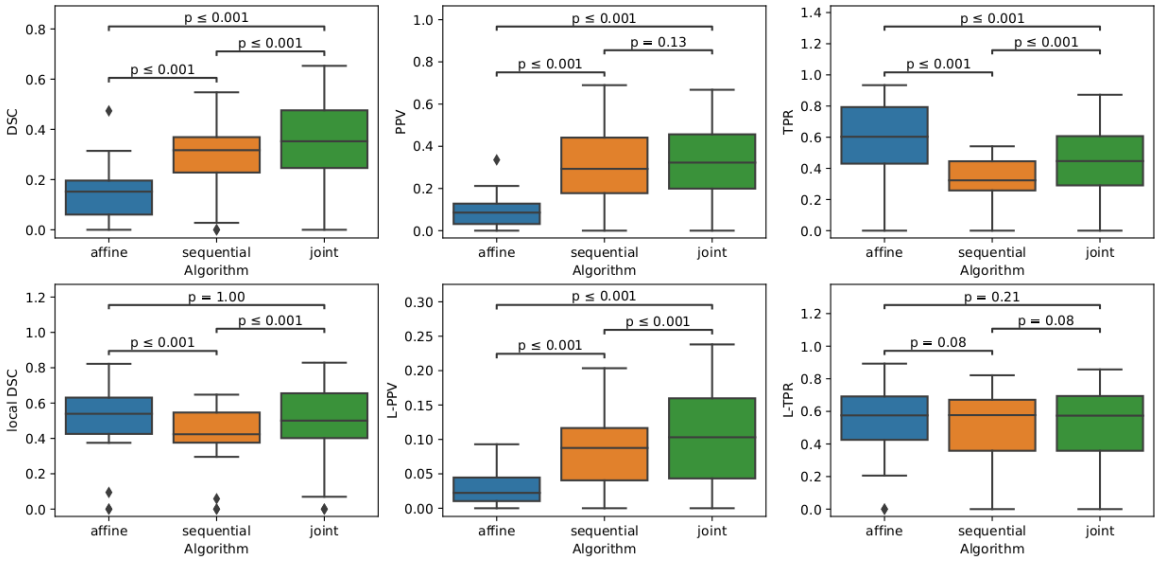


Figure 6.4: Diagrammes en boîte correspondant aux résultats résumés dans les Tableaux 6.2 et 6.3 pour LesjakDB ($N_{\text{Sujet}} = 20$). La signification statistique est évaluée grâce au test de signe de Wilcoxon entre chaque paire de méthodes en appliquant la correction de FDR de Benjamini/Hochberg.

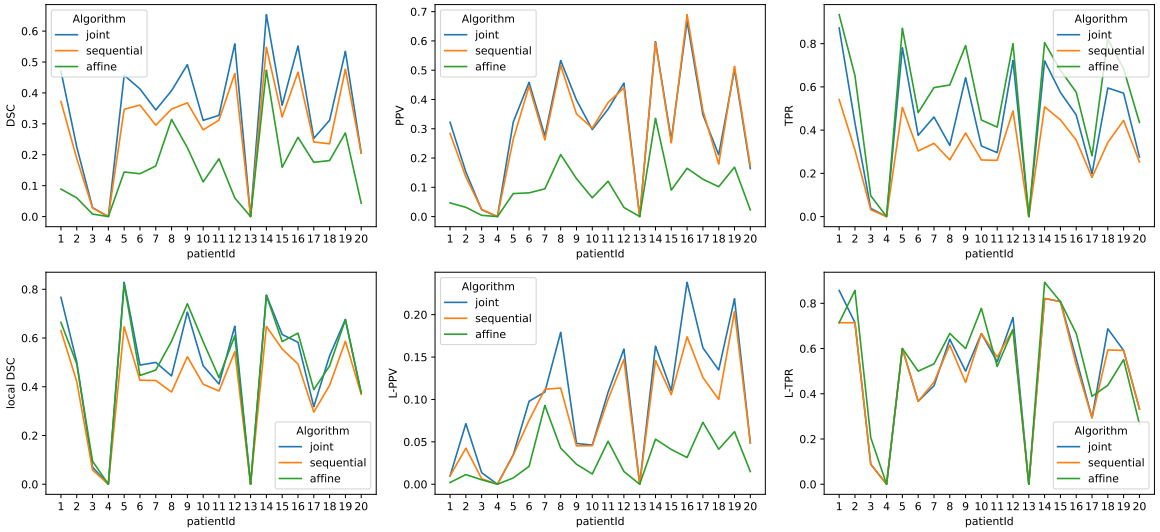


Figure 6.5: Métriques rapportant la performance des trois méthodes pour chaque sujet de LesjakDB.

6.8) qui consiste à échanger l'image de référence et l'image de suivi, de sorte que la vérité terrain correspond désormais à des lésions disparues. La même conclusion peut être tirée du DSC, du PPV et du TPR par rapport à l'expérience MSSEG-2-Change. Le point le plus intéressant concerne le local DSC qui se concentre sur l'évaluation de la

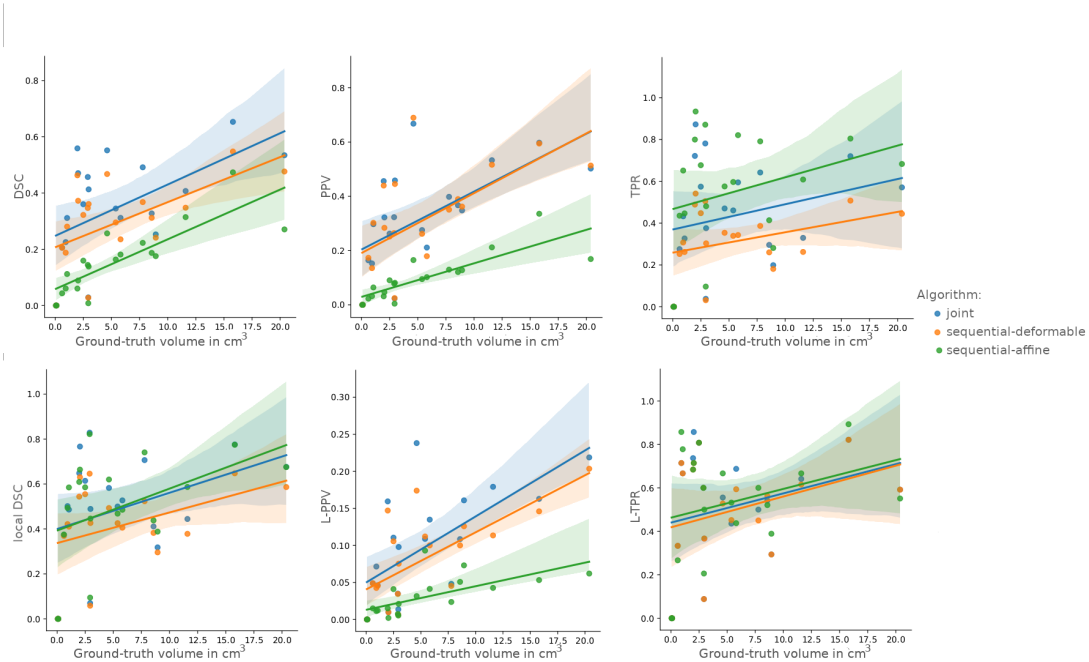


Figure 6.6: Corrélation entre la performance des trois méthodes et la charge lésionnelle évolutive réelle (cm^3) des sujets de LesjakDB.

lésion disparue. Dans ce cas, il n'y a plus de différence significative entre les approches *séquentielle* et *jointe* contrairement à l'expérience MSSEG-2-Change, montrant l'absence d'effet de surcompensation des lésions dans le scénario spécifique de la détection des lésions disparues.

Il convient de noter que tous les résultats présentés dans cette section sont évalués sur les 61 sujets de MSSEG-2-Change (*i.e.,* sujets présentant au moins une nouvelle lésion apparue). En effet, les métriques présentées ne peuvent plus être calculées pour les 39 sujets de MSSEG-2-NoChange, car la carte de détection du changement de vérité terrain est vide. C'est pourquoi nous ne rapportons que le volume des changements détectés pour ce sous-ensemble de MSSEG-2 dans la Fig. 6.9. On peut remarquer que les approches *séquentielle* et *jointe* conduisent à un volume de changements détectés significativement plus faible par rapport à l'approche *affine*, ce qui est en ligne avec les résultats précédents soutenant l'utilisation du recalage déformable pour réduire le nombre de fausses détections. La méthode *jointe* conduit de manière cohérente à des volumes de changements détectés légèrement plus élevés par rapport à la méthode *séquentielle*. Cela est également la conséquence de l'effet de surcompensation des lésions qui affecte l'approche *séquentielle*.

Discussion

Il existe une grande variabilité inter-sujets dans les métriques, ce qui reflète la grande variabilité des cas considérés, notamment en termes de charge lésionnelle et d'atrophie. Dans MSSEG-2, 50

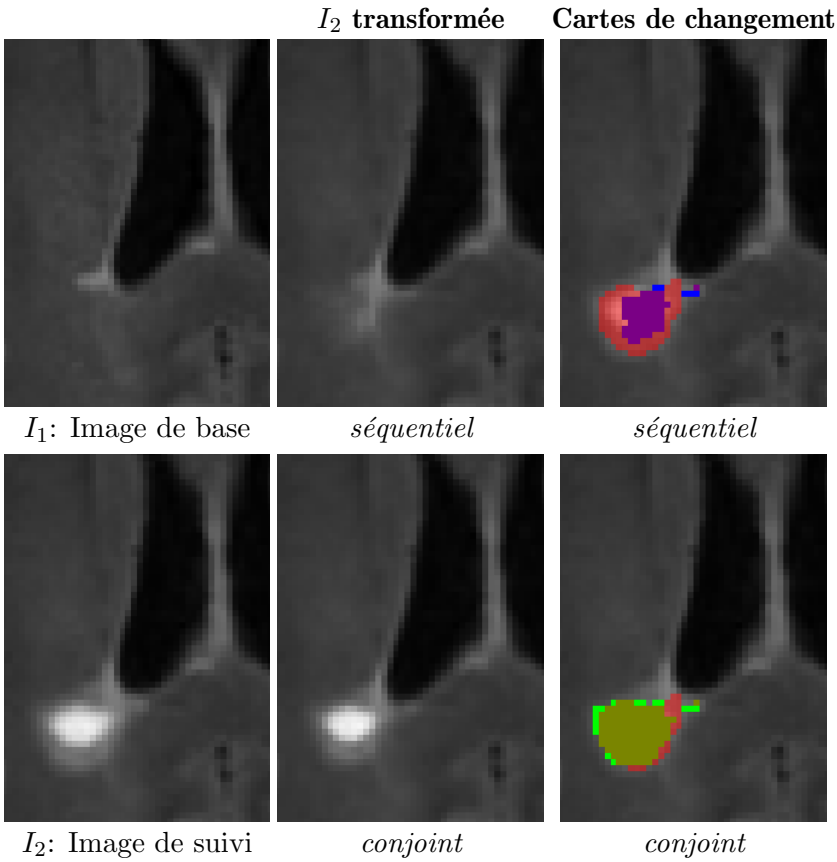
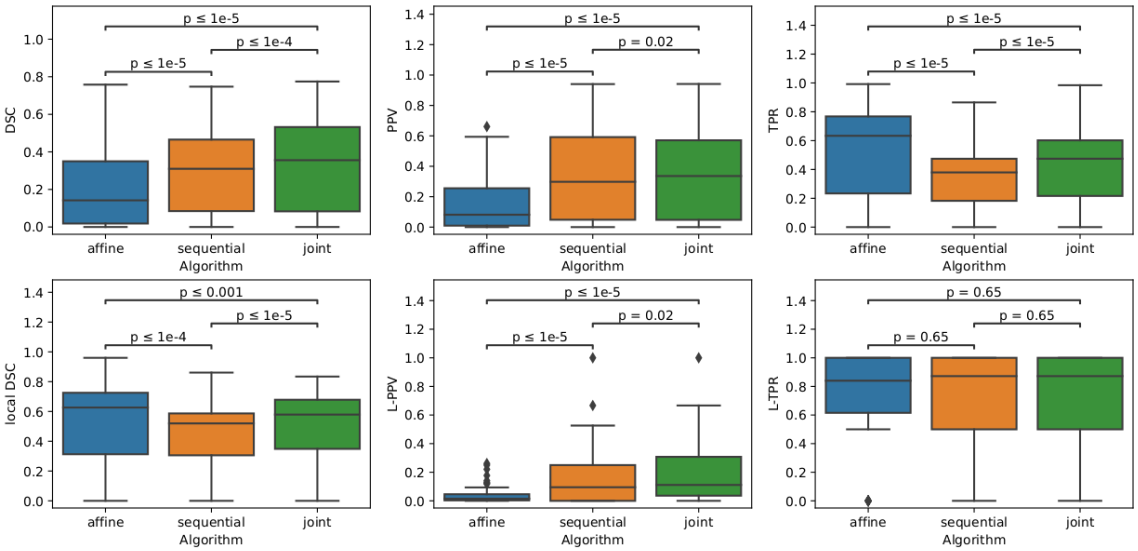


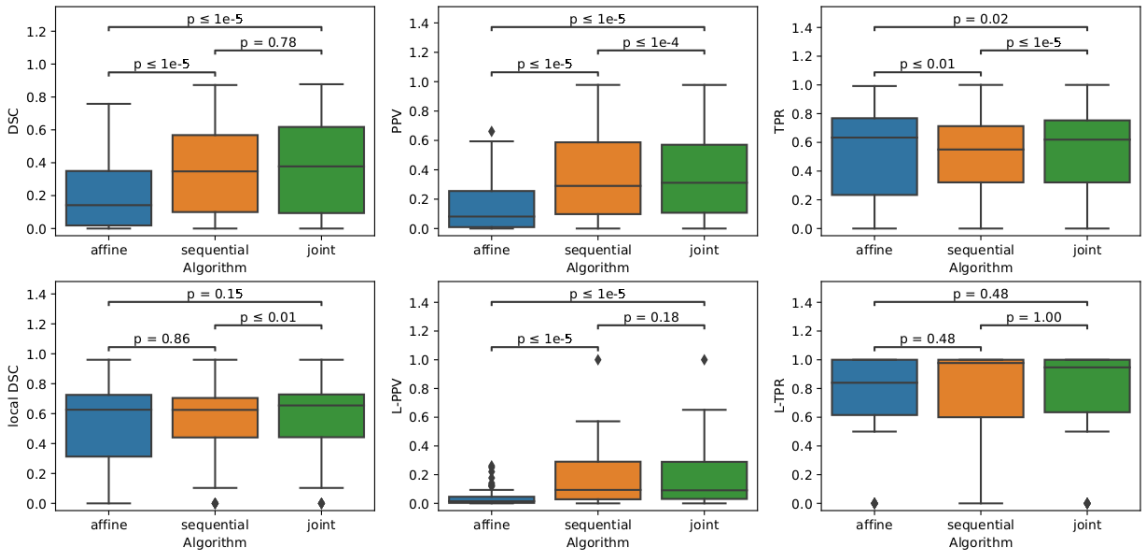
Figure 6.7: Comparaison qualitative des cartes de détection du changement binaire obtenues avec les approches *séquentielle* (violet) et *conjointe* (vert) par rapport à la vérité terrain (en rouge transparent en dessous) sur un sujet sélectionné de l'ensemble de données MSSEG-2.

Influence des hyperparamètres sur les résultats de détection Dans les résultats, nous affirmons que le choix des hyperparamètres, pourvu qu'ils se situent dans une certaine plage de validité, n'affecte pas le classement des résultats ni les propriétés qualitatives des cartes de changement. La figure 6.10 illustre cette affirmation.

Disponibilité et diversité des données et de la vérité terrain La comparaison avec d'autres approches peut être difficile car tous les groupes de recherche utilisent des ensembles de données de taille et de qualité différentes, et le protocole d'annotation manuelle n'est probablement pas normalisé. Il est donc difficile d'évaluer la généralisabilité de chaque approche. Les méthodes d'apprentissage en profondeur commencent seulement à émerger dans le domaine de la détection des changements de lésions de la SEP en raison de ce problème, mais le récent défi MICCAI2021 pour la détection de nouvelles lésions a amélioré la situation en proposant un ensemble de données de 100 patients disponible publiquement. De plus, obtenir des données annotées est très coûteux en temps et en compétences, et les ensembles de données sont souvent relativement petits. Le fait que les vérités terrain soient des images binaires est également une



MSSEG-2-Change



MSSEG-2-Change inverse

Figure 6.8: Boîtes à moustaches correspondant aux résultats résumés dans les Tables 6.2 et 6.3 pour MSSEG-2-Change et MSSEG-2-Change inverse ($N_{Sujet} = 61$). La significativité statistique est évaluée grâce au test des rangs signés de Wilcoxon entre chaque paire de méthodes en appliquant la correction de Benjamini/Hochberg FDR.

limitation, car la délimitation des lésions n'est pas toujours claire.

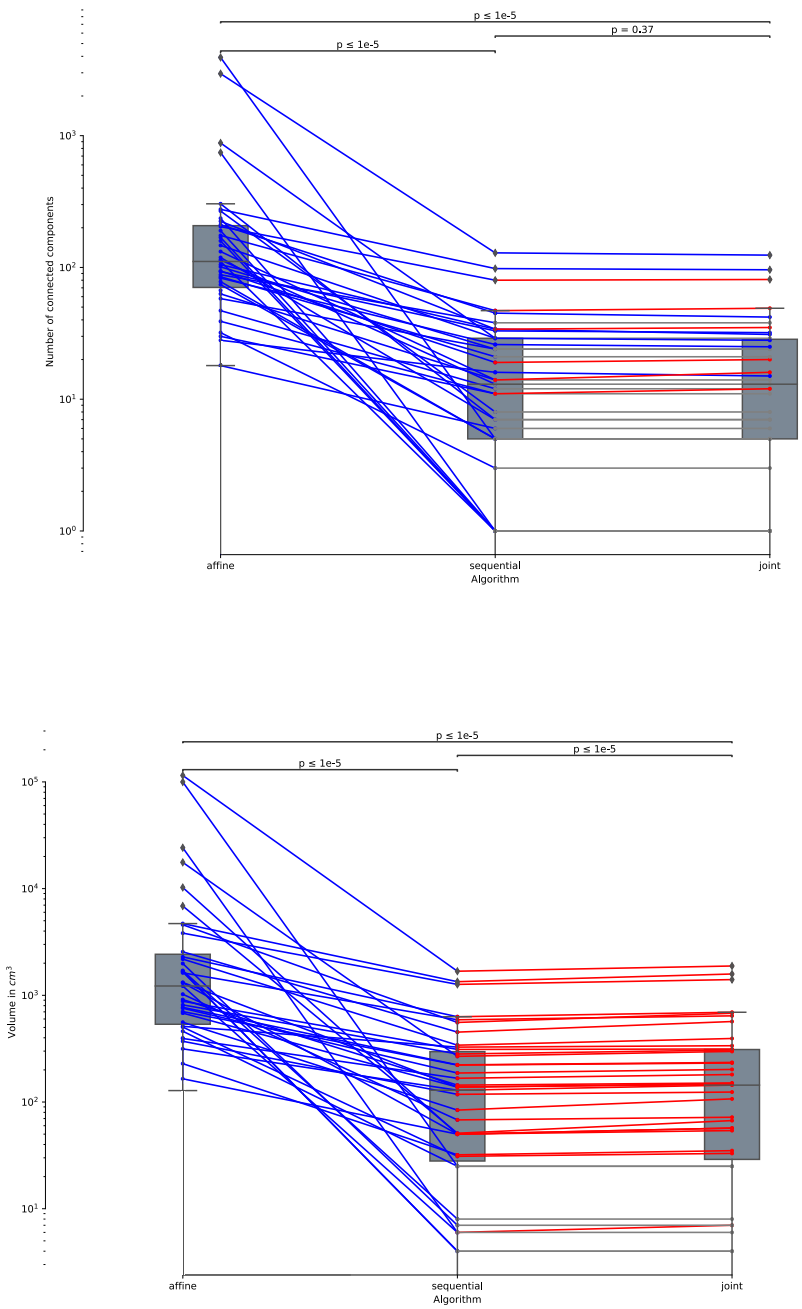


Figure 6.9: Nombre de composants connectés (à gauche) et volume en cm^3 (à droite) des changements détectés calculés pour les 39 sujets de l'ensemble de données MSSEG-2-NoChange (*i.e.*, sujets sans nouvelle lésion apparue). La significativité statistique est évaluée grâce au test des rangs signés de Wilcoxon entre chaque paire de méthodes en appliquant la correction de Benjamini/Hochberg FDR.

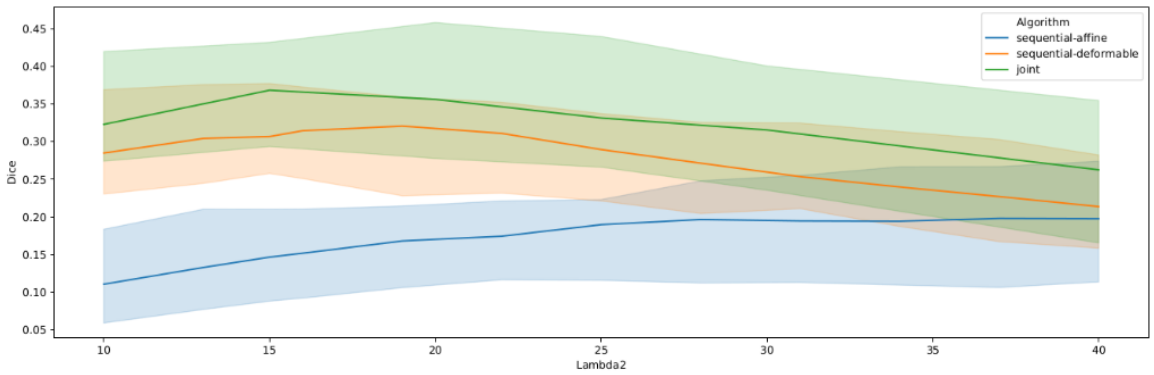


Figure 6.10: Influence du seuil sur la métrique Dice sur l'ensemble de données Lesjak

Conclusion

L'objectif de ce travail de thèse était de proposer une méthode pour détecter les lésions de la sclérose en plaques (SEP) émergentes, disparues, expansives ou rétractiles, tant dans la matière blanche que dans la matière grise. La procédure classique pour détecter les changements dans les lésions de la SEP dans les méthodes basées sur la soustraction et la déformation traite le recalage et la détection des changements de manière séquentielle. Cependant, une détection précise des changements ne peut pas être obtenue sans un recalage approprié, mais la présence de changements dans une image peut entraver le recalage. Nous émettons l'hypothèse que le découplage de ces problèmes est la cause de ces limitations.

Pour résoudre ce problème, nous avons proposé une méthode d'intégration des deux tâches dans un modèle unifié. Notre approche implique de modéliser conjointement ces deux tâches en tant qu'une seule fonction objective. Pour résoudre le problème d'optimisation, nous avons utilisé le cadre de la méthode des multiplicateurs de directions alternées (ADMM) pour le recalage, et la méthode du graph-cut pour la détection des changements. Le cadre ADMM est suffisamment flexible pour gérer différents modèles de recalage avec un faible coût computationnel, ce qui le rend avantageux pour le problème en question.

Nous avons évalué l'approche proposée dans le contexte du suivi des lésions de la SEP, où un recalage déformable est nécessaire pour capturer l'atrophie cérébrale, ce qui peut affecter la performance de la méthode de détection des changements. Les résultats présentés démontrent la performance supérieure du modèle conjoint dans la détection longitudinale des changements, comparativement à ses équivalents séquentiels. Alors que la méthode affine fournissait une bonne couverture des lésions, elle était sujette à des détections de faux positifs en raison de l'atrophie cérébrale. Les méthodes séquentielles et conjointes peuvent corriger les déformations dues à l'atrophie cérébrale, mais la méthode séquentielle échouait à détecter précisément les changements de lésions en raison d'une surcompensation. Cette limitation a été surmontée par l'approche conjointe, qui détectait avec succès des zones de lésions plus étendues. Dans les acquisitions longitudinales sans évolution de la charge lésionnelle, l'approche conjointe

produisait significativement moins de détections que la méthode affine. Ainsi, notre approche conjointe proposée combine la capacité du recalage déformable à corriger l'atrophie cérébrale et la préservation de la forme des lésions pour assurer une détection précise des changements. Cependant, si les lésions sont trop petites et surcompensées lors de la première itération de l'approche conjointe, leur récupération au fil des itérations peut être incomplète, conduisant à des détections moins sensibles.

Bien que l'implémentation actuelle repose sur des hypothèses de modélisation relativement simples, le cadre d'optimisation peut être étendu pour accueillir des modèles plus sophistiqués, incorporant des termes d'adéquation de données alternatifs et des techniques de régularisation. Les extensions possibles du modèle comprennent des termes de données qui présentent une résilience accrue au bruit, l'incorporation d'informations multicanaux, le recalage symétrique ou l'intégration des tâches de normalisation d'intensité dans le modèle conjoint. Les expériences mises en œuvre suggèrent un potentiel d'amélioration de la précision du recalage, ce qui est particulièrement pertinent pour l'estimation de l'atrophie - un domaine non évalué dans cette thèse mais offrant des opportunités de recherche prometteuses pour l'avenir.

Bibliography

U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015. arXiv. doi: 10.48550/ARXIV.1505.04597. URL https://arxiv.org/abs/1505.04597.

Ebru Erbayat Altay, Elizabeth Fisher, Stephen E. Jones, Claire Hara-Cleaver, Jar-Chi Lee, and Richard A. Rudick. Reliability of Classifying Multiple Sclerosis Disease Activity Using Magnetic Resonance Imaging in a Multiple Sclerosis Clinic. *JAMA Neurology*, 70(3):338–344, 03 2013. ISSN 2168-6149. doi: 10.1001/2013.jamaneurol.211. URL https://doi.org/10.1001/2013.jamaneurol.211.

Julia Andresen, Hristina Uzunova, Jan Ehrhardt, Timo Kepp, and Heinz Handels. Image registration and appearance adaptation in non-correspondent image regions for new MS lesions detection. *Frontiers in Neuroscience*, 16, 2022. ISSN 1662-453X. URL https://www.frontiersin.org/articles/10.3389/fnins.2022.981523.

B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2007.06.004. URL https://www.sciencedirect.com/science/article/pii/S1361841507000606. Special Issue on The Third International Workshop on Biomedical Image Registration – WBIR 2006.

Brian B. Avants, Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033–2044, February 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.09.025. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811910012061.

Berke Doga Basaran, Paul M. Matthews, and Wenjia Bai. New lesion segmentation for multiple sclerosis brain images with imaging and lesion-aware augmentation. *Frontiers in Neuroscience*, 16, 2022. ISSN 1662-453X. URL https://www.frontiersin.org/articles/10.3389/fnins.2022.1007453.

Marco Battaglini, Francesca Rossi, Richard A. Grove, Maria Laura Stromillo, Brandon Whitcher, Paul M. Matthews, and Nicola De Stefano. Automated identification of brain new lesions in multiple sclerosis using subtraction images. *Journal of Magnetic Resonance Imaging*, 39(6):1543–1549, 2014. doi: https://doi.org/10.1002/jmri.24293. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.24293.

Elena Bernardis, Kilian M. Pohl, and Christos Davatzikos. Extracting Evolving Pathologies via Spectral Clustering. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, James C. Gee, Sarang Joshi, Kilian M. Pohl, William M. Wells, and Lilla Zöllei, editors, *Information Processing in Medical*

Imaging, volume 7917, pages 680–691. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-38867-5 978-3-642-38868-2. doi: 10.1007/978-3-642-38868-2_57. URL http://link.springer.com/10.1007/978-3-642-38868-2_57.

Ariel Birenbaum and Hayit Greenspan. Multi-view longitudinal CNN for multiple sclerosis lesion segmentation. *Engineering Applications of Artificial Intelligence*, 65: 111–118, October 2017. ISSN 0952-1976. doi: 10.1016/j.engappai.2017.06.006. URL https://www.sciencedirect.com/science/article/pii/S0952197617301252.

Richard Bitar, General Leung, Richard Perng, Sameh Tadros, Alan R. Moody, Josee Sarrazin, Caitlin McGregor, Monique Christakis, Sean Symons, Andrew Nelson, and Timothy P. Roberts. Mr pulse sequences: What every radiologist wants to know but is afraid to ask. *RadioGraphics*, 26(2):513–537, 2006. doi: 10.1148/rg.262055063. URL https://doi.org/10.1148/rg.262055063. PMID: 16549614.

Marcel Bosc, Fabrice Heitz, Jean-Paul Armspach, Izzie Namer, Daniel Gounot, and Lucien Rumbach. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *NeuroImage*, 20(2):643–656, October 2003. ISSN 10538119. doi: 10.1016/S1053-8119(03)00406-3. URL http://linkinghub.elsevier.com/retrieve/pii/S1053811903004063.

Emrah Bostan, Stamatios Lefkimmiatis, Orestis Vardoulis, Nikolaos Stergiopulos, and Michael Unser. Improved variational denoising of flow fields with application to phase-contrast MRI data. *IEEE Signal Processing Letters*, 22(6):762–766, 2014.

Stephen Boyd. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2010. ISSN 1935-8237, 1935-8245. doi: 10.1561/2200000016. URL http://www.nowpublishers.com/article/Details/MAL-016.

Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23 (11):1222–1239, 2001.

Matthew Brett, Alexander P. Leff, Chris Rorden, and John Ashburner. Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. *NeuroImage*, 14(2):486–500, August 2001. ISSN 10538119. doi: 10.1006/nimg.2001.0845. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811901908456.

Wallace J Brownlee, Dan R Altmann, Ferran Prados, Katherine A Miszkiel, Arman Eshaghi, Claudia A M Gandini Wheeler-Kingshott, Frederik Barkhof, and Olga Ciccarelli. Early imaging predictors of long-term outcomes in relapse-onset multiple sclerosis. *Brain*, 142(8):2276–2287, 07 2019. ISSN 0006-8950. doi: 10.1093/brain/awz156. URL https://doi.org/10.1093/brain/awz156.

Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *European conference on computer vision*, pages 25–36. Springer, 2004.

Andrés Bruhn, Joachim Weickert, and Christoph Schnörr. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods. *International Journal of Computer Vision*, 61(3):1–21, February 2005. ISSN 0920-5691. doi: 10.1023/B:VISI.0000045324.43199.43. URL http://link.springer.com/10.1023/B:VISI.0000045324.43199.43.

M. Cabezas, J.F. Corral, A. Oliver, Y. Díez, M. Tintoré, C. Auger, X. Montalban, X. Lladó, D. Pareto, and 'A. Rovira. Improved Automatic Detection of New T2 Lesions in Multiple Sclerosis Using Deformation Fields. *American Journal of Neuroradiology*, 37(10):1816–1823, 2016. ISSN 0195-6108, 1936-959X. doi: 10.3174/ajnr.A4829. URL http://www.ajnr.org/lookup/doi/10.3174/ajnr.A4829.

Mariano Cabezas, Arnau Oliver, Sergi Valverde, Brigitte Beltran, Jordi Freixenet, Joan C. Vilanova, Lluís Ramió-Torrentà, 'Alex Rovira, and Xavier Lladó. BOOST: A supervised approach for multiple sclerosis lesion segmentation. *Journal of Neuroscience Methods*, 237:108–117, November 2014. ISSN 01650270. doi: 10.1016/j.jneumeth.2014.08.024. URL https://linkinghub.elsevier.com/retrieve/pii/S0165027014003161.

Stefano Cerri, Oula Puonti, Dominik S. Meier, Jens Wuerfel, Mark Mühlau, Hartwig R. Siebner, and Koen Van Leemput. A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. *NeuroImage*, 225:117471, 2021. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2020.117471. URL https://www.sciencedirect.com/science/article/pii/S1053811920309563.

Kanglin Chen, Alexander Derksen, Stefan Heldmann, Marc Hallmann, and Benjamin Berkels. Deformable image registration with automatic non-correspondence detection. In Jean-François Aujol, Mila Nikolova, and Nicolas Papadakis, editors, *Scale Space and Variational Methods in Computer Vision*, pages 360–371, Cham, 2015. Springer International Publishing. ISBN 978-3-319-18461-6.

Myra Cheng, Alfiia Galimzianova, **v**Ziga Lesjak, **v**Ziga **v**Spiclin, Christopher B. Lock, and Daniel L. Rubin. A Multi-scale Multiple Sclerosis Lesion Change Detection in a Multi-sequence MRI. In Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel R.S. Tavares, Andrew Bradley, João Paulo Papa, Vasileios Belagiannis, Jacinto C. Nascimento, Zhi Lu, Sailesh Conjeti, Mehdi Moradi, Hayit Greenspan, and Anant Madabhushi, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, volume 11045, pages 353–360. Springer International Publishing, Cham, 2018. ISBN 978-3-030-00888-8 978-3-030-00889-5. doi: 10.1007/978-3-030-00889-5_40. URL http://link.springer.com/10.1007/978-3-030-00889-5_40.

Chris A. Cocosco, Vasken Kollokian, Remi K.-S. Kwan, G. Bruce Pike, and Alan C. Evans. Brainweb: Online interface to a 3D mri simulated brain database. *NeuroImage*, 5:425, 1997.

Olivier Commowick, Frédéric Cervenansky, François Cotton, and Michel Dojat. MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure. *MICCAI-2021*, 13:310–313, 2017. doi: 10.1016/j.dib.2017.05.036.

Olivier Commowick, Audrey Istace, Michaël Kain, Baptiste Laurent, Florent Leray, Mathieu Simon, Sorina Camarasu Pop, Pascal Girard, Roxana Améli, Jean-Christophe Ferré, Anne Kerbrat, Thomas Tourdias, Frédéric Cervenansky, Tristan Glatard, Jérémy Beaumont, Senan Doyle, Florence Forbes, Jesse Knight, April Khademi, Amirreza Mahbod, Chunliang Wang, Richard McKinley, Franca Wagner, John Muschelli, Elizabeth Sweeney, Eloy Roura, Xavier Lladó, Michel M. Santos, Wellington P. Santos, Abel G. Silva-Filho, Xavier Tomas-Fernandez, Hélène Urien, Isabelle Bloch, Sergi Valverde, Mariano Cabezas, Francisco Javier Vera-Olmos, Norberto Malpica, Charles Guttmann, Sandra Vukusic, Gilles Edan, Michel Dojat, Martin Styner, Simon K. Warfield, François Cotton, and Christian Barillot. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. *Scientific Reports*, 8(1), December 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-31911-7. URL http://www.nature.com/articles/s41598-018-31911-7.

Olivier Commowick, Frédéric Cervenansky, François Cotton, and Michel Dojat. MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure. In *MICCAI 2021 - 24th International Conference on Medical Image Computing and Computer Assisted Intervention*, page 126, Strasbourg, France, September 2021. URL https://hal.inria.fr/hal-03358968.

Stefan Denner, Ashkan Khakzar, Moiz Sajid, Mahdi Saleh, Ziga Spiclin, Seong Tae Kim, and Nassir Navab. Spatio-temporal learning from longitudinal data for multiple sclerosis lesion segmentation. 2020. doi: 10.48550/ARXIV.2004.03675. URL https://arxiv.org/abs/2004.03675.

Yago Diez, Arnau Oliver, Mariano Cabezas, Sergi Valverde, Robert Martí, Joan Carles Vilanova, Lluís Ramió-Torrentà, 'Alex Rovira, and Xavier Lladó. Intensity Based Methods for Brain MRI Longitudinal Registration. A Study on Multiple Sclerosis Patients. *Neuroinformatics*, 12(3):365–379, July 2014. ISSN 1539-2791, 1559-0089. doi: 10.1007/s12021-013-9216-z. URL http://link.springer.com/10.1007/s12021-013-9216-z.

E. Dufresne, D. Fortun, B. Kumar, S. Kremer, and V. Noblet. Joint registration and change detection in longitudinal brain mri. In *ISBI*, Apr 2020. doi: 10.1109/ISBI45749.2020.9098626. URL http://publis.icube.unistra.fr/4-DFKK20.

E. Dufresne, D. Fortun, S. Kremer, and V. Noblet. A unified framework for focal intensity change detection and deformable image registration. application to

the monitoring of multiple sclerosis lesions in longitudinal 3d brain mri. *Frontiers in Neuroimaging*, 1:1–15, 2022a. doi: 10.3389/fnimg.2022.1008128. URL `http://publis.icube.unistra.fr/2-DFKN22`.

E. Dufresne, D. Fortun, S. Kremer, and V. Noblet. Approche conjointe de recalage et détection de changement appliquée à l'analyse d'irm cérébrales longitudinales. In *GRETSI*, Jan 2022b. URL `http://publis.icube.unistra.fr/5-DFKN22`.

Colm Elliott, Douglas L. Arnold, D. Louis Collins, and Tal Arbel. Temporally Consistent Probabilistic Detection of New Multiple Sclerosis Lesions in Brain MRI. *IEEE Transactions on Medical Imaging*, 32(8):1490–1503, August 2013. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2013.2258403. URL `http://ieeexplore.ieee.org/document/6502722/`.

Denis Fortun, Patrick Bouthemy, and Charles Kervrann. Aggregation of local parametric candidates with exemplar-based occlusion handling for optical flow. *Computer Vision and Image Understanding*, 145:81–94, 2016. ISSN 1077-3142. doi: https://doi.org/10.1016/j.cviu.2015.11.020. URL `https://www.sciencedirect.com/science/article/pii/S1077314215002660`. Light Field for Computer Vision.

Denis Fortun, Martin Storath, Dennis Rickert, Andreas Weinmann, and Michael Unser. Fast piecewise-affine motion estimation without segmentation. *IEEE Transactions on Image Processing*, 27(11):5612–5624, 2018.

Onur Ganiler, Arnau Oliver, Yago Diez, Jordi Freixenet, Joan C. Vilanova, Brigitte Beltran, Llu'is Rami'o-Torrent'a, 'Alex Rovira, and Xavier Llad'o. A subtraction pipeline for automatic detection of new appearing multiple sclerosis lesions in longitudinal studies. *Neuroradiology*, 56(5):363–374, May 2014. ISSN 0028-3940, 1432-1920. doi: 10.1007/s00234-014-1343-1. URL `http://link.springer.com/10.1007/s00234-014-1343-1`.

Ali Gooya, Kilian M. Pohl, Michel Bilello, George Biros, and Christos Davatzikos. Joint Segmentation and Deformable Registration of Brain Scans Guided by a Tumor Growth Model. In Gabor Fichtinger, Anne Martel, and Terry Peters, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, volume 6892, pages 532–540. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-23628-0 978-3-642-23629-7. doi: 10.1007/978-3-642-23629-7_65. URL `http://link.springer.com/10.1007/978-3-642-23629-7_65`. Series Title: Lecture Notes in Computer Science.

Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in medicine & biology*, 46(3):R1, 2001.

Sebastian Hitziger, Wen Xin Ling, Thomas Fritz, Tiziano D'Albis, Andreas Lemke, and Joana Grilo. Triplanar U-Net with lesion-wise voting for the segmentation of new lesions on longitudinal MRI stud-

ies. *Frontiers in Neuroscience*, 16, 2022. ISSN 1662-453X. URL https://www.frontiersin.org/articles/10.3389/fnins.2022.964250.

David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Nicha Chitphakdithai, and James S. Duncan. Non-rigid Registration with Missing Correspondences in Preoperative and Postresection Brain Images. In Tianzi Jiang, Nassir Navab, Josien P. W. Pluim, and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, volume 6361, pages 367–374. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-15704-2 978-3-642-15705-9. doi: 10.1007/978-3-642-15705-9_45. URL http://link.springer.com/10.1007/978-3-642-15705-9_45. Series Title: Lecture Notes in Computer Science.

J. E. Iglesias, Cheng-Yi Liu, P. M. Thompson, and Zhuowen Tu. Robust Brain Extraction Across Datasets and Comparison With Publicly Available Methods. *IEEE Trans. Med. Imaging*, 30(9):1617–1634, September 2011. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2011.2138152. URL http://ieeexplore.ieee.org/document/5742706/.

Fabian Isensee, Jens Petersen, André Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Köhler, Tobias Norajitra, Sebastian J. Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *CoRR*, abs/1809.10486, 2018. URL http://arxiv.org/abs/1809.10486.

Saurabh Jain, Annemie Ribbens, Diana M. Sima, Melissa Cambron, Jacques De Keyser, Chenyu Wang, Michael H. Barnett, Sabine Van Huffel, Frederik Maes, and Dirk Smeets. Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework. *Frontiers in Neuroscience*, 10, December 2016. ISSN 1662-453X. doi: 10.3389/fnins.2016.00576. URL http://journal.frontiersin.org/article/10.3389/fnins.2016.00576/full.

Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *NeuroImage*, 62(2):782–790, August 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.09.015. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811911010603.

Reda Abdellah Kamraoui, Boris Mansencal, José V. Manjon, and Pierrick Coupé. Longitudinal detection of new MS lesions using deep learning. *Frontiers in Neuroimaging*, 1, 2022. ISSN 2813-1193. URL https://www.frontiersin.org/articles/10.3389/fnimg.2022.948235.

Ulrike W. Kaunzner and Susan A. Gauthier. Mri in the assessment and monitoring of multiple sclerosis: an update on best practice. *Therapeutic Advances in Neu-*

rological Disorders, 10(6):247–261, 2017. doi: 10.1177/1756285617708911. URL https://doi.org/10.1177/1756285617708911. PMID: 28607577.

Arno Klein, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, D. Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P. Woods, J. John Mann, and Ramin V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, July 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.12.037. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811908012974.

Dongjin Kwon, Marc Niethammer, Hamed Akbari, Michel Bilello, Christos Davatzikos, and Kilian M. Pohl. PORTR: Pre-Operative and Post-Recurrence Brain Tumor Registration. *IEEE Transactions on Medical Imaging*, 33(3):651–667, March 2014. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2013.2293478. URL https://ieeexplore.ieee.org/document/6678314/.

Pamela J. LaMontagne, Tammie LS. Benzinger, John C. Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G. Vlassenko, Marcus E. Raichle, Carlos Cruchaga, and Daniel Marcus. OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. 2019. doi: 10.1101/2019.12.13.19014902. URL https://www.medrxiv.org/content/early/2019/12/15/2019.12.13.19014902. Publisher: Cold Spring Harbor Laboratory Press _eprint: https://www.medrxiv.org/content/early/2019/12/15/2019.12.13.19014902.full.pdf.

Christian Thode Larsen, J Eugenio Iglesias, and Koen Van Leemput. N3 bias field correction explained as a bayesian modeling method. In *Bayesian and grAphical Models for Biomedical Imaging: First International Workshop, BAMBI 2014, Cambridge, MA, USA, September 18, 2014, Revised Selected Papers*, pages 1–12. Springer, 2014.

R Hc Lazeron, J B Boringa, M Schouten, B Mj Uitdehaag, E Bergers, J Lindeboom, M J Eikelenboom, P H Scheltens, F Barkhof, and C H Polman. Brain atrophy and lesion load as explaining parameters for cognitive impairment in multiple sclerosis. *Multiple Sclerosis Journal*, 11(5):524–531, October 2005. ISSN 1352-4585, 1477-0970. doi: 10.1191/1352458505ms1201oa. URL http://journals.sagepub.com/doi/10.1191/1352458505ms1201oa.

vZiga Lesjak, Franjo Pernu**v**s, Bo**v**stjan Likar, and **v**Ziga **v**Spiclin. Validation of White-Matter Lesion Change Detection Methods on a Novel Publicly Available MRI Image Database. *Neuroinformatics*, 14(4):403–420, October 2016. ISSN 1539-2791, 1559-0089. doi: 10.1007/s12021-016-9301-1. URL http://link.springer.com/10.1007/s12021-016-9301-1.

E. B. Lewis and N. C. Fox. Correction of differential intensity inhomogeneity in longitudinal MR images. *Neuroimage*, 23(1):75–83, 2004.

Xavier Lladó, Onur Ganiler, Arnau Oliver, Robert Martí, Jordi Freixenet, Laia Valls, Joan C. Vilanova, Lluís Ramió-Torrentà, and 'Alex Rovira. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology*, 54(8):787–807, August 2012. ISSN 0028-3940, 1432-1920. doi: 10.1007/s00234-011-0992-6. URL http://link.springer.com/10.1007/s00234-011-0992-6.

Chao Lu, Sudhakar Chelikani, and James S. Duncan. A Unified Framework for Joint Segmentation, Nonrigid Registration and Tumor Detection: Application to MR-Guided Radiotherapy. In Gábor Székely and Horst K. Hahn, editors, *Information Processing in Medical Imaging*, volume 6801, pages 525–537. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-22091-3 978-3-642-22092-0. doi: 10.1007/978-3-642-22092-0_43. URL http://link.springer.com/10.1007/978-3-642-22092-0_43. Series Title: Lecture Notes in Computer Science.

C. McNamara, G. Sugrue, B. Murray, and P.J. MacMahon. Current and emerging therapies in multiple sclerosis: Implications for the radiologist, part 1—mechanisms, efficacy, and safety. *American Journal of Neuroradiology*, 38(9):1664–1671, 2017. ISSN 0195-6108. doi: 10.3174/ajnr.A5147. URL http://www.ajnr.org/content/38/9/1664.

V. Noblet, C. Heinrich, F. Heitz, and J-P. Armspach. A topology preserving non-rigid registration method using a symmetric similarity function - application to 3-D brain images. In *European Conference on Computer Vision (ECCV)*, pages 546–557, Prague, Czech Republic, 2004. doi: 10.1007/b97871. URL http://publis.icube.unistra.fr/4-NHHA04.

V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armspach. 3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. *IEEE Transactions on Image Processing*, 14 (5):553–566, May 2005. ISSN 1057-7149. doi: 10.1109/TIP.2005.846026. URL http://ieeexplore.ieee.org/document/1420387/.

Vincent Noblet, Christian Heinrich, Fabrice Heitz, and Jean-Paul Armspach. An efficient incremental strategy for constrained groupwise registration based on symmetric pairwise registration. *Pattern Recognition Letters*, 33(3):283–290, February 2012. ISSN 01678655. doi: 10.1016/j.patrec.2011.11.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0167865511003801.

LG Nyúl, JK Udupa, and X Zhang. New variants of a method of MRI scale standardization. *IEEE transactions on medical imaging*, 19(2):143—150, February 2000. ISSN 0278-0062. doi: 10.1109/42.836373. URL https://doi.org/10.1109/42.836373.

Sarah Parisot, Hugues Duffau, Stéphane Chemouny, and Nikos Paragios. Joint tumor segmentation and dense deformable registration of brain MR images. In Nicholas Ayache, Hervé Delingette, Polina Golland, and Kensaku Mori, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, volume 7511, pages 651–658. Springer Berlin Heidelberg. ISBN 978-3-642-33417-7 978-3-642-33418-4. doi:

10.1007/978-3-642-33418-4_80. URL http://link.springer.com/10.1007/978-3-642-33418-4_80. Series Title: Lecture Notes in Computer Science.

P. Pieperhoff, M. Südmeyer, L. Hömke, K. Zilles, A. Schnitzler, and K. Amunts. Detection of structural changes of the human brain in longitudinally acquired MR images by deformation field morphometry: Methodological analysis, validation and application. *NeuroImage*, 43(2):269–287, November 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.07.031. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811908008367.

Andrew J. Plassard, L. Taylor Davis, Allen T. Newton, Susan M. Resnick, Bennett A. Landman, and Camilo Bermudez. Learning implicit brain MRI manifolds with deep learning. In Elsa D. Angelini and Bennett A. Landman, editors, *Medical Imaging 2018: Image Processing*, page 56, Houston, United States, March 2018. SPIE. ISBN 978-1-5106-1637-0 978-1-5106-1638-7. doi: 10.1117/12.2293515. URL https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10574/2293515/Learning-implicit-brain-MRI-manifolds-with-deep-learning/10.1117/12.2293515.full.

D. Rey, G. Subsol, H. Delingette, and N. Ayache. Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. *Med Image Anal*, 6(2):163–179, Jun 2002.

Petter Risholm, Eigil Samsett, Ion-Florin Talos, and William Wells. A non-rigid registration framework that accommodates resection and retraction. *Information Processing in Medical Imaging: Proceedings of the ... Conference*, 21:447–458, 2009. ISSN 1011-2499. doi: 10.1007/978-3-642-02498-6_37.

Mostafa Salem, Mariano Cabezas, Sergi Valverde, Deborah Pareto, Arnau Oliver, Joaquim Salvi, 'Alex Rovira, and Xavier Lladó. A supervised framework with intensity subtraction and deformation field features for the detection of new T2-w lesions in multiple sclerosis. *NeuroImage: Clinical*, 17: 607–615, 2018. ISSN 22131582. doi: 10.1016/j.nicl.2017.11.015. URL https://linkinghub.elsevier.com/retrieve/pii/S2213158217302954.

Mostafa Salem, Sergi Valverde, Mariano Cabezas, Deborah Pareto, Arnau Oliver, Joaquim Salvi, 'Alex Rovira, and Xavier Llad'o. A fully convolutional neural network for new T2-w lesion detection in multiple sclerosis. *NeuroImage: Clinical*, 25: 102149, 2020. ISSN 2213-1582. doi: https://doi.org/10.1016/j.nicl.2019.102149. URL https://www.sciencedirect.com/science/article/pii/S2213158219304954.

Paul Schmidt, Viola Pongratz, Pascal Küster, Dominik Meier, Jens Wuerfel, Carsten Lukas, Barbara Bellenberg, Frauke Zipp, Sergiu Groppa, Philipp G. Sämann, Frank Weber, Christian Gaser, Thomas Franke, Matthias Bussas, Jan Kirschke, Claus Zimmer, Bernhard Hemmer, and Mark Mühlau. Automated segmentation of changes in FLAIR-hyperintense white matter lesions in

multiple sclerosis on serial magnetic resonance imaging. *NeuroImage: Clinical*, 23:101849, 2019. ISSN 22131582. doi: 10.1016/j.nicl.2019.101849. URL https://linkinghub.elsevier.com/retrieve/pii/S2213158219301998.

Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E.J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, Marilena De Luca, Ivana Drobnjak, David E. Flitney, Rami K. Niazy, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, J. Michael Brady, and Paul M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23:S208–S219, January 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.07.051. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811904003933.

A. Sotiras, C. Davatzikos, and N. Paragios. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013. doi: 10.1109/TMI.2013.2265603.

Radu Stefanescu, Olivier Commowick, Grégoire Malandain, Pierre-Yves Bondiau, Nicholas Ayache, and Xavier Pennec. Non-rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Christian Barillot, David R. Haynor, and Pierre Hellier, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, volume 3216, pages 704–711. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-22976-6 978-3-540-30135-6. doi: 10.1007/978-3-540-30135-6_86. URL http://link.springer.com/10.1007/978-3-540-30135-6_86.

E.M. Sweeney, R.T. Shinohara, C.D. Shea, D.S. Reich, and C.M. Crainiceanu. Automatic Lesion Incidence Estimation and Detection in Multiple Sclerosis Using Multisequence Longitudinal MRI. *American Journal of Neuroradiology*, 34(1):68–73, January 2013. ISSN 0195-6108, 1936-959X. doi: 10.3174/ajnr.A3172. URL http://www.ajnr.org/lookup/doi/10.3174/ajnr.A3172.

Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, and James C. Gee. N4ITK: Improved N3 bias correction with robust B-spline approximation. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 708–711, Rotterdam, Netherlands, 2010. IEEE. ISBN 978-1-4244-4125-9. doi: 10.1109/ISBI.2010.5490078. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5490078.

Liliana Valencia, Albert Clèrigues, Sergi Valverde, Mostafa Salem, Arnau Oliver, 'Alex Rovira, and Xavier Llad'o. Evaluating the use of synthetic T1-w images in new T2 lesion detection in multiple sclerosis. *Frontiers in Neuroscience*, 16, 2022. ISSN 1662-453X. URL https://www.frontiersin.org/articles/10.3389/fnins.2022.954662.

Kicky G. van Leeuwen, Steven Schalekamp, Matthieu J. C. M. Rutten, Bram van Ginneken, and Maarten de Rooij. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European Radiology*, 31(6):3797–3804, June 2021. ISSN 0938-7994, 1432-1084. doi: 10.1007/s00330-021-07892-z. URL https://link.springer.com/10.1007/s00330-021-07892-z.

Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45 (1):S61–S72, 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.10.040. URL http://linkinghub.elsevier.com/retrieve/pii/S1053811908011683.

Christoph Vogel, Stefan Roth, and Konrad Schindler. An evaluation of data costs for optical flow. In *DAGM Symposium on Pattern Recognition*, pages 343–353, 2013.

M. P. Wattjes, M. D. Steenwijk, and M. Stangel. MRI in the Diagnosis and Monitoring of Multiple Sclerosis: An Update. *Clinical Neuroradiology*, 25(2): 157–165, October 2015. ISSN 1869-1447. doi: 10.1007/s00062-015-0430-y. URL https://doi.org/10.1007/s00062-015-0430-y.

A. Yezzi, L. Zöllei, and T. Kapur. A variational framework for integrating segmentation and registration through active contours. *Medical Image Analysis*, 7(2): 171–185, June 2003. ISSN 1361-8415. doi: 10.1016/S1361-8415(03)00004-5. URL https://www.sciencedirect.com/science/article/pii/S1361841503000045.

C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv-l1 optical flow. In Fred A. Hamprecht, Christoph Schnörr, and Bernd Jähne, editors, *Pattern Recognition*, pages 214–223, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-74936-3.