A Generative Approach to Simultaneous Diffeomorphic Registration and Lesion Segmentation of Neuroimages

Loïc Muhirwa

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science Mathematics and Statistics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The M.Sc. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

Image segmentation and image registration are two fundamental problems in computer vision and medical image processing. In image segmentation, one seeks to partition an image into meaningful segments by assigning a label to each pixel indicating which segment it belongs to. In image registration, one seeks to recover a spatial transformation that geometrically aligns two or more images, which allows downstream image analyses in which the registered images share a coordinate system. Image processing pipelines typically apply these procedures sequentially even though the segmentation of an image could improve its registration and registration of an image could improve its segmentation. With an appropriate parametrization, one can view these two tasks as an inference problem in which the spatial transformation and segmentation are latent variables. In this work, registration and segmentation are integrated through a hierarchical Bayesian generative framework. The framework models the data generating process of a set of magnetic resonance (MR) images of ischemic stroke lesioned brains. Under this framework, we simultaneously estimate a lesion tissue segmentation along with a spatial diffeomorphic transformation that maps a subject image into spatial correspondence with a healthy template image. The framework is evaluated on two-dimensional images both real and synthetic. Experimental results on real MR images show that simultaneous segmentation and registration can significantly improve the accuracy of lesion segmentation as well as the accuracy of registration near the lesion.

Dedications

I dedicate this work to my best friends

Ted Thompson
David Mammarella
Vanessa Susevski

Your friendships have been the greatest joy of my life

Acknowledgement

To my advisor, Dr. Tanya Schmah, this work is in no small part thanks to you. Thank you for your insightful guidance and your tireless support. Whether it was night or day, week or weekend, you were always available to prepare me for the rigours of this process.

To my first supervisor and friend, Jessica Andrews, thank you for championing academic excellence in your team and allowing an inexperienced 22-year-old to pursue graduate studies while simultaneously working for you.

To my parents, Sévérien and Landrada, thank you for leaving your native country and everything you have ever known with nothing to your name but hopes for a better life for my siblings and me. I owe you everything I am and everything I will ever be.

To my partner, Vanessa Susevski, it's not lost on me that so much success in my life is thanks to so much sacrifice in yours. From my first day of high school to my last day of graduate school, you have been in every way all that anyone could be.

To the spirits of Thomas Bayes, Carl Friedrich Gauss, Jacob Bernoulli and Pierre-Simon marquis de Laplace, today I stand on your shoulders.

I will end with a quote by Stefan Banach. *"Mathematics is the most beautiful and most powerful creation of the human spirit."*

Contents

List of Figures

List of Tables

List of Algorithms

CHAPTER 1

Introduction

Computer vision and image analysis techniques are widely used in medical imaging applications such as lesion detection, diagnosis, health monitoring and prognosis prediction. These applications are typically implemented through data pipelines in which image processing techniques are sequentially applied to a set of images. Image segmentation and image registration are such processes.

Image segmentation is the process of partitioning an image into meaningful segments. This process is ubiquitous in image processing and analysis since many applications require substructures of digital images to be identified and annotated, making it a fundamental preprocessing procedure. A canonical example of image segmentation is binary segmentation between the foreground and background of an image; it's easy to imagine how broadly applicable such a segmentation would be. In medical imaging, one might need to segment magnetic resonance (MR) or computerized tomography (CT) images of an organ into distinct anatomical structures or segment different tissue types. In neuroimaging, in particular, one might be interested in segmenting the human brain into major tissue (white and grey matter) or segmenting between the health status of a tissue (healthy and lesioned). This work falls in the latter case and considers the automatic segmentation of ischemic stroke lesions in human brains from MR images.

Image registration is the process of superimposing two or more images in order to geometrically align some substructures of interest. Many image processing and image analysis require that the images in question have some level of spatial correspondence, making image registration a fundamental task. In multi-temporal image analysis, for instance, a set of images at different time intervals could be used to detect motion or tumour growth. These images would need to have a spatial correspondence otherwise the image misalignments could be interpreted as the motion of image contents. Another common application is in multi-modal image analysis; in oncologic imaging, it is common to acquire multiple images with different modalities—MR and CT—of the same organ from the same patient, under this scenario the images would need to be geometrically aligned before any downstream image analysis. Even in conditions where images are taken with the same sensor (mono-modal), image registration could still be a fundamental task. Consider a functional imaging study in which a neuroscientist studies brain regions' response (neural activation sites) to some stimuli. The researcher would first need to have the neuroimages of their subject images in spatial correspondence otherwise image misalignment could introduce some level of ambiguity between the activation sites across subject images. Even under conditions in which the image capturing apparatus is fixed, such as taking images with the same camera and conditions in which the contents of the images are fixed, such as taking images of immobile objects like mountains and a stabilized camera on a tripod, some post-processing image registration might still be needed. Both the image capturing method and the conditions under which the image is captured could create spatial artefacts that geometrically corrupt the substructures of an image [57]. In such scenarios, image regis-

tration would correct these spatial artefacts by mapping the corrupted images to a shared coordinate space.

Before registering images, one must consider the class of spatial transformations that would sufficiently account for the geometric misalignment among images. One can often find a correspondence between the physical properties of the objects within the images and the class of transformations needed to align these objects. When doing mono-modal intra-subject registration of images containing rigid bodies, such as bone tissue of the same subject, a class of global linear transformations (translation, scaling, rotation and shearing) is sufficient. Registration problems under this class of transformations are sometimes called affine or linear registration. Using the terms linear and affine interchangeably could potentially lead to some confusion since is not strictly true in an algebraic sense, however this thesis will use the terms synonymously as this convention is consistent with the registration literature.

On the other hand, inter-subject registration of images of non-rigid bodies, such as the cortical lobes of a human brain of the same subject, a class of non-linear transformations is necessary. More generally, non-linear transformations are necessary for inter-subject registration to account for the non-linear anatomic variability across subjects. In this thesis, the term *deformation* refers to an invertible transformation that is possibly non-affine. Registration problems or algorithms in which the class of admissible transformations are not all affine are called deformable or non-linear registration. Unlike linear transformations that are globally specified over an entire image, deformations can model spatially varying local non-linear deformations. Many applications use a combination of linear and non-linear registration, typically starting with a linear registration to account for global geometric properties and then subsequently applying non-linear registration to account for local geometric properties. In neuroimaging, the head or skull is typically linearly registered as a preprocessing step to non-linear registration; this is the approach the proposed framework uses. Given how vast the class of deformations is, we restrict the class by imposing certain desirable smoothness properties, one of which is topological preservation. This means we want the deformation to preserve the topology of the objects being registered by not allowing folding, tearing or pinching.

In this work, we consider the relationship between segmentation and registration and we unify both processes in a joint generative model and an iterative parameter estimation scheme that alternates between improving segmentation and improving registration, i.e. block coordinate descent. This is because the segmentation of an image can enhance the registration and the registration of an image can enhance the segmentation [5, 10]. To understand why, consider the registration of two brain images given that we know their tissue segmentations. The additional information from the segmentation would allow the construction of a registration performance metric that not only promotes the alignment of anatomic structures but also the alignment of the two segmentations. Conversely, con-

sider the lesion segmentation of a brain image, that has been pre-registered to a healthy brain. Since the lesion tissue of the first image would be superimposed on the healthy tissue of the second image, the tissue dissimilarity could be used as a lesion detection criterion [12].

To formalize these ideas, we need a mathematical representation of an image. There are many ways to represent images, depending on the applications, some are more convenient than others. This document presents many approaches to segmentation and registration making a single mathematical representation of an image difficult. Nevertheless, a primary representation will be adopted, which we will deviate away from a few times in the document for notational convenience or completeness in the cases where an image is understood as a discrete object as opposed to a continuous one. Formally, an image can be represented as a map from some image domain to an intensity domain as follows

$$I : \Omega \to \mathbb{R} \tag{1.0.1}$$

where the image domain Ω is a compact and simply connected subset of $\mathbb{R}^d$ for $d \in \{2, 3\}$ for 2D or 3D images also known as *volumes*. This definition allows us to alter quantities defined on Ω by geometrically transforming Ω. Moreover, without loss of generality, this document only considers images with unidimensional intensity values, i.e. grayscale images; this is implicit in the definition since the image maps to $\mathbb{R}$. It is worth noting that this is not a simplifying assumption since MR and CT images are naturally grayscaled.

The rest of this thesis is organized as follows. In chapters 2 and 3 we give an overview of the general theory of image segmentation and image registration, respectively. In chapter 4 we introduce the general theory of Markov random fields, which we initially considered as a foundation for our approach and believe is a promising extension of the proposed framework. In chapter 5 we propose a generative framework that integrates image registration and image segmentation along with inference algorithms to infer the model's latent variables and learn the model parameters and we present experimental results in which the framework was applied to synthetic image data. In chapter 6, we present some applications and experiments of our proposed algorithm applied to a medical neuroimaging ischemic stroke lesion dataset.

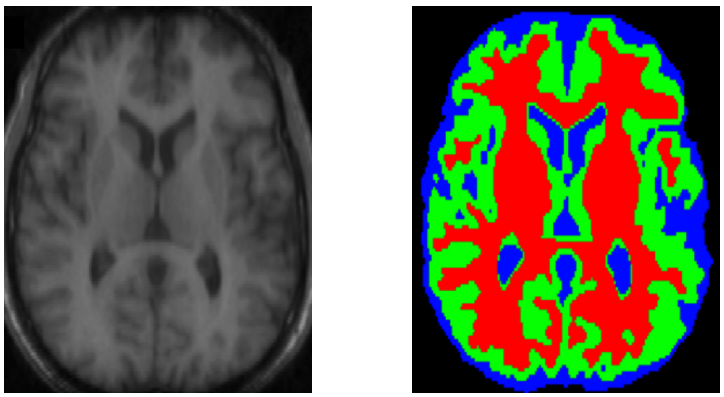
CHAPTER 2

Image Segmentation

Digital image segmentation can be broadly classified into two categories, manual segmentation and automated segmentation. In manual segmentation, a human manually annotates the different segments in a digital image, whereas in automated segmentation an automated algorithm segments the image. Manual segmentation comes with various challenges, we present three. Firstly, in many applications, the human who annotates the images needs to be a domain expert—this makes manual segmentation operationally challenging and costly. In medical neuroimaging, for instance, some of the common segmentation types are tissue type segmentation where brain tissue is categorized into three major tissue types, white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF), see Figure 2.2. Manually annotating even one neuroimage volume of a patient would require a considerable amount of time and the expert knowledge of a radiologist. Secondly, the image processing task might require annotating an unmanageable quantity of images, on the order of hundreds or even thousands, which quickly becomes intractable when done manually. Thirdly, manual segmentations tend to vary greatly across annotators, even in the case where the annotators are domain experts [43]. Moreover, certain scanners produce images that could be automatically segmented with relative ease. A CT neuroimage for instance has intensities with fixed physical correspondence, making automated segmentation a straightforward task. For those reasons and more, automated segmentation is preferred in practice. In this chapter, we will broadly categorise automated segmentation methods into three categories, generative model-based methods, deep learning methods and a broad category that contains methods that are not included in the first two. Conceptually, a segmentation Z of the image I can be represented as the following map

$$Z : \Omega \to \{0, ..., K-1\} \tag{2.0.1}$$

where K is the number of distinct image segments.



(a) Axial slice (b) Tissue segmentation

Figure 2.2: A segmentation of an axial brain slice. Red, green and blue correspond to white matter, grey matter and cerebrospinal fluid respectively. (Figure taken from [62, Fig. 1])

2.1 Generative Model-Based Methods

In generative model-based methods, the segmentation of an image is modelled as a statistical inference problem. More specifically, a generative model of the image is specified, in which the segmentation is a latent variable and segmenting an image corresponds to inferring this variable.

2.1.1 Univariate Models

In a point-wise fashion, univariate methods model image intensities from different segments as sub-populations from a Finite Mixture Model (FMM) with different intensity distributions describing the data generating processes of different segments. Under this modelling approach image intensities belonging to specific segments are drawn from distinct mixtures. A special case of FMM that is often used in image segmentation is a Gaussian mixture model (GMM) where the mixture distributions are Gaussians [5, 10, 18]; we will use a GMM to demonstrate how a generative model can be used to segment an image. The image is represented by a random variable I with is a collection of independent random variables $I(\mathbf{x})$ with support $\mathbb{R}$ and let $Z(\mathbf{x})$ be random variable with support $\{0,1\}^K$ representing the intensity value and mixture assignment for some image domain element $\mathbf{x} \in \Omega$ respectively and let $\boldsymbol{\pi}$ be a random variable with support $[0,1]^K$ be mixture probabilities where entry $\boldsymbol{\pi}_k$ indicates the probability of belonging to the k^{th} image segment. Assuming a Bayesian setting, $Z(\mathbf{x})|\boldsymbol{\pi} \sim \textbf{Categorical}(\boldsymbol{\pi})$ is a latent variable such that event $Z(\mathbf{x})_k = 1$ indicates that $\mathbf{x}$ belongs to the k^{th} image segment and there is a Dirichlet prior on mixture probabilities, i.e. $\boldsymbol{\pi} \sim \textbf{Dirichlet}(\beta)$. The joint intensity and segment assignment probability density function at a specific $\mathbf{x} \in \Omega$ has the following form

$$\mathbb{P}(I(\mathbf{x}) = x, Z(\mathbf{x}) = \mathbf{z}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \mathbb{P}(I(\mathbf{x}) = x \,|Z(\mathbf{x}) = \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma})\mathbb{P}(Z(\mathbf{x}) = \mathbf{z}|\boldsymbol{\pi}) f_\pi(\boldsymbol{\pi}) \quad (2.1.1)$$

$$\propto \prod_{k=1}^{K} [\mathcal{N}(\mu_k, \sigma_k)]^{\mathbf{z}_k} \prod_{k=1}^{K} \boldsymbol{\pi}_k^{\mathbf{z}_k} \frac{\Gamma(K\beta)}{\Gamma(\beta)} \prod_{k=1}^{K} \boldsymbol{\pi}_k^{\beta-1} \quad (2.1.2)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are mean and variance parameters respectively, for each mixture.

2.1.2 Multivariate Models

In contrast to univariate methods, multivariate methods model the intensity distribution across an entire image domain, while accounting for the long-range dependencies among pixels. Markov Random Fields (MRFs) are a class of models that specify a distribution across the image domain by first discretizing the domain such that

$$\Omega = \{1, ..., M\}^d \quad (2.1.3)$$

for some $M \in \mathbb{N}$ and $d \in \{2, 3\}$. Following this discretization, the segmentation and intensities are indexed by vertices of an undirected graph and with adjacent vertices corresponding to adjacent pixels [24]. A detailed review of MRFs is given in Chapter 4. We first assume that the mean intensity estimates for each segment class are given, let μ_k be the mean intensity estimate for the k^{th} image segment. We can then define a functional which corresponds to the negative log probability of the model and has the following form

$$H[Z] = \sum_{k=1}^{K} \sum_{\mathbf{x} \in \Omega} \mathbb{1}\left(Z(\mathbf{x}) = k\right) \left[\frac{I(\mathbf{x}) - \mu_k}{\sigma_k}\right]^2 + \lambda \sum_{\mathbf{x} \in \Omega} \sum_{\mathbf{y} \in N(\mathbf{x})} \delta_Z(\mathbf{x}, \mathbf{y}) \tag{2.1.4}$$

where $N(\mathbf{x})$ is the set of vertices adjacent to $\mathbf{x}$, $\mathbb{1}(\cdot)$ has the following form

$$\mathbb{1}(P) = \begin{cases} 1 & \text{if proposition } P \text{ is true} \\ 0 & \text{otherwise} \end{cases} \tag{2.1.5}$$

and $\delta_Z(\cdot, \cdot)$ has the following form

$$\delta_Z(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } Z(\mathbf{x}) \neq Z(\mathbf{y}) \\ 0 & \text{otherwise .} \end{cases} \tag{2.1.6}$$

In practice, μ_k and σ_k^2 are typically obtained through a partial labelling (semi-supervised) done by a domain expert. In Equation 2.1.4 the first double summation penalizes a $Z(\cdot)$ which labels pixels whose intensities deviate greatly from the mean intensity in that segment class. For segment class, the distance from the mean is normalized by the standard deviation so that the proximity to the mean across segment classes is comparable. The second double summation favours a $Z(\cdot)$ which gives neighbouring pixels the same label and λ is a parameter that balances both double summations.

2.1.3 Inference and Learning

As previously mentioned, in the context of generative models, image segmentation is a statistical problem in which the segmentation is inferred and the parameters governing the generative model are learned. In this sub-section, we give examples of common inference problems in image segmentation.

MLE and MAP

In the case where we have access to a tractable likelihood or posterior distribution, one can do this inference by a maximum likelihood estimation (MLE) or maximum a posteriori (MAP) estimation of the segment assignments. More formally, assuming a Bayesian context, suppose we have access to a tractable and reasonably well-behaved posterior $p(Z|I)$

where Z and I are the segmentation MAP and the image, respectively. A MAP estimate of the segmentation MAP would have the following form

$$\hat{Z}_{MAP} = \underset{Z}{\operatorname{argmax}}\, p(Z|I) \ . \tag{2.1.7}$$

In the case of MRFs, with the help of a result known as *Hammersley-Clifford-Theorem* [26] which we present in Chapter 4, it can be shown that the energy functional in Equation 2.1.4) uniquely determines a well-defined posterior distribution such that

$$\hat{Z}_{MAP} = \underset{Z}{\operatorname{argmin}}\, E[Z] \ . \tag{2.1.8}$$

The posterior distribution of a segmentation is not always easy to directly sample from, this is particularly true for MRFs. In these scenarios, graph-based Markov chain Monte Carlo (MCMC) methods are typically used. More specifically, Gibbs sampling is generally used for MRFs [30] since they are a special case of a conditional random field (CRF) making it relatively easy to specify each segment assignment as a conditional probability.

Variational Inference

Rather than sampling an intractable posterior, one can use a method known as variational inference (VI) to approximate the posterior with a distribution that comes from a family of tractable distributions. This family of tractable distributions are called variational distributions—named after *variational calculus*. Once the family of distributions are specified one can approximate the posterior by finding the variational distribution that optimizes some metric between the posterior and itself. The most common metric used to measure the similarity between two distributions is the Kullback-Leibler (KL) divergence and it is defined as follows

$$KL(q||p) = \mathbb{E}\left[\log \frac{q(x)}{p(x)}\right] = \int q(x) \log \frac{q(x)}{p(x)} \tag{2.1.9}$$

where $q(\cdot)$ and $p(\cdot)$ are probability density distributions over the same support. In the special case of variational inference $q(\cdot)$ is an approximate density and $p(\cdot)$ is the true. Inferring the latent segmentation through this distributional approximation can be formulated as a variational Bayesian expectation maximization (VBEM) inference problem [9]. For a deeper analysis of VI the interested reader can consult section 4 of [11].

2.2 Deep Learning Methods

In recent years deep learning (DL) methods have been successfully applied to many learning tasks and in computer vision, in particular, they have been shown to outperform previous

state-of-the-art machine learning techniques [61]. Loosely inspired by computational models of biological learning, DL methods allow efficient and highly parallelizable computational models of multiple processing layers which implicitly learn data representations [25]. The structural configurations of these processing layers are known as architectures and some of the architectures that are prominent in computer vision include generative adversarial networks (GAN) [15], recurrent neural networks (RNN) [50] and convolutional neural networks (CNN) [21], with CNNs performing particularly well in image segmentation tasks. The authors of [28] applied a 3D CCN to brain lesion segmentation and were able to improve on the state-of-the-art and get top performance on public benchmark datasets BRATS 2015 [35] and ISLES 2015 [34] which are public datasets used in brain lesion segmentation challenges.

2.2.1 Convolutional Neural Networks

Presently, Convolutional Neural Networks (CNNs) are considered the state-of-the-art networks for supervised DL image segmentation problems [37]. Their architecture is inspired by a hierarchical receptive field model of the visual cortex and generally includes the composition of three types of layers. Convolutional layers, where a kernel (filter) is convolved over inputs to extract a hierarchy of features, nonlinear layers which allow inputs to be mapped to feature spaces and pooling layers which reduce the spatial resolution by aggregating local information. Each layer is made of processing units that are locally connected and these local connections are called receptive fields. The layers are typically composed to form a multi-resolution pyramid in which higher-level layers learn features from wider receptive fields. The model parameters are typically learned through a stochastic version of the backpropagation algorithm [48], which is a gradient-based optimization routine that efficiently propagates the gradient of the residual through the network.

2.3 Other Methods

Many segmentation methods cannot naturally be categorised into generative models or deep learning methods. In this section, we provide three examples of such classes of methods, intensity thresholding, region-based and deformable shape methods. Without loss of generality, for the remainder of the chapter, we will assume that the segmentation task is binary.

2.3.1 Intensity Thresholding

Intensity thresholding is a conceptually simple image segmentation method in which a pixel is classified into one of the segment classes if its intensity is within a specified interval. These methods are typically used to segment between the foreground and the background of an image since in many applications of image processing, the average grayscale value of pixels

belonging to an object of interest is substantially different than the average grayscale value of the pixels belonging to the background [53]. More formally, for a given scalar threshold γ, we have the following segmentation map

$$Z(\mathbf{x}) = \begin{cases} 1 & \text{if } I(\mathbf{x}) > \gamma \\ 0 & \text{otherwise .} \end{cases} \tag{2.3.1}$$

Otsu's Method

To fully automate thresholding, one can implement an algorithm that automatically finds an optimal threshold γ. Many such algorithms use the histogram of the image intensities to estimate an intensity PMF in order to find an optimal γ, we will present one of them—*Otsu's method* [42]. This method is closely related to Fisher's Discriminant Analysis (FDA) with the main difference being that Otsu's method is unsupervised, whereas FDA is supervised. Suppose $p(\cdot)$ and $p(\cdot)$ are the intensity CDF and PMF respectively of some image. Further, suppose G is the set of intensities and that the intensity values are bounded. We can define the mean and variance of segment class 0 as a function of the threshold as follows

$$\mu_0(\gamma) = \sum_{g \in G : g \leq \gamma} g \cdot p(g) \tag{2.3.2}$$

and

$$\sigma_0^2(\gamma) = \sum_{g \in G : g \leq \gamma} (g - \mu_0(\gamma))^2 \cdot p(g) \tag{2.3.3}$$

the mean and variance of the second segment class are defined similarly by taking the sum over intensities that are greater than the threshold, namely

$$\mu_1(\gamma) = \sum_{g \in G : g > \gamma} g \cdot p(g) \tag{2.3.4}$$

and

$$\sigma_1^2(\gamma) = \sum_{g \in G : g > \gamma} (g - \mu_1(\gamma))^2 \cdot p(g) \ . \tag{2.3.5}$$

Otsu's method is an algorithm that exhaustively searches for the threshold that minimizes the intra-class variance *Otsu's method* [42], defined as a weighted sum of variances of the two classes. The optimal γ is retrieved by solving the following optimization problem

$$\gamma^* = \underset{\gamma}{\operatorname{argmax}} \left\{ \frac{P(\gamma)[1 - P(\gamma)][\mu_0(\gamma) - \mu_1(\gamma)]^2}{P(\gamma)\sigma_0^2(\gamma) + [1 - P(\gamma)]\sigma_1^2(\gamma)} \right\} . \tag{2.3.6}$$

2.3.2 Region-Based Methods

Region-based methods perform segmentation by attempting to find regions within an image such that within each region, some notion of similarity is maximized while simultaneously minimizing the region boundaries.

Mumford-Shah Functional Model

The Mumford-Shah functional model, first proposed by the authors of [39], is a segmentation model that retrieves an image segmentation by computing an optimal piecewise-smooth or piecewise-constant function approximation to the original image. The method aims to find a partition of the image domain Ω.

$$\Omega = \Omega_0 \cup \Omega_1 \cup \mathcal{B} \tag{2.3.7}$$

where Ω_0 and Ω_1 are image domain partitions belonging to each respective segment class and $\mathcal{B}$ is the set of edge pixels, pixels belonging to the shared boundaries between the image segments. The objective is to find a piece-wise function that is constant within the Ω_k but varies discontinuously and greatly across most of the boundary $\mathcal{B}$ between different Ω_k. Rather than mapping pixels to the label set $\{0, 1\}$, we will consider the following piecewise-constant segmentation map

$$Z(\mathbf{x}) = \begin{cases} a & \text{if } \mathbf{x} \in \Omega_0 \\ b & \text{if } \mathbf{x} \in \Omega_1 \end{cases} \tag{2.3.8}$$

where $a, b \in \mathbb{R}$ and we will consider the following functional

$$E[Z, \mathcal{B}] = \int_{\Omega \setminus \mathcal{B}} \left(I(\mathbf{x}) - Z(\mathbf{x})\right)^2 d\mathbf{x} + \lambda |\mathcal{B}| \ . \tag{2.3.9}$$

The integral in Equation 2.3.9 attempts to quantify how well the piece-wise constant segmentation approximates the original image. The second term, the cardinality of the edge-pixel set, quantifies the number of allowable edges within the segmentation and λ is again a balancing parameter. To retrieve the optimal segmentation, one has to minimize the functional in Equation 2.3.9. In principle, the segmentation that minimizes this functional, will approximate the original image well, all the while minimizing the boundaries between the segment, which effectively minimized the number of disconnected segments and maximizes the area (or volume) of the connected segments [45].

2.3.3 Deformable Shape Methods

Deformable shape methods perform segmentation by relying on the registration of an image and an annotated template image, sometimes called an *atlas*. Once both images are aligned,

regions that are pre-segmented on the template allow automatic segmentation of the source image. Suppose ϕ is a deformation that geometrically aligns image contents of a source image I and an atlas T such that

$$I = T \circ \phi^{-1} \, . \tag{2.3.10}$$

In Chapter 3, we give a more rigorous formulation of such a deformation. Further, suppose that T has a corresponding segmentation Z_T. Under this formulation, the segmentation for the image I would be

$$Z_I = Z_T \circ \phi^{-1} \, . \tag{2.3.11}$$

2.4 Evaluation Methods

Generally, supervised segmentation evaluation methods attempt to quantify the degree of overlap between an estimated and ground truth segmentation. Using the map notation of a segmentation in 2.0.1, we can equivalently understand the segmentation as a set with the image of its map, i.e. $Z(\Omega)$.

The Dice score is one of the most popular and conceptually easy to understand segmentation evaluation methods. For two segmentations A and B, the Dice score is calculated as follows

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|} \, , \tag{2.4.1}$$

which is the size of the overlap relative to the mean region size [59]. Where the numerator accounts for the segmentations' overlap and the denominator accounts for the sizes of the segmentations. The Jaccard coefficient is another segmentation evaluation method and is calculated as follows

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \, . \tag{2.4.2}$$

J and D are equivalent [59] and are related through the following expression

$$D = \frac{2J}{1 + J} \, . \tag{2.4.3}$$

D is known to yield higher values for larger volumes. Another segmentation evaluation method is the average Hausdorff distance, which is especially recommended for segmentation tasks with complex boundaries and small thin segments and compared to the Dice score, the average Hausdorff distance has the advantage of accounting for localisation when considering segmentation performance [7]. For two segmentations A and B, which are

non-empty subsets of a metric space $\big(S, d\big)$, the Average Hausdorff distance is calculated as follows

$$H(A,B) = \frac{1}{2}\Big(\frac{1}{|A|}\sum_{x\in A}\min_{y\in B} d(x,y) + \frac{1}{|B|}\sum_{y\in B}\min_{x\in A} d(x,y)\Big)\,. \tag{2.4.4}$$

For more evaluation methods, the interested reader can consult [59].

CHAPTER 3

Image Registration

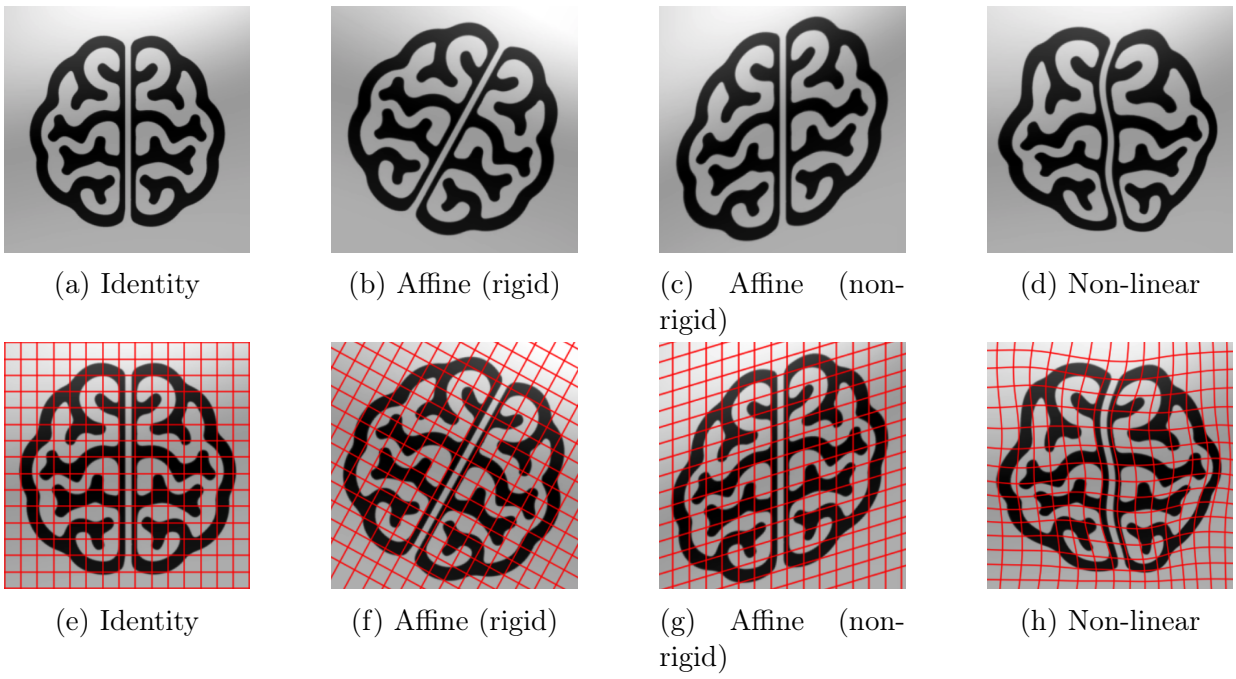
In this chapter, we introduce the general theory of image registration and its applications. As previously mentioned, image registration is the process of geometrically aligning two or more images. Once the substructures are aligned the set of images is said to be registered and the images are now in spatial correspondence. To achieve this correspondence one of the images is held fixed while all other images are *warped* such that they align with the fixed image. The fixed image is called the *target, reference* or *template* image and the warped images are called the *source* or *moving* images. Without loss of generality, we will consider the registration of two images, one fixed and one moving. Before registration, we need to specify a spatial transformation model and a method of quantifying image misalignment.

3.1 Spatial Transformation Between Image Domains

Throughout this thesis, we will assume that images being registered have been affinely co-registered to a common space, Ω, beforehand. This image preprocessing step is typical in registration procedures since it allows us to omit large affine transformations at registration time and also allows us to assume that the deformations are mapping between the same common space. A spatial transformation of an image is a mapping from the image domain to itself.

$$\phi : \Omega \to \Omega \tag{3.1.1}$$

With this formulation, a transformed moving image is represented by the composition $I \circ \phi$ which is also known as the pull-back of I by ϕ. Since ϕ is a transformation of the image domain Ω, it can alter quantities—such as intensities—defined on Ω. If ϕ is bijective it can be viewed as a change of coordinates from source image coordinates to template coordinates or more generally viewed as an image transformation that changes the appearance of a source image, into the appearance of the target image. Consider the source and target images in Figure 3.4, where we are looking to warp an image of a square into an image of a circle. Subfigure 3.4 (e) shows the registered pull-back image, i.e. $I \circ \phi$ and subfigures (e) and(f) demonstrate the corresponding warped coordinates assuming we started with a regular grid coordinate system.



(a) Identity (b) Affine (rigid) (c) Affine (non-rigid) (d) Non-linear

(e) Identity (f) Affine (rigid) (g) Affine (non-rigid) (h) Non-linear

Figure 3.2: Different types of image spatial transformations. Transformations (a)-(c) are linear and (d) is non-linear.

3.1.1 Affine Registration

Affine registration refers to the case where the transformation space is restricted to affine transformations, more formally ϕ takes the following form

$$\phi(\mathbf{x}) = Q\mathbf{x} + \mathbf{q} \tag{3.1.2}$$

where $\mathbf{x} \in \Omega$, Q is a d by d matrix and $\mathbf{q}$ is a d-dimensional vector. This class of transformations can adequately align a set of images when the misalignment between them can be modelled as an affine transformation such as a translation, rotation, scaling, shearing, reflection or any combination of them. Moreover, from an engineering point of view, affine transformations are relatively easy to implement due to their relatively low dimensional parametrization and relative ease of imposing sufficient invertibility conditions.

3.1.2 Non-Linear Registration

Non-linear registration (or more correctly, non-affine registration) allows transformations that are not affine. The class of admissible transformations is typically very high-dimensional, often infinite-dimensional, making these registration schemes much more complicated and

computationally demanding. We now outline several such approaches, focussing on the transformation classes.

Low-Frequency Basis Functions

We can start by considering transformation classes that can model global non-linear variability between images. We can parametrize these deformations by formulating them as linear combinations of low-frequency basis functions such as Fourier bases, sine and cosine transform basis functions, B-splines and piecewise affine or trilinear basis functions [6, 4]. Such a transformation model has the following form

$$\phi(\mathbf{x}_i) = \mathbf{x}_i + \sum_{j=1}^{p} \lambda_{ij} b_j(\mathbf{x}_i) \tag{3.1.3}$$

where $\mathbf{x}_i$ is the i^{th} entry of vector $\mathbf{x} \in \Omega$ and b_j is the j^{th} basis function. Under this setting, deformations of d-dimensional images can be parametrized by the following matrix

$$\Lambda = \begin{bmatrix} \lambda_{11} & \dots & \lambda_{1p} \\ \vdots & \ddots & \vdots \\ \lambda_{d1} & \dots & \lambda_{dp} \end{bmatrix} \tag{3.1.4}$$

where p is the number of distinct basis functions. The parameter matrix Λ is chosen such that it minimizes an objective function, also known as the *energy*. The term energy comes from Elastic Registration Theory where the fixed and moving images are modelled as a continuous dense elastic plane sheet and under this perspective, Elasticity Theory is used to register the images. In particular, the registration problem is seen as a balancing of internal and external forces (image intensity differences) acting on elastic objects and the cost function is synonymous with the potential energy of this system; when the internal and external forces are balanced, the registration is complete [17].

Radial Basis Functions

The linear elasticity assumption in the previous section allows for global non-linear deformations but doesn't allow for highly localized deformations. This is in part because there is nothing local about the parametrization; the matrix Λ is defined globally and is fixed over the entire image domain. Starting with the transformation model in Equation 3.1.3 we can define a transformation model that can handle localized deformations. First, we choose a countable set of points from Ω, let's call this set of points $P \subset \Omega$. These points will allow us to define localized deformations by transforming these points and interpolating the deformation field across all of Ω. The points in P are also called *control points* or *landmarks*.

For any given point in the image domain, the deformation field will be interpolated in a way that considers the point's proximity to the control points. The distance between a given point and a control point will be computed through a radial basis function and the transformation model has the following form

$$\phi(\mathbf{x}_i) = \mathbf{x}_i + \sum_{j=1}^{p} \lambda_{ij} b_j(\mathbf{x}_i) + \sum_{\mathbf{p} \in P} \alpha_{\mathbf{p}} \kappa(\|\mathbf{x} - \mathbf{p}\|) \; . \tag{3.1.5}$$

Note that Equation 3.1.5 is similar to 3.1.3 but with an extra term. $\mathbf{p}$ is a control point, $\alpha_{\mathbf{p}}$ is a parametric coefficient and κ is the radial basis function. Typically under this transformation model, the basis functions are polynomials up to a specified degree and κ is a Gaussian kernel or any positive-definite kernel. The number of control points and how they are distributed introduce an added level of complexity in registration schemes that use this class of transformation models. Since the control points introduce localization, they should be well distributed across the image domain Ω, if localized deformations are to be expected across the image domain and if not, more control points can be placed in areas of high localized deformation. Moreover, the higher the number of control points the higher the dimension of the parametrization.

Large Deformation Diffeomorphic Metric Mapping

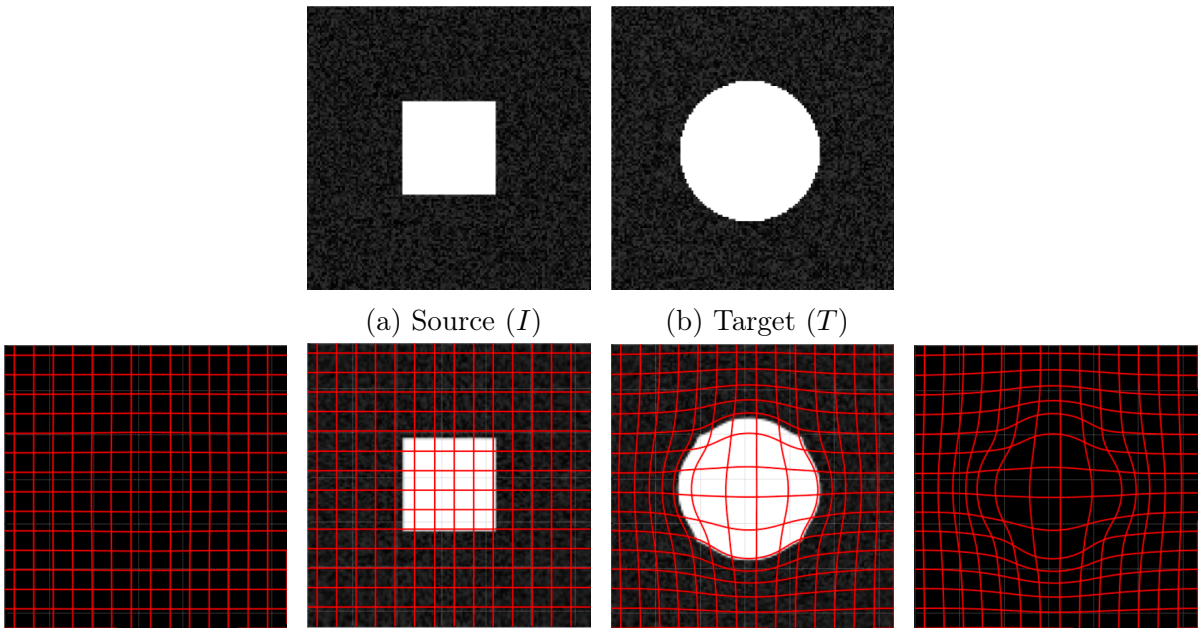
Though the radial basis based transformation models allow for some level of localized deformation they are not a very principled way of achieving this end. This is due to the fact that under that transformation model, one cannot give sufficient conditions to guarantee desirable properties of the deformation field. By carefully sampling control points and regularizing the deformation field we can guarantee the invertibility of the resulting deformation [49] but this is not a simple task and it requires hard constraints that might compromise registration performance. Moreover, regularizing harshly under the basis function transformation models yields globally smooth deformations which result in small local deformations. This motivates the desire for a principled registration approach that retrieves transformations with attractive theoretical guarantees such as invertibility and topological invariance, while also performing well by allowing large deformations and high localization. The Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework is such a framework. In LDDMM, the registration problem corresponds to a variational problem that finds a geodesic (shortest path) on the Riemannian manifold of diffeomorphisms on Ω, i.e. $\mathbf{Diff}(\Omega)$. It's not self-evident how one can optimize over this group; one approach is to introduce a time variable t and solve a system of partial differential equations. More specifically, it can be shown that one way to generate deformations from $\mathbf{Diff}(\Omega)$ is to solve the following non-stationary transport equation

$$\frac{\partial \, \phi(\mathbf{x}, t)}{\partial t} = v(\phi(\mathbf{x}, t), t) \;\; \text{such that} \;\; \phi(\mathbf{x}, 0) = Id \;\; \text{and} \;\; t \in [0, 1] \tag{3.1.6}$$

where $v(\phi(\mathbf{x}, t), t)$ is a non-stationary smooth velocity vector field belonging to a Hilbert space $\mathcal{H}$. Under this setting, the final diffeomorphism is obtained through the following time-integration

$$\phi(\mathbf{x}, 1) = \int_0^1 v(\phi(\mathbf{x}, t), t)dt + Id \ . \tag{3.1.7}$$

This integral describes a path on $\mathbf{Diff}(\Omega)$ that starts at Id and ends at $\phi(\mathbf{x}, 1)$. A technique known as *geodesic shooting* [66] exploits the fact that the time-dependent velocity field, as a solution to the first-order PDE in Equation 3.1.6, is uniquely specified by its initial velocity i.e. $v(Id, 0)$. Intuitively, geodesic shooting is analogous to repeatedly adjusting the angle of a cannon in order to specify a projectile trajectory. Optimizing over initial conditions (instead of over the time-varying vector field) significantly reduces the space-time algorithmic complexity during runtime since we would not have to compute and store an entire time series of velocity fields [58 ,10 , 2 , 36].



(a) Source (I) (b) Target (T)

(c)Identity map (Id) (d) Source ($I \circ Id$) (e) Source ($I \circ \phi$) (f) Warp (ϕ)

Figure 3.4: Deformable registration. Source and target images are to be registered such that the image of a square is warped onto an image of a circle along with the source images with superimposed coordinates.

3.2 Quantifying Image Alignment

If we formalize the misalignment as some real-valued functional E that acts on the warped image $I \circ \phi$ and target image J, we can then conceptualize the image registration problem as the following abstract optimization problem

$$\phi^* = \underset{\phi}{\operatorname{argmin}}\, E[I \circ \phi, J] \tag{3.2.1}$$

where ϕ^* is an optimal transformation that aligns the two images. The quantity E is called an *energy function*, *cost function* or a *similarity measure*. The latter is potentially misleading since we are minimizing, so we really mean E is a *dissimilarity* measure; nonetheless the terminology is common. Similarity measures can broadly be categorized into two types, point-wise measures and global measures. For simplicity, In this section we will assume that the image domain Ω is discretized and therefore the similarity measures will be defined such that we sum over elements of Ω rather than integrating over them.

3.2.1 Point-wise Similarity Measures

Point-wise similarity measures compute the similarity of two images using point-wise intensity values—that is, intensities at individual pixels. These similarity measures are effective when the assumption that the intensity distribution is similar across images holds as in the case of images with the same modalities. Two common point-wise similarity measures are the normalized sum of squared differences (SSD) and the normalized sum of absolute differences (SAD) and they have the following forms

$$E_{SSD}(I, J) = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} (I(\mathbf{x}) - J(\mathbf{x}))^2 \tag{3.2.2}$$

and

$$E_{SAD}(I, J) = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} |I(\mathbf{x}) - J(\mathbf{x})| \,. \tag{3.2.3}$$

These two similarity measures behave similarly and are easy to implement and interpret. SAD is slightly more robust to outlying intensities, while SSD over-penalizes them [24].

Another measure that can be viewed as a scaled and translated E_{SSD} is the correlation coefficient (CC) of two images and is calculated as follows

$$E_{CC}(I, J) = \frac{\sum_{\mathbf{x} \in \Omega} (I(\mathbf{x}) - \mu_I)(J(\mathbf{x}) - \mu_J)}{\sqrt{\sum_{\mathbf{x} \in \Omega} (I(\mathbf{x}) - \mu_I)^2} \sqrt{\sum_{\mathbf{x} \in \Omega} (J(\mathbf{x}) - \mu_J)^2}} \tag{3.2.4}$$

such that

$$\mu_I = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega} I(\mathbf{x}) \quad \text{and} \quad \mu_J = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega} J(\mathbf{x}) \tag{3.2.5}$$

where μ_I and μ_J are intensity means for images I and J respectively. E_{CC} is a measure of the linear relationship between the image intensities and since we are considering dissimilarity, the objective function would be $E = 1 - E_{CC}$.

3.2.2 Global Similarity Measures

Using information-theoretic measures we can construct similarity measures that are able to model more complex statistical relationships between image intensities; one such measure is the *mutual information* (MI) of two images. We must first estimate the intensity probability distribution for each image using the image intensities as a random sample. The Statistics literature gives many ways to estimate probability mass functions (PMFs) from finite samples. With these PMFs, the Shannon marginal entropy is defined as follows

$$\mathbf{H}(I) = -\sum_{\mathbf{x}\in\Omega} p(\mathbf{x}) \log p(\mathbf{x}) \tag{3.2.6}$$

where $p(\cdot)$ is the estimated marginal intensity distribution for image an I. The Shannon joint entropy is defined as follows

$$\mathbf{H}(I, J) = -\sum_{\mathbf{x}\in\Omega}\sum_{\mathbf{y}\in\Omega} p(\mathbf{x}, \mathbf{y}) \log p(\mathbf{x}, \mathbf{y}) \tag{3.2.7}$$

where $p(\cdot, \cdot)$ is the estimated joint intensity distribution for images I and J. These entropies are statistical measures of *information* or *disorder*. The MI of two images is defined as follows

$$E_{MI}(I, J) = \mathbf{H}(I) + \mathbf{H}(J) - \mathbf{H}(I, J) \ . \tag{3.2.8}$$

A normalized variant of MI is the entropy correlation coefficient (ECC)

$$E_{ECC}(I, J) = 2 - \frac{2\mathbf{H}(I, J)}{\mathbf{H}(I) + \mathbf{H}(J)} \ . \tag{3.2.9}$$

These measures work particularly well when doing multi-modal registration when the intensities across images are assumed to come from very different distributions.

3.2.3 Regularized Similarity

So far we've only considered cost functions that only account for the similarity between the moving image and the target image, however, it would be desirable to introduce another term that penalizes deformations with undesirable properties, this is called regularizing. Without regularization, it would be possible to introduce unnecessary deformations that only reduce the cost function by a small amount [4]; if we only optimize for image similarity and lead the deformation unrestricted, we could retrieve a deformation that seems to align the images but that is not smooth or does not preserve topology. Regularizing allows us to define a term that restricts the solution space and that is independent of the image data. To do so, we add a term to the optimization problem in Equation 3.2.1 as follows

$$\phi^* = \underset{\phi}{\operatorname{argmin}}\, E_{data}[I \circ \phi, J] + E_{reg}[\phi] \tag{3.2.10}$$

where E_{reg} is the regularizer term. In this section, we revert back to the continuous definition of the image domain where Ω is a compact and simply connected subset of $\mathbb{R}^d$. To present concrete examples of regularizations, suppose we are under a deformation model that can be expressed as the following deformation field

$$\phi(\mathbf{x}) = \mathbf{x} - \mathbf{u}\ . \tag{3.2.11}$$

This transformation model allows for many deformations that might be undesirable and so regularization is necessary if one hopes to retrieve a plausible deformation. Three common regularization schemes are based on the minimization of the *membrane*, *bending*, *linear elastic* energies and have the following form

$$E_{membrane}[\phi] = \int_\Omega \sum_{j=1}^{d}\sum_{i=1}^{d} \left(\frac{\partial \mathbf{u}_j}{\partial \mathbf{x}_i}\right)^2 d\mathbf{x} \tag{3.2.12}$$

$$E_{bending}[\phi] = \int_\Omega \sum_{k=1}^{d}\sum_{j=1}^{d}\sum_{i=1}^{d} \left(\frac{\partial^2 \mathbf{u}_k}{\partial \mathbf{x}_i \mathbf{x}_j}\right)^2 d\mathbf{x} \tag{3.2.13}$$

$$E_{elastic}[\phi] = \int_\Omega \sum_{j=1}^{d}\sum_{i=1}^{d} \frac{\lambda_o}{2}\left(\frac{\partial \mathbf{u}_i}{\partial \mathbf{x}_i}\right)\left(\frac{\partial \mathbf{u}_j}{\partial \mathbf{x}_j}\right) + \frac{\mu_o}{4}\left(\frac{\partial \mathbf{u}_i}{\partial \mathbf{x}_j} + \frac{\partial \mathbf{u}_j}{\partial \mathbf{x}_i}\right)^2 d\mathbf{x} \tag{3.2.14}$$

where $d\mathbf{x} = d\mathbf{x}_1 \dots d\mathbf{x}_d$. The membrane, bending and linear elastic energies are based on the Laplacian, Hessian and Jacobian of the deformation fields respectively [4] and λ_o and μ_o are Lamé elastic material constants [14].

Under the LDDMM diffeomorphic transformation model from Section 3.1.2 we can regularize the corresponding velocity vector field to be as smooth as possible. Assuming

the time-dependent velocity vector field v (from equation 3.1.6) belongs to a Reproducing Kernel Hilbert space (RKHS) $\mathcal{H}$ with norm $\|\cdot\|_{\mathcal{H}}$ we can define the following regularization term

$$E_{velocity}[v(\cdot,0)] = \int_{\Omega}\int_{0}^{1} \|v(\phi(\mathbf{x},t),t)\|_{\mathcal{H}}^{2} dt d\mathbf{x} \tag{3.2.15}$$

and the choice of reproducing kernel associated with $\mathcal{H}$ allows us to model spatial regularizations with different characteristics [58]; this regularizer penalizes diffeomorphisms that are far from the identity.

3.2.4 Generative Model Perspective

We now explain how the dissimilarity measures (*energies*) defined above can be viewed as the negative log of a well-defined probability distribution over a discretized space of deformations. We will borrow the idea of a *Gibbs distribution* (also known as a *Boltzmann distribution*) from statistical physics [47]. In particular, a Gibbs measure can be used to relate the *probability* of the state of a system to its *energy* [47]; the underlying idea is that the states with the lesser energies are the most probable. Using this idea, we can model the similarity measures in Section 3.2 as the energy (or Hamiltonian) of a system, where the state space of a system corresponds to the space of deformations. To give a concrete example, we will consider the case where the deformations come from the group $\mathbf{Diff}(\Omega)$. We first note that E_{reg} can be understood as a map

$$E_{reg} : \mathbf{Diff}(\Omega) \to [0,\infty) \tag{3.2.16}$$

and we consider a finite measure μ defined on all of $\mathbf{Diff}(\Omega)$. Since $\mathbf{Diff}(\Omega)$ is non-compact, the existence of this finite measure has to be carefully established, using techniques in [44]. In practice however, it suffices to use only a compact subset of $\mathbf{Diff}(\Omega)$ parametrized by a compact set of vector fields (using the geodesic parametrization described earlier), and on this set a finite measure clearly exists. Under this setting we can further consider a random variable Φ with support $\mathbf{Diff}(\Omega)$ and under this abstract setting we can define a Gibbs measure which is equivalent to a probability density function with the following form

$$\mathbb{P}(\Phi = \phi) = \frac{1}{Z_{reg}(\beta)} \exp\big\{ -\beta E_{reg}[\phi]\big\} \tag{3.2.17}$$

where β is a free parameter and $Z_{reg}(\beta)$ is partition function corresponding to a normalizing constant with the following form

$$Z_{reg}(\beta) = \int \exp\big\{ -\beta E_{reg}[\phi]\big\} d\mu(\phi)\ . \tag{3.2.18}$$

This distribution is well-define since $E_{reg} \geq 0$ and $\mu < \infty$ are sufficient conditions for $Z_{reg}(\beta)$ to be finite. Moreover, given a deformation ϕ, a source image I which is a realization of a random variable $\mathbf{I}$ and a target image T we can define the following likelihood function

$$\mathcal{L}_I(\phi) = \mathbb{P}(\mathbf{I} = I|\ \phi, T) \tag{3.2.19}$$

$$= \frac{1}{Z_{data}} \exp\big\{ - E_{data}[I \circ \phi, T]\big\}\ . \tag{3.2.20}$$

where

$$Z_{data} = \int_{\Omega} \exp\big\{ - E_{data}\big[I \circ \phi(\mathbf{x}), T(\mathbf{x})\big]\big\} d\mathbf{x}\ . \tag{3.2.21}$$

Therefore the following optimization problem

$$\phi_{MLE} = \underset{\phi}{\operatorname{argmax}}\, \mathcal{L}_I(\phi) \tag{3.2.22}$$

is equivalent to the optimization problem in equation 3.2.1; that is, from a generative model perspective, an optimal deformation for registration is the Maximum Likelihood Estimate (MLE). If we include a regularized cost function we can then consider a Bayesian framework where the likelihood becomes a posterior distribution and the MLE becomes a Maximum a Posteriori (MAP) estimate, this can be achieved by considering the following posterior

$$\mathbb{P}(\Phi = \phi|\mathbf{I} = I, T) = \frac{\exp\big\{ - E_{data}[I \circ \phi, T] - \beta E_{reg}[\phi]\big\}}{Z_{data} Z_{reg}(\beta) p(I)}\ . \tag{3.2.23}$$

CHAPTER 4

Markov Random Fields

In this chapter, our presentation of Markov random fields loosely follows [24] which provides a good overview of this topic generally and its applications to image registration problems in particular. As mentioned in the previous chapter, the segmentation component in our framework could be enhanced through carefully modelling local dependencies. In particular, we would like a model that naturally penalizes adjacent pixels that do not share the same segmentation label; Markov random fields (MRFs) are such models. In addition to modelling the segmentation component of our framework, they can also be used to model the registration component and by imposing some hard constraints on the displacement field during training, we can guarantee the retrieved deformation to be diffeomorphic [13]. Furthermore, the energy functional of these MRFs can be linearized [23] turning our simultaneous estimation framework into an iterative efficient linear programming optimization problem that recovers the lowest potential of the energy functional. In this chapter, we introduce the general theory of MRFs and how they can be used to model image segmentation and image registration problems through efficient linear programming.

4.1 Random Fields

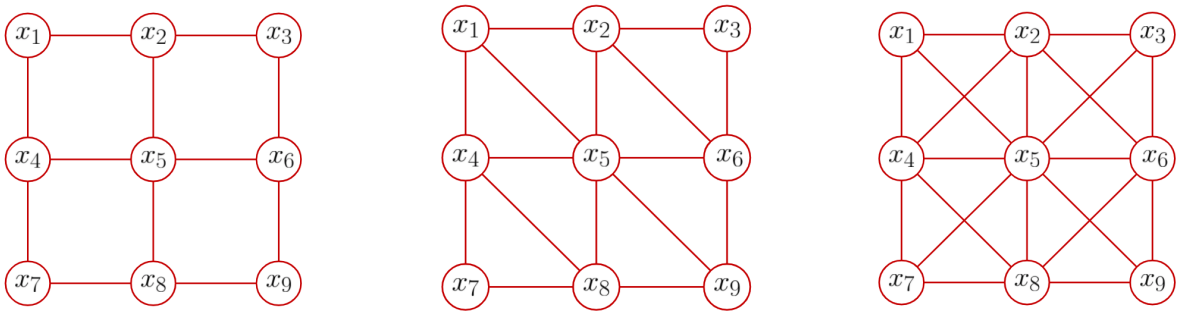
In its most general form, a random field is a set of random variables indexed by elements of a reference space. A random field is a generalization of a unidimensional indexed stochastic process. A stochastic process is a set of random variables indexed by $\mathbb{R}$, typically interpreted as *time*. Rather than thinking of the stochastic behaviour of random variables over time (a process), we can more generally think of the stochastic behaviour of random variables over a space (a field) or index set. Random fields are often considered in graphical models, where the set of random variables is indexed by the vertex set of some specified simply connected (no loops or multiple edges) undirected graph and the edge set of the graph represents some relationship among the vertices. In Section 2 we will come to see that the edge set, or more specifically the neighbourhood structure of this graph, represents some special relationship between the vertices which translates into a regularization we would like to model into our vision problem. Although the general definition of a random field doesn't impose this graph-theoretic approach, the literature often does; this is due to their wide applications in graphical models. That is, when a random field is said to be a set of random variables, these random variables are thought to be indexed by the vertex set of some undirected graph. For the remainder of this chapter, we will think of random fields in this way.

Consider a random field $\mathbf{X}$, which denotes a finite set of random variables and is indexed by a set of vertices V of an undirected graph, further let $X_i \in \mathbf{X}$ be a random variable with support L_i; for simplicity, we will assume L_i is discrete. The form of the joint distribution is given in Section 4.3 and for now, we will begin by first considering the marginal distribution over L_i, denoted $\mathbb{P}(X_i = x_i)$. More generally we can define a joint

probability distribution over the Cartesian product of the supports. Let $\mathcal{X} = \times_{i \in V} L_i$, the Cartesian product of all supports, the joint probability is then denoted by $\mathbb{P}(\mathbf{X} = \mathbf{x})$ for some $\mathbf{x} \in \mathcal{X}$. $\mathcal{X}$ is also known as a labelling or configuration of the random field and $\mathcal{X}$ is the label space or configuration space. In our case, we will be using random fields to model some real-world vision problem, in which the solution will be reduced to a labelling optimization problem, thus searching through $\mathcal{X}$ for some optimal labelling $\mathbf{x}$.

4.2 Neighbourhood Structures

Let $G = (V, E)$ be the graph corresponding to the random field in the previous section, that is V is the vertex set such that there is a one-to-one correspondence between the vertices and the random variables in $\mathbf{X}$ and E is the edge set. It is natural to consider the neighbourhood structure of the corresponding graph since the interconnectivity of the graph is meant to model some underlying relationships between the random variables. We define two types of subsets of V in order to specify this connectivity, namely neighbourhood sets and cliques. Two random variables $X_i, X_j \in \mathbf{X}$ are said to be neighbours if and only if $(i, j) \in E$ (i.e. their corresponding vertices in V are adjacent). For some random variable $X_i \in \mathbf{X}$, let $N(i) \subset V$ denote the set of all neighbours of X_i. A clique is a subset $C \subset V$ such that every pair of distinct vertices in C are adjacent to each, that is, cliques are either single nodes or sets of pair-wise adjacent nodes (i.e. inducing complete sub-graphs of G). Let $\mathcal{C}$ denote the collection of all cliques of G. Furthermore, the *order* of a random field is the cardinality of the maximal clique minus one. For instance, a random field that corresponds to a regular grid graph (every interior node has exactly 4 neighbours) has an order of 1 since the maximal clique is a pair of adjacent nodes. In practice, the order of a random field significantly impacts the computational cost and for this reason, first-order random fields are the most common. For the remainder of this chapter, we only consider first-order random fields, i.e. $|C| = 2$.



(a) First-order (b) Second-order (b) Third-order

Figure 4.2: Random fields with different orders.

4.3 Markov Random Fields

We consider a particular class of random fields, Markov random fields.

Definition 4.3.1 (Markov Random Field)**.** *A random field* $\mathbf{X}$ *is said to be a Markov random field, if and only if it satisfies the following 2 properties*

$$\mathbb{P}(X_i = l \mid \mathbf{X} - \{X_i\}) = \mathbb{P}(X_i = l | N(i)) \qquad (Markovian\ property) \tag{4.3.1}$$

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) > 0 \ \ \forall\, \mathbf{x} \in \mathcal{X} \qquad (Positivity\ property)\ . \tag{4.3.2}$$

The first property states that the conditional probability of any given variable in the field is equivalent to the conditional probability of that variable given its neighbours. Meaning that the variables depend only on their neighbours, this equivalently implies that any two nodes that are not neighbours are conditionally independent. The second property ensures that the distribution can be specified by the potentials in a Gibbs distribution.

Specifying the distribution of a Markov random field (MRF) in terms of the local conditional probabilities is not an easy task. Because of this in practice when MRFs are used for modelling, their probabilities are specified using a theoretical result known as the Hammersley-Clifford-Theorem [26]. Before presenting this result, we first define another particular class of random fields, namely the Gibbs random fields (GRFs).

Definition 4.3.2 (Gibbs Random Field)**.** *A random field* $\mathbf{X}$ *is said to be a Gibbs random field, if and only if its probability distribution has the following form*

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) = \frac{1}{M} \prod_{C \in \mathcal{C}} \exp\big\{ -\psi_C(\mathbf{x}) \big\} \tag{4.3.3}$$

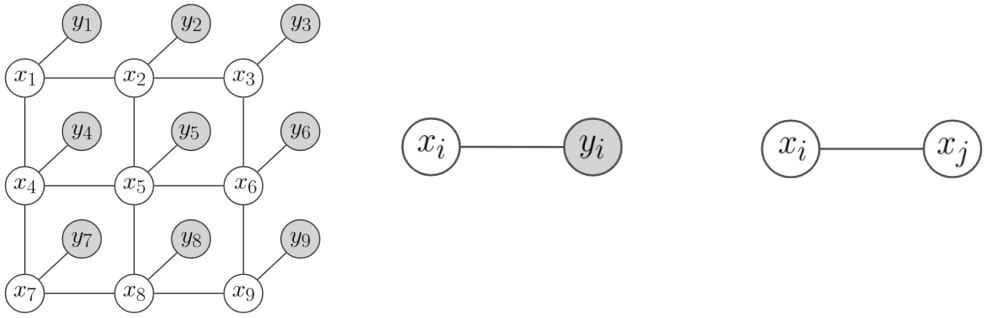
where M is a normalizing constant, $\psi(\cdot)_C$ are called potential functions and $\mathcal{C}$ is the set of cliques in the graph corresponding to $\mathbf{X}$. From Equation 4.3.3 it is clear that the probability distribution of a Gibbs random field can be uniquely specified by these potential functions. Now we are ready to state the Hammersley-Clifford-Theorem.

Theorem 4.3.3 (Hammersley-Clifford-Theorem)**.** *A random field* $\mathbf{X}$ *is a Markov random field if and only if it is a Gibbs random field.*

Rather than specifying the conditional probabilities when modelling with MRFs, we can simply specify the potential functions $\psi(\cdot)_C$.

4.4 Inference with Markov Random Fields

Before going over how MRFs can be used to model image registration and image segmentation problems, we will first give a general methodology for modelling inference problems with MRFs. We first define a first-order Markov random field over the set $\mathbf{X} \cup \mathbf{Y}$, which is the union of two sets of random variables where the set $\mathbf{X}$ is latent and $\mathbf{Y}$ is observed. The random field has a corresponding graph $G = (V, E)$ and set of cliques $\mathcal{C}$. Each variable in $\mathbf{X}$ has a corresponding variable in $\mathbf{Y}$ such that for a pair of corresponding variables $X_i \in \mathbf{X}$ and $Y_j \in \mathbf{Y}$ $(i, j) \in E$ and indices of $\mathbf{X}$ form a regular grid sub-graph of G, this structure is illustrated in Figure 4.4 (a) for 2 sets of 9 latent and observed variables.



(a) Markov random field (b) Likelihood clique (c) Prior clique

Figure 4.4: Sub-figure (a) displays a graphical representation of the a Markov random field over $\mathbf{X} \cup \mathbf{Y}$ where $|\mathbf{X}| = |\mathbf{Y}| = 9$. The x_i and y_i are realizations of random variables $X_i \in \mathbf{X}$ and $X_i \in \mathbf{X}$ respectively. The observed variables are represented by shaded nodes and latent variables are represented as non-shaded nodes. Sub-figure (b) displays the first type of clique, which depends on a latent variable and an observed variable. Sub-figure (b) displays the second type of clique, which depends on two neighbouring latent variables.

P(X=

The joint distribution of this MRF can be written as the following Gibbs distribution

$$\mathbb{P}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) = \frac{1}{M} \prod_{C \in \mathcal{C}} \exp\big\{ -\psi_C(\mathbf{x}, \mathbf{y}) \big\} \,. \tag{4.4.1}$$

Moreover, since an index of a variable in $\mathbf{Y}$ is adjacent to exactly one index of a variable in $\mathbf{X}$ the set of cliques can be partitioned into two sets, one including cliques with indices of variables in $\mathbf{Y}$ and one without, we will denote these partitions with $\mathcal{C}_y$ and $\mathcal{C}_x$ respectively. This partition leads to the following factorization of the joint distribution

$$\mathbb{P}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) = \frac{1}{M} \prod_{C \in \mathcal{C}_y} \exp\big\{ -\psi_C(\mathbf{x}, \mathbf{y}) \big\} \prod_{C \in \mathcal{C}_x} \exp\big\{ -\psi_C(\mathbf{x}, \mathbf{y}) \big\} \tag{4.4.2}$$

and since cliques in $\mathcal{C}_x$ do not depend on the observed variables in $\mathbf{Y}$, we can rewrite the joint distribution as follows

$$\mathbb{P}(\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y}) = \frac{1}{M}\prod_{C\in\mathcal{C}_y}\exp\{-\psi_C(\mathbf{x},\mathbf{y})\}\prod_{C\in\mathcal{C}_x}\exp\{-\psi_C(\mathbf{x})\}\,. \tag{4.4.3}$$

The first and second product corresponds to the likelihood and prior of the MRF respectively and the $\psi_C(\mathbf{x},\mathbf{y})$ and $\psi_C(\mathbf{x},\mathbf{y})$ are called the *likelihood* and *prior potentials*.

4.4.1 MAP estimates

Typically, a Bayesian framework is employed when modelling with MRFs, in which a Maximum a Posteriori (MAP) estimate of the random field is inferred. The idea is to model the problem such that $\hat{\mathbf{X}} = \text{argmax}_{\mathbf{x}\in\mathcal{X}}\, p(\mathbf{X}=\mathbf{x}|\mathbf{Y})$ corresponds to a reasonable solution. Intuitively we are looking for the most probable configuration or labelling of the random field given we observed a set of random variables $\mathbf{Y}$. The posterior $p(\mathbf{X}|\mathbf{Y})$ is proportional to the product of $p(\mathbf{Y}|\mathbf{X})$ (the likelihood) and $p(\mathbf{X})$ (the prior). Since the MAP estimate is obtained by optimizing over the configuration space $\mathcal{X}$ and the distribution $p(\mathbf{Y})$ (the evidence) is free of $\mathbf{X}$ we can equivalently restate our problem as $\hat{\mathbf{X}} = \text{argmax}_{\mathbf{x}\in\mathcal{X}}\, p(\mathbf{Y}|\mathbf{X})p(\mathbf{Y})$. As a consequence of the Gibbs-Markov equivalence, we can specify the distribution of an MRF by specifying the likelihood and prior potential functions its corresponding Gibbs distribution.

4.4.2 MRF Likelihood Distribution

The likelihood models a relationship between the latent random field and the observed data. In the case of first-order random fields, every clique in $\mathcal{C}_y$ corresponds to exactly one corresponding pair of latent and observed variables and therefore the likelihood potentials can be indexed by the vertex set as follows

$$\mathbb{P}(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{x}) \propto \prod_{i\in V}\exp\big\{-\psi_i(x_i,y_i)\big\} \tag{4.4.4}$$

$$= \exp\Big\{-\sum_{i\in V}\psi_i(x_i,y_i)\Big\}\,. \tag{4.4.5}$$

Since the likelihood is a decreasing function of the sum of the likelihood potentials, i.e. low potentials are more probable, the likelihood potentials are usually modelled as a similarity metric between the data and the latent variables. This is similar to Gaussian parametric regression where the likelihood function is the exponential of the negative mean squared errors (MSE) and the Maximum Likelihood Estimate (MLE) corresponds to minimizing this error. The choice of likelihood potentials follows directly from the modelling assumptions of how the MRF relates to the observed data.

4.4.3 MRF Prior Distribution

The prior models some prior assumptions about the random field, independent of the observed data. The prior also helps to restrict the solution space to more likely solutions and in some cases (as we will see with image registration) helps to make the optimization feasible. The prior distribution has the following form

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) \propto \prod_{\{i,j\} \in \mathcal{C}_x} \exp\big\{ -\psi_{ij}(x_i, x_j)\big\} \,. \tag{4.4.6}$$

Because of the Markovian property of MRFs, random variables that aren't neighbours are conditionally independent, thus we can rewrite Equation 4.4.6 in terms of neighbouring vertices as follows

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) \propto \prod_{i \in V} \prod_{j \in N(i)} \exp\{-\psi_{ij}(x_i, x_j)\} \tag{4.4.7}$$

$$= \exp\Big\{ \sum_{i \in V} \sum_{j \in N(i)} \{-\psi_{ij}(x_i, x_j)\Big\} \,. \tag{4.4.8}$$

Specifying the potentials in terms of neighbouring random variables is not only mathematically convenient, but it is also often more intuitive to the modelling assumptions and easier to interpret. For instance, in image segmentation where the random variables might be segment labels of individual pixels, neighbouring pixels could be assumed to belong to the same image segment, thus this assumption would naturally translate into a prior potential function which would be some distance metric between neighbouring pixel labels.

4.4.4 MRF Energy Function

Looking at equations 4.4.4 and 4.4.6 it is clear that the same labelling that maximizes the product of the prior and the likelihood, is the same labelling that minimizes the sum of the prior potentials and likelihood potentials, this sum is called the energy of the MRF and it is equivalent to the negative log probability of the MRF up to an additive constant. This is the same energy discussed in Chapter 3, Section 3.2.4 and the distribution of an MRF is indeed a Gibbs distribution. Consequently finding a MAP estimate reduces to finding a labelling that minimizes the energy.

Let $\mathbf{X} = \{X_i\}_{i=1}^N$ be an MRF. To simplify the notation, we can omit the observed random variable from the likelihood potentials and simply let $\psi_i(x_i)$ and $\psi_{ij}(x_i, x_j)$ be the likelihood and prior potentials respectively where x_i and x_j are labels. The energy function has the following form

$$E(x_1, \ldots, x_N) = \sum_{i \in V} \psi_i(x_i) + \sum_{i \in V} \sum_{j \in N(i)} \psi_{ij}(x_i, x_j) \,. \tag{4.4.9}$$

Under this form $\psi_i(\cdot)$ and $\psi_{ij}(\cdot,\cdot)$ are also known as *unitary* and *pair-wise* potentials, respectively.

4.4.5 Linearization of the Energy

Searching the configuration space for an optimal labelling is computationally intensive. One way to approximate the optimal labelling is by relaxing the optimization of the energy into a continuous linear program (LP). This is attractive for at least 2 reasons, firstly the LP relaxation can often be formulated such that a unique solution exists [23] and secondly continuous linear programs can be solved reasonably fast [31]. Following a linearization procedure proposed in [23], we first demonstrate how the labelling problem can be turned into an integer linear program (ILP) and then relax the ILP into a continuous LP.

Let L be the set of possible labels for any given variable in the random field. Consider 2 binary functions $\phi_i : L \to \{0,1\}$ and $\phi_{ij} : L \times L \to \{0,1\}$ such that

$$\phi_i(\alpha) = \begin{cases} 1 & \text{if} \quad x_i = \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{4.4.10}$$

and

$$\phi_{ij}(\alpha,\beta) = \begin{cases} 1 & \text{if } x_i = \alpha \text{ and } x_j = \beta \\ 0 & \text{otherwise.} \end{cases} \tag{4.4.11}$$

where $\phi_i(\cdot)$ and $\phi_{ij}(\cdot,\cdot)$ are indicator functions for the labels of a single variable and a pair of variables respectively. Using these indicator functions, the MRF energy can be rewritten as

$$E(x_1,\ldots,x_N) = \sum_{i\in V}\sum_{\alpha\in L}\psi_i(\alpha)\phi_i(\alpha) + \sum_{i\in V}\sum_{j\in N(i)}\sum_{\alpha,\beta\in L}\psi_{ij}(\alpha,\beta)\phi_{ij}(\alpha,\beta)\ . \tag{4.4.12}$$

Now the energy is expressed as a linear combination of $\phi_i(\cdot)$ and $\phi_{ij}(\cdot,\cdot)$, the original optimization problem is equivalent to minimizing this new expression subject to the following equality constraints

$$\sum_{\alpha\in L}\phi_i(\alpha) = 1 \qquad \forall i \in V \tag{4.4.13}$$

$$\sum_{\alpha\in L}\phi_{ij}(\alpha,\beta) = \phi_i(\beta) \qquad \forall\beta\in L \quad \forall i\in V \quad \forall j\in N(i) \tag{4.4.14}$$

$$\sum_{\beta\in L}\phi_{ij}(\alpha,\beta) = \phi_i(\alpha) \qquad \forall\alpha\in L \quad \forall i\in V \quad \forall j\in N(i)\ . \tag{4.4.15}$$

$$\tag{4.4.16}$$

Since $\phi_i(\cdot)$ and $\phi_{ij}(\cdot,\cdot)$ are integer-valued this is an integer linear program (ILP) and ILPs are typically NP-hard. However, this form gives us a basis on which we can relax the problem into a continuous LP. In particular, rather than keeping $\phi_i(\cdot)$ and $\phi_{ij}(\cdot,\cdot)$ as indicator functions, we can allow them to take any non-negative value, i.e. $\phi_i(\cdot) \geq 0$ and $\phi_{ij}(\cdot,\cdot) \geq 0$. Thus the new optimization problem becomes

$$\min \sum_{i \in V} \sum_{\alpha \in L} \psi_i(\alpha)\phi_i(\alpha) + \sum_{i \in V} \sum_{j \in N(i)} \sum_{\alpha,\beta \in L} \psi_{ij}(\alpha,\beta)\phi_{ij}(\alpha,\beta) \tag{4.4.17}$$

subject to:

$$\sum_{\alpha \in L} \phi_i(\cdot) \geq 0 \qquad \forall i \in V \tag{4.4.18}$$

$$\sum_{\alpha \in L} \phi_{ij}(\cdot,\cdot) \geq 0 \quad \forall i \in V \tag{4.4.19}$$

$$\sum_{\alpha \in L} \phi_i(\alpha) = 1 \qquad \forall i \in V \tag{4.4.20}$$

$$\sum_{\alpha \in L} \phi_{ij}(\alpha,\beta) = \phi_i(\beta) \qquad \forall \beta \quad \forall i \in V \quad \forall j \in N(i) \tag{4.4.21}$$

$$\sum_{\beta \in L} \phi_{ij}(\alpha,\beta) = \phi_i(\alpha) \qquad \forall \alpha \quad \forall i \in V \quad \forall j \in N(i)\ . \tag{4.4.22}$$

The authors of [31] propose a primal-dual approximation algorithm for retrieving a solution for this program that is an f-approximation of the real optimum, meaning all instances of the problem produces a result whose value is within a factor of f of the value of the optimal solution, where $f = 2\frac{d_{max}}{d_{min}}$, $d_{max} = \max_{\alpha \neq \beta} \phi_{ij}(\alpha,\beta)$ and $d_{min} = \min_{\alpha \neq \beta} \phi_{ij}(\alpha,\beta)$.

4.5 Modelling with MRFs

In this section, we provide examples of MRFs being applied to image segmentation and image registration.

4.5.1 MRFs in Image Segmentation

In Chapter 2 Equation 2.1.4, we presented an objective function that can be used to retrieve an MRF-based image segmentation. In this section, we give further detail on this kind of image segmentation modelling while using some new concepts proposed in this chapter. Once again, we consider a binary segmentation problem. Consider an image with N pixels, let $\mathbf{Y} = \{y_i\}_{i=1}^N$ where y_i is the grayscale intensity value associated with the i^{th} pixel and consider a latent MRF $\mathbf{X} = \{x_i\}_{i=1}^N$ such that

$$x_i = \begin{cases} 1 & \text{if the } i^{th} \text{ pixel belongs to the foreground} \\ 0 & \text{if the } i^{th} \text{ pixel belongs to the background .} \end{cases} \tag{4.5.1}$$

As previously stated, specifying an MRF in a modelling exercise reduces to specifying the unitary and pairwise potentials. In knowledge-based segmentation, we use some prior knowledge about the segment classes in order to help the construction of the unitary potentials. By taking a small sample of grayscale intensity values from both classes we can estimate some simple statistics such as the mean and the variance of each segment class. Let $\{y_i^{(1)}\}_{i=1}^n$ and $\{y_i^{(0)}\}_{i=1}^m$ be samples of pixels that belong to the foreground and the background respectively. We can then estimate the following class specific parameters

$$\hat{\mu}_0 = \frac{1}{m}\sum_{i=1}^{m} y_i^{(0)} \tag{4.5.2}$$

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} y_i^{(1)} \tag{4.5.3}$$

$$\hat{\sigma}_0^2 = \frac{1}{m}\sum_{i=1}^{m} (y_i^{(0)} - \hat{\mu}_0)^2 \tag{4.5.4}$$

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i^{(1)} - \hat{\mu}_1)^2 \tag{4.5.5}$$

Now that we have some statistical information about each one of our classes, we can define the unitary potential of a pixel as a statistical distance between some attribute of the pixel and some arbitrary class. A common choice is the squared standard distance of a pixel's grayscale intensity value and a particular class. Formally the unitary potential would have the following form

$$\psi_i(\alpha) = \left(\frac{y_i - \hat{\mu}_\alpha}{\hat{\sigma}_\alpha}\right)^2 \tag{4.5.6}$$

where $\alpha \in \{0, 1\}$. Another reasonable assumption would be that neighbouring pixels should more often than not belong to the same image segment. We can incorporate this assumption by constructing a pair-wise potential that penalizes neighbours who do not have the same label. The pair-wise potentials are also known as *smoothness* terms since they allow us to restrict the space of possible configurations such that only configurations with a certain smoothness criterion are considered. One straightforward modelling choice would be to penalize neighbouring pixels that do not share the same label. We can model this assumption with the following potential

$$\psi_{ij}(\alpha, \beta) = \lambda_{ij}\delta_{ij}(\alpha, \beta) \tag{4.5.7}$$

such that

$$\delta_{ij}(\alpha, \beta) = \begin{cases} 1 & \text{if } \alpha \neq \beta \\ 0 & \text{otherwise} \end{cases} \tag{4.5.8}$$

where λ_{ij} is a hyperparameter specifying how much the pair of neighbours (i, j) are penalized for not sharing labels. If we fix λ_{ij} for all pairs, we get the following non-linear energy

$$E = \sum_{i \in V} \left(\frac{y_i - \hat{\mu}_\alpha}{\hat{\sigma}_\alpha} \right)^2 + \sum_{i \in V} \sum_{j \in N(i)} \lambda_{ij} \delta_{ij}(\alpha, \beta) \tag{4.5.9}$$

which we can then linearize following the steps in Section 4.4.5.

4.5.2 MRFs in Image Registration

Reverting to our discretized map notation of an image, i.e. $I : \Omega \to \mathbb{R}$ where Ω is a discretized image domain. We now consider two images, I and T, and want to retrieve a deformation that aligns them by warping one (the source image) onto the other (the target image). Before modelling such a problem with an MRF we first need to specify a deformation model and apply some discretization. We can express a general deformation in the following way

$$\phi(\mathbf{x}) = \mathbf{x} + T(\mathbf{x}) \quad \text{where } T(\mathbf{x}) = \sum_{\mathbf{y} \in \Omega - \{\mathbf{x}\}} \mathbf{w}(\|\mathbf{x} - \mathbf{y}\|)\mathbf{d_y} + \mathbf{w}(0)\mathbf{d_x} \, . \tag{4.5.10}$$

Thus expressing the deformation as a translation by a displacement vector $\mathbf{T}(\cdot)$ which is the result of a linear combination of displacement vectors at each point in Ω. $\mathbf{w}(\cdot)$ works as a weighting function that accounts for the contribution of each displacement vector. Therefore, specifying a $\mathbf{d_y}$ $\forall \mathbf{y} \in \Omega$ specifies a displacement vector field on the image domain Ω.

Let's further consider a discretization of the deformation space $\mathbf{D} = \{\mathbf{d^1}, ..., \mathbf{d^m}\}$, to turn this into a labelling problem, we can let $\alpha \in \{1, ..., m\}$ indicate a particular displacement vector $\mathbf{d}^\alpha \in \mathbf{D}$. Now we can rewrite the deformation at a particular point as a function of the label at that same point as follows

$$\phi(\mathbf{x}, \alpha) = \mathbf{x} + \sum_{\mathbf{y} \in \Omega - \{\mathbf{x}\}} \mathbf{w}(\|\mathbf{x} - \mathbf{y}\|)\mathbf{d_y} + \mathbf{w}(0)\mathbf{d}^\alpha_{\mathbf{x}} \, . \tag{4.5.11}$$

We could consider a unitary potential that is a weighted squared $L^2 - norm$ between the target and source image.

$$\psi_i(\alpha) = \sum_{\mathbf{y} \in \Omega} \kappa(x_i, \mathbf{y}) \|T(x_i) - I \circ \phi(x_i, \alpha)\|^2_{L^2} \tag{4.5.12}$$

where $\kappa(\cdot, \cdot)$ is some kernel distributing the error over the whole image domain. A reasonable assumption would be that the resulting labelling assigns similar displacement vectors

to neighbouring labels; this would allow us to retrieve a smooth deformation. More formally, we can incorporate this assumption in the pair-wise potential with the following regularization

$$\psi_{ij}(\alpha, \beta) = \lambda \|\mathbf{d}_i^\alpha - \mathbf{d}_j^\beta\| \tag{4.5.13}$$

where λ is a hyperparameter. Putting all of this together gives us the following energy

$$E = \sum_{i \in V} \sum_{\mathbf{y} \in \Omega} \kappa(x_i, \mathbf{y}) \|T(x_i) - I \circ \phi(x_i, \alpha)\|_{L^2}^2 + \lambda \sum_{i \in V} \sum_{j \in N(i)} \|\mathbf{d}_i^\alpha - \mathbf{d}_j^\beta\| \,. \tag{4.5.14}$$

4.6 Initial Experiments

Our work initially considered an MRF-based simultaneous image registration and image segmentation framework in which MRF inference was done through constrained linear programs implemented using GEKKO [8], a Python optimization package. The implementation of the registration component, in particular, was found to be difficult and slow to converge. The latter limitation was crucial since an iterative approach to registration and segmentation would have to perform reasonably fast to justify combining both these processes. Moreover, though we could put hard constraints on the optimization in order to guarantee a diffeomorphic registration, it was unclear to us how such deformations would compare to LDDMM which is considered the state-of-the-art for registration [29]. With the emergence of open-source deep learning registration frameworks, and in particular the recent availability of the mermaid package that efficiently implements LDDMM through automatic differentiation, we decided to take a different approach and considered a simultaneous registration and segmentation approach with a deep learning registration approach. The initial GEKKO-based MRF optimization took hours to converge and our current deep learning implementation converges under minutes.

CHAPTER 5

Joint Registration and Segmentation

The classical approach to solving medical imaging problems that require image segmentation and image registration is to apply the two procedures sequentially as part of two distinct components of a data pipeline without accounting for their co-dependencies. In recent years, within the field of neuroimaging methodology, there has been a paradigm shift towards combining these two procedures into one integrated approach. Integrating these two procedures in an iteratively circular fashion is natural since the quality of the registration can be enhanced by a pre-segmented image and the quality of the segmentation can be enhanced by a pre-registered image and therefore yielding better results than a sequential application of both procedures. The first appearance of this approach in the literature was in [5], in which the authors proposed a generative model that combines non-linear image registration, brain tissue segmentation and intensity bias correction, they called this framework *Unified Segmentation*. The registration was modelled by a low-dimensional linear parametrization of discrete cosine transforms (DCT) and has the same form as the transformation model in Equation 3.1.3 of Chapter 3, moreover, the segmentation was modelled with a Gaussian mixture model. The authors of [10] proposed a similar generative approach with the addition of a diffeomorphic registration through the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework. The main disadvantages of these integrated approaches are that the models tend to be more complex and contain more parameters and therefore they are more difficult to implement, and generally more computationally expensive during inference, especially if diffeomorphic registration is used. Although joint registration and segmentation approaches come with their own challenges and disadvantages, they can produce more accurate results than their sequential counterpart [5].

Symbol	Description
$I_i \in \mathcal{F}(\Omega, \mathbb{R})$	i^{th} Subject image in subject coordinates
$J_i \in \mathcal{F}(\Omega, \mathbb{R})$	Subject image in template coordinates
$Z_i \in \mathcal{F}(\Omega, \{0, 1\})$	i^{th} Subject lesion mask
$\phi_i \in \mathcal{F}(\Omega, \Omega)$	i^{th}) Subject deformation
$\epsilon_i \in \mathcal{F}(\Omega, \mathbb{R})$	i^{th} Subject noise
$T_i' \in \mathcal{F}(\Omega, \mathbb{R})$	i^{th} Subject noisy template
$T \in \mathcal{F}(\Omega, \mathbb{R})$	Template
$\Theta \in \mathbb{R}^p \times \mathbb{R}^p$	Lesion intensity distribution parameter with p mixtures
$\Gamma \in \mathcal{F}(\Omega, [0, 1])$	Lesion spatial prior
$\lambda \in \mathbb{R}$	Deformation prior parameter

Table 5.1: Notation. $\mathcal{F}(A, B)$ denotes the set of all functions from A to B.

5.1 Forward Model

We proposed an approach that combines image segmentation and image registration through a hierarchical Bayesian generative model. One way to understand an image data generative model is to specify the image generating process through which the images were generated; this will give insight into how high-level features (like image appearance) relate to low-level features (like pixel intensities) [27]. Conceptually, we can specify the conditional probability distribution that generates the data, this is also called the *forward model*. Moreover, the generated images can be used as an initial validation since they can be used to assess whether the generative model produces plausible images or not.

Let us consider a set of lesioned brain images from M subjects $\mathcal{D} = \{I_i\}_{i=1,\dots,M}$ and a healthy template image T such that

$$I_i : \Omega \to \mathbb{R} \tag{5.1.1}$$

and

$$T : \Omega \to \mathbb{R} \tag{5.1.2}$$

where Ω is a common coordinate system to which the subject images have been affinely co-registered beforehand. The subjects have corresponding deformations $\mathcal{G} = \{\phi_i\}_{i=1,\dots,M}$ which are sampled from the diffeomorphic group $\mathbf{Diff}(\Omega)$. ϕ_i can be viewed as a change of coordinates from subject coordinates to template coordinates or equivalently viewed as an image transformation that changes the appearance of a subject image, into the appearance of the template image; thus ϕ_i can be used to define a deterministic relationship between the subject image in template space, denoted by J_i, and the same image in subject space, denoted by I_i. J_i is the left action of ϕ_i on I_i, that is

$$J_i := \phi_i \cdot I_i \tag{5.1.3}$$

and

$$\phi \cdot I_i := I_i \circ \phi_i^{-1} \, . \tag{5.1.4}$$

Furthermore, the subjects have corresponding lesion masks, noisy templates and noise parameters denoted by $\mathcal{Z} = \{Z_i\}_{i=1,\dots,M}$, $\mathcal{T} = \{T_i'\}_{i=1,\dots,M}$ and $\mathcal{N} = \{\epsilon_i\}_{i=1,\dots,M}$ respectively. T_i' is a noisy variant of T where the noise is derived from ϵ_i. Z_i is a map that assigns a binary label to elements of Ω indicating whether they belong to a lesion or healthy tissue.

$$Z_i : \Omega \to \{0, 1\} \tag{5.1.5}$$

where label 0 and 1 indicates that the region belongs to the healthy tissue class and lesioned tissue class respectively. Moreover, Z_i depends on a lesion spatial prior Γ. Γ is a tissue

probability map that assigns a probability of belonging to the lesioned tissue class at every location in template space; this prior probability map is learned beforehand from a labelled dataset and is, therefore, a fixed model parameter.

$$\Gamma : \Omega \to [0, 1] \, . \tag{5.1.6}$$

From a generative perspective, an image in $\mathcal{D}$ is generated as follows, starting with a healthy brain template image, we generate a noisy template along with a lesion mask and following this a lesioned brain image is generated in the template space. We then sample a warp and deterministically apply this warp to the lesioned brain image to bring it in subject space. The difference in appearance across images is assumed to be accounted for by a combination of noise, lesion tissue assignments and deformations. $\mathcal{Z}$, $\mathcal{T}$ and $\mathcal{N}$ are sets of latent variables whereas T and Γ are model parameters; for a complete list of the notation refer to table 5.1. Given the model parameters and latent variables, images are conditionally independent across subjects and under the same subject, the image and deformation are also conditionally independent. More formally, $\mathcal{D}$ and $\mathcal{G}$ are random samples from random variables $\mathbf{J}$ and Φ respectively and their joint distribution probability density function has the following form

$$\mathbb{P}(\mathcal{D}, \mathcal{G}|\ \mathcal{Z}, \mathcal{N}, \mathcal{T}, T, \Theta, \Gamma, \lambda) = \prod_{i=1}^{M} p_{\mathbf{J},\Phi}(J_i, \phi_i|\ Z_i, \epsilon_i, T_i', T, \Theta, \Gamma, \lambda) \tag{5.1.7}$$

where λ is a deformation prior hyperparameter. By the conditional independence assumption across subjects, the generative process is reduced to sampling an image-deformation pair at a time. Since the image and deformation are also conditionally independent, we have the following factorization

$$p_{\mathbf{J},\Phi}(J_i, \phi_i|\ Z_i, \epsilon_i, T_i', T, \Theta, \Gamma, \lambda) = \mathbb{P}(\mathbf{J} = J_i|\ Z_i, \epsilon_i, T_i', T, \Theta, \Gamma)\mathbb{P}(\Phi = \phi_i|\lambda) \tag{5.1.8}$$

where the first factor can be further factorized in the following way

$$\mathbb{P}(\mathbf{J} = J_i|\ Z_i, \epsilon_i, T_i', T, \Theta, \Gamma) = \mathbb{P}(\mathbf{J} = J_i|\ Z_i, T', \Theta)\mathbb{P}(\mathbf{T}' = T'|T, \epsilon_i)\mathbb{P}(\mathbf{Z} = Z_i|\Gamma) \, . \tag{5.1.9}$$

The lesion mask prior distribution is modelled by a Markov random field (MRF) which we presented in Chapter 4 and its coordinates are conditionally independent given the lesion map Γ. More concretely, the probability density function has the following form

$$\mathbb{P}(Z_i(\mathbf{x}) = 1|\Gamma(\mathbf{x})) \propto \Gamma(\mathbf{x}) \exp\left\{ -\gamma \sum_{\mathbf{y} \in N(\mathbf{x})} \delta\left(Z_i(\mathbf{x}), Z_i(\mathbf{y})\right) \right\} \tag{5.1.10}$$

where $\mathbf{x} \in \Omega$ and

$$\delta(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{otherwise} \end{cases} \tag{5.1.11}$$

and where $N(\mathbf{x})$ is a sample of points in some neighbourhood of $\mathbf{x}$ and δ penalizes nearby regions that do not share the same tissue assignments, this is because lesions typically come in connected regions and γ is a hyperparameter.

Measuring noise in MRI is a complex topic (see [1]), but the "noise" in our model accounts for more than just true measurement noise, it also accounts for imperfect image registration. The noisy template is calculated as follows

$$T_i' = T + \epsilon_i \tag{5.1.12}$$

where ϵ_i is a smooth noise function drawn from a Gaussian process indexed by Ω.

$$\epsilon_i : \Omega \to \mathbb{R} \, . \tag{5.1.13}$$

A Gaussian process can be entirely characterized by its mean and covariance structure [51]; the Gaussian process in question has the following mean

$$\mathbb{E}[\epsilon_i(\mathbf{x})] = 0 \quad \text{for} \quad \mathbf{x} \in \Omega \tag{5.1.14}$$

and the following covariance structure

$$\mathbb{E}[\epsilon_i(\mathbf{x})\epsilon_i(\mathbf{y})] = \kappa(\mathbf{x}, \mathbf{y}) \quad \text{for} \quad \mathbf{x}, \mathbf{y} \in \Omega \, . \tag{5.1.15}$$

Therefore

$$\epsilon_i \sim \mathcal{GP}\big(0, \kappa(\mathbf{x}, \mathbf{y})\big) \tag{5.1.16}$$

where κ is a positive semi-definite kernel defined by

$$\kappa = \kappa_0 + \kappa_1 \tag{5.1.17}$$

where

$$\kappa_0(\mathbf{x}, \mathbf{y}) = \theta_1 \exp\big(-\theta_2 \|\mathbf{x} - \mathbf{y}\|_2^2\big) \tag{5.1.18}$$

and

$$\kappa_1(\mathbf{x}, \mathbf{y}) = \theta_3 \exp\big(-\theta_4 \|\mathbf{x}' - \mathbf{y}\|_2^2 - \theta_4 \|\mathbf{x} - \mathbf{y}'\|_2^2\big) \, . \tag{5.1.19}$$

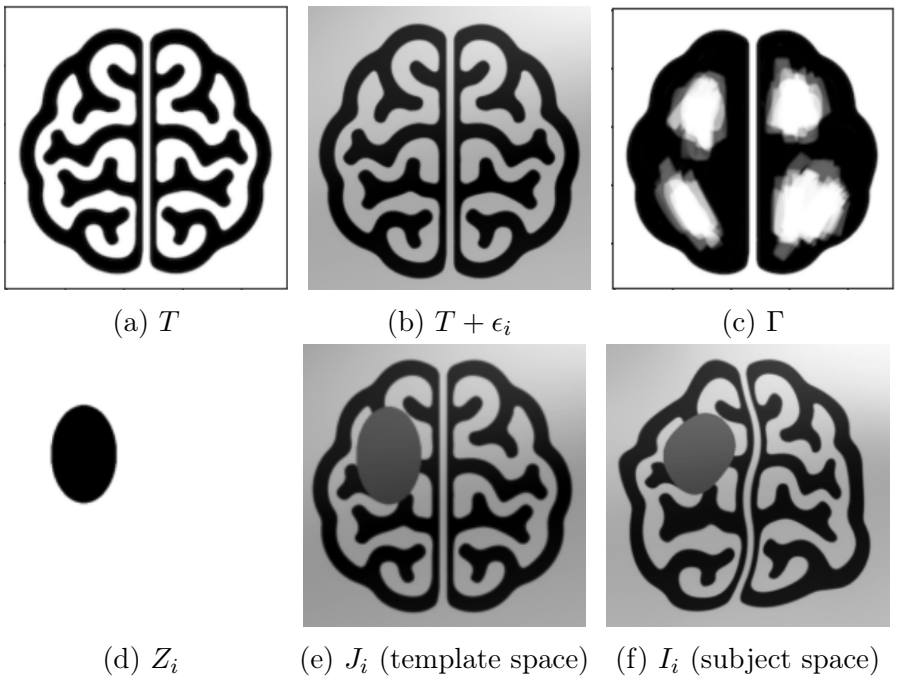
Where $\theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ are hyperparameters and $(\mathbf{x}, \mathbf{x}')$ and $(\mathbf{y}, \mathbf{y}')$ are pairs of bilaterally corresponding locations in Ω. κ_0 accounts for the smoothness of the noise and κ_1 accounts for the bilateral symmetry between the left hemisphere and the right hemisphere and assumes that the coordinates are such that $\mathbf{x}' = -\mathbf{x}$ and $\mathbf{y}' = -\mathbf{y}$.

Once we have the lesion mask and noisy template, we can generate the subject image in template space. We can then define the relationship between intensities J_i and T_i' as follows

$$J_i(\mathbf{x}) = \begin{cases} T_i'(\mathbf{x}) & \text{if } Z_i(\mathbf{x}) = 0 \\ \eta & \text{if } Z_i(\mathbf{x}) = 1 \end{cases} \qquad \text{where } \eta \sim p(\eta; \Theta) \tag{5.1.20}$$

$p(\eta; \Theta)$ is a distribution characterizing the intensity of lesion tissue.

Figure 5.2 demonstrates the image generating process at different stages under a toy template example.



(a) T (b) $T + \epsilon_i$ (c) Γ

(d) Z_i (e) J_i (template space) (f) I_i (subject space)

Figure 5.2: Forward model generating a sample from a toy template example.

5.2 Inference and Parameter Optimization

The key inference problem will be to compute the posterior probability of the latent variables and model parameters, given a set of images. We will begin our inference discussion by considering inference and latent variable estimation procedures in which the model parameters are fixed and known.

5.2.1 Simplified Model

To facilitate inference, we propose a simplified model since the one proposed in section 5.1 would be very difficult to optimize. In particular, we will be simplifying the segmentation

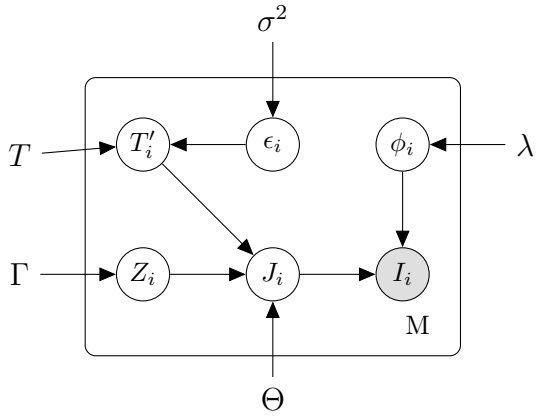


Figure 5.3: Graphical representation of the forward model. Observed variables I_i are represented by shaded nodes. Latent variables and model parameters are represented as non-shaded nodes. The rectangular plate denotes repeated variables. Symbols referring to all variables and parameters are listed in Table 5.1

and noise model. Rather than modelling the smooth noise through a Gaussian process, the noise model will be simplified to a point-wise Gaussian noise model and therefore will have the following form

$$\epsilon_i(\mathbf{x}) \sim \mathcal{N}\big(0, \sigma^2\big) \quad \text{where} \quad \mathbf{x} \in \Omega \, . \tag{5.2.1}$$

It's important to note that this noise model doesn't account for the difference in appearance across images, but rather, it provides a way for us to probabilistically evaluate the model performance. Under this simplification, we will assume that the images have been preprocessed such that they contain little to no smooth varying noise. Moreover, to simplify the lesion prior, we will remove the MRF and instead use a fixed lesion prior map Γ, with

$$\mathbb{P}(Z_i(\mathbf{x}) = 1) = \Gamma(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \Omega \, . \tag{5.2.2}$$

Using the model proposed in section 1, we can derive the following posterior distribution

$$\mathbb{P}(Z_i(\mathbf{x}) = z |\, J_i(\mathbf{x})) = \frac{\mathbb{P}(J_i(\mathbf{x}) |\, Z_i(\mathbf{x}) = z) p(Z_i(\mathbf{x}) = z)}{\mathbb{P}(J_i(\mathbf{x}))} \tag{5.2.3}$$

where

$$J_i(\mathbf{x}) |\, Z_i(\mathbf{x}) = 1 \sim \mathbf{GMM}(\cdot; \Theta) \tag{5.2.4}$$

$$J_i(\mathbf{x}) |\, Z_i(\mathbf{x}) = 0 \sim \mathcal{N}\Big(\mu = T_i(\mathbf{x}), \sigma^2\Big) \tag{5.2.5}$$

$$Z_i(\mathbf{x}) \sim \mathbf{Bernoulli}(\Gamma(\mathbf{x})) \tag{5.2.6}$$

and $\mathbf{GMM}(\cdot;\Theta)$ is a Gaussian mixture model parametrized by Θ which is the parameter in Equation 5.1.20 that characterizes the intensity of lesioned tissue. The intensities of lesioned tissue will be more likely to come from $\mathbf{GMM}(\cdot;\Theta)$ and the intensities of healthy tissue will be more likely to resemble the template. Given an intensity $J_i(\mathbf{x})$, a lesion mask can be estimated in a point-wise fashion by selecting the tissue assignment that maximizes the posterior, i.e.

$$\hat{z} = \underset{z\in\{0,1\}}{\operatorname{argmax}} \mathbb{P}(Z_i(\mathbf{x}) = z|\ J_i(\mathbf{x}))\ . \tag{5.2.7}$$

5.2.2 Maximum A Posteriori Estimation

Under the setting proposed, the simplest latent variable estimator would be the Maximum A Posteriori (MAP) estimator, i.e.

$$\left(\mathcal{G}^*, \mathcal{Z}^*, \boldsymbol{\epsilon}^*\right)_{MAP} = \underset{\mathcal{G},\mathcal{Z},\boldsymbol{\epsilon}}{\operatorname{argmax}} \prod_{i=1}^{M} \mathbb{P}(\Phi = \phi_i, \mathbf{Z} = Z_i, \boldsymbol{\epsilon} = \epsilon_i|\ \mathcal{D}, T, \Theta, \Gamma, \lambda) \tag{5.2.8}$$

where $\boldsymbol{\epsilon} = \{\epsilon_i\}_{i=1,...,M}$. We can attempt to solve the optimization in Equation 5.2.8 using a simple gradient-based algorithm. Since the parameters can naturally be partitioned into blocks, we can efficiently implement a block coordinate descent algorithm where we iteratively optimize each block of parameters while holding the rest of the parameters fixed. Let

$$l(\mathcal{G}, \mathcal{Z}, \boldsymbol{\epsilon}) = -\log \prod_{i=1}^{M} \mathbb{P}(\Phi = \phi_i, \mathbf{Z} = Z_i, \boldsymbol{\epsilon} = \epsilon_i|\ \mathcal{D}, T, \Theta, \Gamma, \lambda) \tag{5.2.9}$$

and consider the following block partition of the parameters of interest $\{\mathcal{G}, \mathcal{Z}, \boldsymbol{\epsilon}\}$; under this setting the block coordinate descent algorithm goes as follows

Algorithm 1 Block Coordinate Descent for MAP Estimation

Input: $\eta_{\mathcal{G}}, \eta_{\mathcal{Z}}, \eta_{\boldsymbol{\epsilon}}$ ▷ Parameter specific learning rate
Output: MAP estimate

procedure BLOCKCOORDINATEDESCENT($\eta_{\mathcal{G}}, \eta_{\mathcal{Z}}, \eta_{\boldsymbol{\epsilon}}$)
  $\left(\mathcal{G}, \mathcal{Z}, \boldsymbol{\epsilon}\right) \leftarrow \left(\mathcal{G}^{(0)}, \mathcal{Z}^{(0)}, \boldsymbol{\epsilon}^{(0)}\right)$ ▷ Initialize
  for $\theta \in \{\mathcal{G}, \mathcal{Z}, \boldsymbol{\epsilon}\}$ **do**
    while `<stopping criterion is not met>` **do**
      for $i \in \{1, ..., N_\theta\}$ **do**
        $\theta_i \leftarrow \theta_i - \eta_\theta \frac{\partial}{\partial \theta_i} l(\theta)$
  return $\left(\mathcal{G}, \mathcal{Z}, \boldsymbol{\epsilon}\right)$ ▷ This will be the estimate $\left(\mathcal{G}^*, \mathcal{Z}^*, \boldsymbol{\epsilon}^*\right)_{MAP}$

In Algorithm 1, $\{\eta_{\mathcal{G}}, \eta_{\mathcal{Z}}, \eta_{\boldsymbol{\epsilon}}\}$ are block parameter specific learning rates, θ_i is the i^{th} index of parameter $\theta \in \{\mathcal{G}, \mathcal{Z}, \boldsymbol{\epsilon}\}$ and N_θ is the total number of individual indices in θ. Moreover, using an expectation maximization (EM) strategy [38] we can simultaneously infer the latent variables and learn model parameters.

5.2.3 Implementation

The current implementation assumes the model parameters λ, σ^2 and Θ are known and uses a 2-block coordinate descent to infer the deformation parameters (block 1) and segmentation parameters (block 2). Θ is the parameters of a 2-Gaussian mixture model that characterizes the lesion intensity distribution and is learned from a training set (see Section 6.2) before applying our algorithm. σ^2 is a parameter governing the noise of the template and was set to 0.05 and had to be chosen to be low enough such that that the intensity values in the subject and template images match well, while allowing our segmentation algorithm to infer mismatching regions. λ is a weight associated with the deformation regularizer, in terms of the deformation model it is equivalent to the free parameter β presented in Chapter 3, Equation 3.2.17; larger values of λ correspond to smoother deformations. Its value was set to 2, we found that this value allowed for deformations that effectively warp one image onto another, while being reasonably smooth.

Deformation inference

The image registration component is implemented with the help of a registration toolkit named `mermaid` that implements registration routines through deep convolutional neural networks (CNN) and automatic differentiation [64,65,41,54, 55]. We modified the source code such that the registration networks could account for a lesion mask. More specifically, we added an option to allow the calculation of the objective function to ignore pixels/voxels in a given mask. The network takes the subject, the template and a lesion mask as inputs and has a deep encoder-decoder architecture (see Figure 5.4) that infers a LDDMM modelled deformation through the prediction of an initial momentum $m = Lv$, where v is the velocity field discussed in Chapter 3 and in Equation 3.1.6 and L is a differential operator chosen such that v is sufficiently smooth and the regularizer norm has the following form

$$\|v\|_L^2 := \langle v, Lv \rangle_L \, . \tag{5.2.10}$$

In `mermaid` the operator L is approximated through a smoothing filter constructed from multiple Gaussians. This network efficiently approximates the geodesic calculation while retaining the theoretical properties of LDDMM and the Bayesian probabilistic version (implemented using dropout) provides deformation field uncertainty and it is trained with accelerated stochastic gradient descent [40]. We will refer to it as the *registration network* and denote it as follows

$$LDDMMShootingVectorMomentumMapNet : (I, T, Z, m^0, \lambda) \rightarrow (\phi, m) \tag{5.2.11}$$

where I, T, Z, m^0, λ, ϕ and m are the subject, the template, the lesion mask, the initial momentum, the deformation prior parameter, the deformation and the predicted momentum, respectively. The name of the network is taken from the `mermaid` software package. After every iteration we store the current momentum and before every iteration we use the previous momentum as our initial starting point.

Segmentation Inference

The image segmentation component implements a MAP estimate of the posterior in Equation 5.2.3; this process is conditioned on the warped subject image, the lesion intensity parameter and lesion spatial prior. Since Θ and σ^2 are not learned at runtime, the estimate can be retrieved through direct function calls of the posterior. We will refer to this component as the *mask estimation* and denote it as follows

$$SegEst : (I \circ \phi, T, \Theta, \Gamma, \sigma^2) \rightarrow Z \tag{5.2.12}$$

where $I \circ \phi$, T, Θ, Γ and Z denote the warped subject, the template, the lesion intensity parameter and lesion spatial prior, respectively. As a post-processing step to the segmentation inference, we considered the application of two morphological image processing techniques, erosion and dilation [60]. We sequentially apply two erosion and dilation steps to the estimated masks. The erosion operations help remove small disconnected regions in the segmentation, that are likely false positives. Since our work considers segmentations of largely connected subregions of an image, dilation operations help connect large regions in the mask which would help reduce false negatives. This computational step is meant to be a placeholder for a later principled use of MRF segmentation (see Chapter 7).

Inference Algorithms

To assess the performance of our simultaneous registration and segmentation inference approach, we considered various algorithms that use different components of our approach, below we define these algorithms.

Definition 5.2.1 (**Reg**). *An image registration algorithm that infers a diffeomorphic deformation that warps a subject onto a template without segmentation information. This algorithm infers the deformation through the LDDMMShootingVectorMomentumMapNet presented in Section 5.2.3.*

Definition 5.2.2 (**Reg-True-Mask**). *An image registration algorithm that infers a diffeomorphic deformation that warps a subject onto a template with the use of a ground truth segmentation. This algorithm infers the deformation through the LDDMMShootingVectorMomentumMapNet presented in Section 5.2.3.*

Definition 5.2.3 (**Seg**). *An image segmentation algorithm that infers a segmentation a subject image. This algorithm infers the latent segmentation through a MAP estimate of the segmentation model proposed in Equation 5.2.3.*

Definition 5.2.4 (**RegSeg**). *A simultaneous registration and segmentation inference algorithm. The algorithm iteratively infers the deformation and segmentation such that each registration step utilizes current segmentation information and each segmentation step utilizes current deformation information. This algorithm infers the deformation through the LDDMMShootingVectorMomentumMapNet presented in Section 5.2.3 and infers the latent segmentation through a MAP estimate of the segmentation model proposed in Equation 5.2.3 without any post-processing morphological operations.*

Definition 5.2.5 (**RegSeg-Morph**). *A simultaneous registration and segmentation inference algorithm. The algorithm iteratively infers the deformation and segmentation such that each registration step utilizes current segmentation information and each segmentation step utilizes current deformation information. This algorithm infers the deformation through the LDDMMShootingVectorMomentumMapNet presented in Section 5.2.3 and infers the latent segmentation through a MAP estimate of the segmentation model proposed in Equation 5.2.3 while sequentially applying two erosion and dilation steps to the estimated segmentation as a post-processing step.*

We ran some initial tests (see Figure 5.4) and on this basis we chose 25 iterations. In future it would be interesting to run for a much large number of iterations or to use a stopping condition based on convergence of the registration and segmentation metrics.

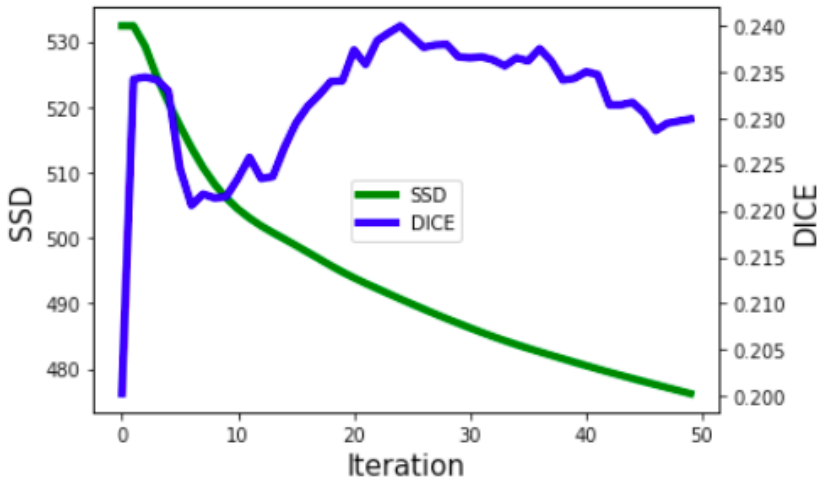


Figure 5.4: Segmentation and registration performance for subject with ID = 2 from the dataset presented in Chapter 6.

Since all the algorithms above could be understood as a sub-routine of *RegSeg-Morph*, we give its pseudo code below. A graphical representation of this algorithm is displayed in figure 5.5.

Algorithm 2 RegSeg-Morph

Input: $\mathcal{D}, T, \Theta, \Gamma, \lambda, \sigma^2$

Output: Deformation and segmentation estimates

procedure REGSEG-MORPH($\mathcal{D}, T, \Theta, \Gamma, \lambda, \sigma^2$)
  for $i \in \{1, ..., |\mathcal{D}|\}$ **do**
    $t \leftarrow 0$ ▷ Initialize iteration counter
    while $t < 25$ **do**
      $Z_i \leftarrow SegEst(I_i, \phi_i, T, \Theta, \Gamma, \sigma^2)$
      $Z_i \leftarrow MorphologicalErosion(Z_i, 2)$
      $Z_i \leftarrow MorphologicalDilation(Z_i, 2)$
      $Z_i \leftarrow SegEst(I_i, T, \Theta, \Gamma, \sigma^2)$
      $(\phi_i, m_i) \leftarrow LDDMMShootingVectorMomentumMapNet(I_i, Z_i, T, m_i, \lambda)$
      $t \leftarrow t + 1$
  return $\{(\phi_j, Z_j)\}_{j=1}^{|\mathcal{D}|}$

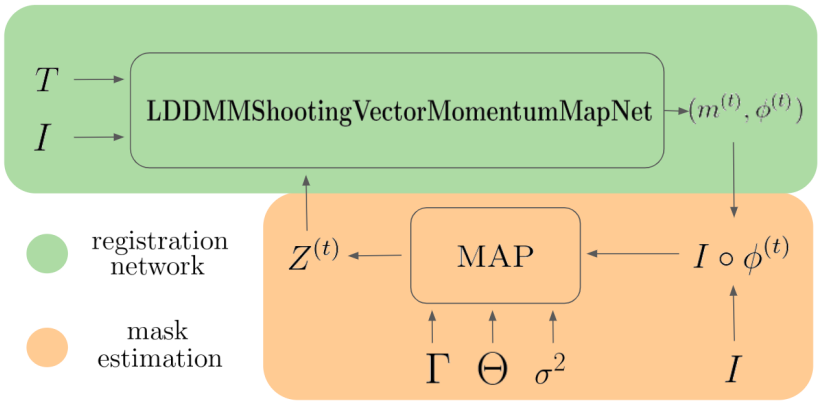


Figure 5.5: A graphical representation of the RegSeg algorithm.

5.3 Synthetic Experiments

In this section, we demonstrate some applications of RegSeg on synthetic data. We first start by demonstrating some general examples, that are not tailored to a specific application, in which a simple geometric object is registered to another while simultaneously retrieving a segmentation of a subregion of the subject image. Secondly, we will apply the algorithm to a cartoon image of a lesioned and template brain. These initial experiments were conducted to demonstrate different components of the algorithm. Each application consists of a template image, a subject image and a corresponding ground truth binary seg-

mentation of the subject image. The segmentation is the ground truth label we are trying to infer. In the lesioned brain registration literature the lesion segmentation is sometimes referred to as a *lesion mask* or simply a *mask* and assigns a label of 0 and 1 to lesion and healthy tissue respectively. This is because the segmentation is typically used to mask out the lesion tissue's contribution to the similarity measure during registration. To be consistent with Chapter 6, we will borrow this terminology and refer to the ground truth segmentation as a true mask and we will also refer to the region in the subject image segmented by the mask as a lesion. In all experiments, the datasets include the true masks, which we use in our evaluations.

To evaluate the registration performance, we will consider 3 scenarios, registration with no mask, registration with the ground truth mask and registration under the simultaneous estimation algorithm. Comparing the first and second scenarios allows us to evaluate and demonstrate how much masking image abnormalities improves registration performance. Comparing the latter two scenarios allows us to evaluate the registration component of our approach against registration under the knowledge of the ground truth segmentation.

For each scenario, once the optimization is complete and a transformation ϕ is retrieved, we will evaluate how well it aligns the two images by pulling the template image into subject coordinates via ϕ^{-1} and therefore all evaluations will take place in subject space. Once in subject coordinates, the images will be compared by taking a sum of squared differences (SSD) masked with the ground truth segmentation and will be calculated as follows

$$SSD = \sum_{\mathbf{x}\in\Omega} \Big(1 - Z(\mathbf{x})\Big)\Big(I(\mathbf{x}) - T \circ \phi^{-1}(\mathbf{x})\Big)^2 \tag{5.3.1}$$

where Ω is the discretized image domain, Z is the ground truth lesion mask, I is the subject image and $T \circ \phi^{-1}$ is the template image in subject coordinates. We expect that the registration will underperform over healthy tissue that is near the lesion. Since lesion tissue accounts for a small proportion of total tissue, an SSD over the entire healthy tissue could potentially lead to a falsely perceived similar performance across all 3 scenarios.

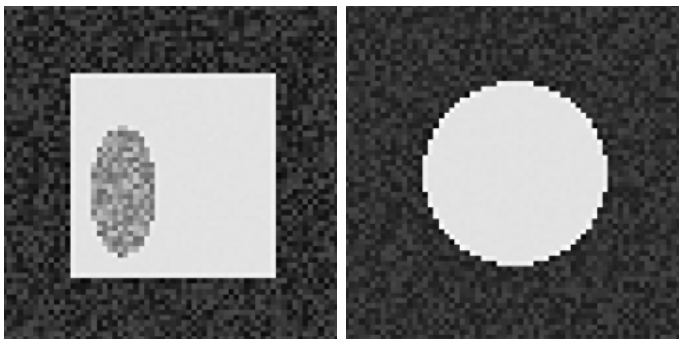
To overcome this effect, we additionally report SSD over a region of healthy tissue near the lesion boundary. This region is acquired by morphologically dilating the lesion by 30 pixels and then removing the original lesion area thus leaving only the dilated region.

These evaluations could be performed in either coordinates but given that the ground truth lesion mask is canonically defined in subject coordinates, we decided to perform them in subject coordinates so that only the template image needs to be warped. Otherwise, we would have to warp both the subject image and the ground truth lesion mask into template space.

To evaluate the segmentation performance, we will consider 2 scenarios, segmentation with and without RegSeg. The first infers a segmentation estimate under our segmentation model, without iteratively being conditioned on the registration estimates; we will refer to this model as *Seg*. The second scenario iteratively updates the segmentation estimate while accounting for the evolving deformation results from the registration component. Comparing performance across both scenarios allows us to evaluate the performance gained between sequential segmentation and simultaneous segmentation that accounts for registration results. For each scenario, the Dice score (introduced in Chapter 2) of the estimated and true masks will be computed.

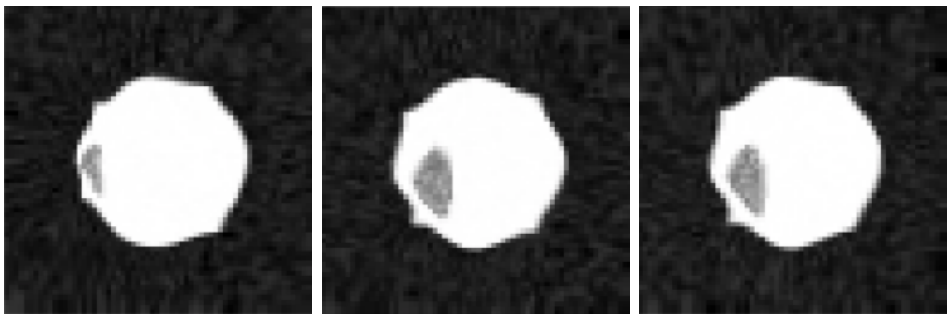
5.3.1 Experiment 1: Shapes

The first experiment considers warping a square onto a circle with the square having a subregion that forms an ellipse and has no counterpart on the circle. This simple example helps demonstrate one of the main challenges of registering two images in the presents of severe mismatch of subregions. Since the mismatched region in the subject image does not have a natural match in the template image, the registration will try to shrink the region resulting in a deformation field that is not smooth near the region boundary, this is portrayed in Figure 5.9 (a).

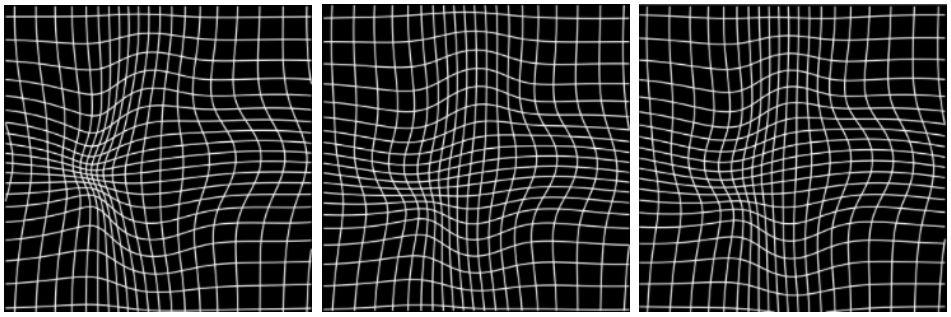


(a) Subject (b) Template

Figure 5.7: Warping a square onto a circle. The square hypo-intense pixels form an ellipse.

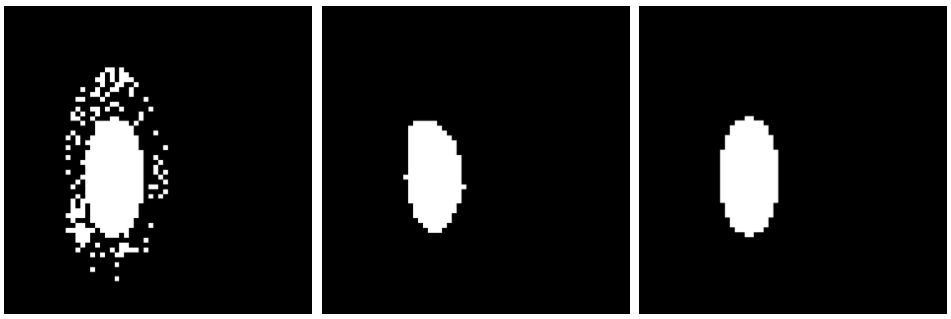


(a) $I \circ \phi$ (Reg) (b) $I \circ \phi$ (RegSeg-True-Mask) (c) $I \circ \phi$ (RegSeg)



(d) ϕ (Reg) (e) ϕ (RegSeg-True-Mask) (f) ϕ (RegSeg)

Figure 5.9: The first row displays the warped subject images ($I \circ \phi$) and the second row displays the corresponding warps (ϕ).



(a) Mask estimate (Seg) (b) Mask estimate (RegSeg-Morph) (c) True mask

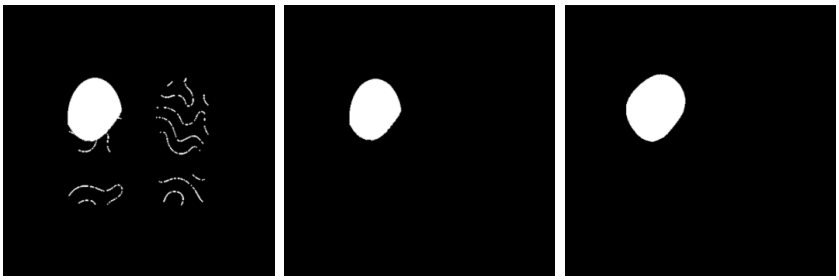
Figure 5.11: The first mask is estimated with the *Seg* algorithm; meaning, under the segmentation model but without iteratively being conditioned on the registration estimates. The second mask is estimated with the *RegSeg* algorithm; mean it is iteratively estimated under the segmentation model while accounting for the evolving deformation results from the registration component.

5.3.2 Experiment 2: Cartoon Brain

The second experiment considers warping a lesioned cartoon brain onto a symmetric cartoon template. The segmentation model depends on the intensity dissimilarity between subject and template. Early in the registration this dissimilarity is higher since the misaligned gyri and sulci lead to overlapping grey and white matter, this is portrayed in Figure 5.15 (a).

(a) Subject (b) Template (c) Warped subject

Figure 5.13: Warping a cartoon brain onto a cartoon template. Both images have added smooth noise and the subject image has hypo-intense pixels to simulate a lesion.



(a) Mask estimate (Seg) (b) Mask estimate (RegSeg-Morph) (c) True mask

Figure 5.15: The first mask is estimated under our the segmentation model but without iteratively being conditioned on the registration estimates. The second mask is estimated iteratively while accounting for the evolving deformation results from the registration component.

Experimental Results

Application	Region	Reg	Reg-True-Mask	RegSeg
Shapes	*Whole image*	31.0	25.3	26.1
	Near boundary	3.18	1.98	2.01
Cartoon brain	*Whole image*	2100	1870	1900
	Near boundary	167	100	113

Table 5.2: Registration results (lower is better). Given is the sum of squared differences (SSD) between the source image in subject coordinates, i.e. I, and the template image in subject coordinates, i.e. $T \circ \phi^{-1}$. SSD is computed outside the segmentation mask this corresponds to the healthy tissue in the case of lesion segmentation. Moreover, it is computed for two distinct regions, one near the segmentation boundary and one over the entire image.

Application	Seg	RegSeg	Percentage Increase
Shapes	0.7934	0.9204	16.01%
Cartoon brain	0.8817	0.9756	10.65%

Table 5.3: Segmentation results (higher is better). Given is the Dice score between true and estimated lesion region, estimated using two algorithms *Seg* and *RegSeg*.

CHAPTER 6

Experiments on Neuroimages

In this final chapter, we demonstrate experiments involving the application of RegSeg to a medical neuroimaging ischemic stroke lesion dataset. We will use terminology consistent with Chapter 5 Section 3 and refer to pairs of images as *subject* and *template* and refer to the ground truth segmentation as a true mask. To evaluate the registration performance, we will follow the same protocol as in Chapter 5 Section 3, we will consider 3 scenarios, registration with no mask, registration with the ground truth mask and registration with RegSeg and all evaluations will take place in subject space using a sum of squared differences (SSD) masked with the ground truth segmentation. In the MR application, this means that we will warp the healthy template image onto the subject image and compare how well the healthy tissue regions align. The SSD will again be computed over the entire healthy tissue and a region of healthy tissue near the lesion boundary. Evaluating the segmentation performance will also follow the same protocol as in Chapter 5 Section 3, in which we compare segmentation estimate results acquired through different algorithms with the ground truth segmentation through their Dice score.

6.1 Dataset

The final experiment considers warping a stroke lesion subject image onto a healthy symmetric template image. The subject images we used were collected through a clinical routine acquisition of a 28 subject sub-acute ischemic stroke lesion segmentation (SISS) dataset from the 2015 Ischemic Stroke Lesion Segmentation (ISLES) [34], a medical image segmentation challenge at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). This segmentation challenge aims to provide a platform for a fair and direct comparison of methods for ischemic stroke lesion segmentation from multi-spectral MRI images. Each sample was segmented by 1 expert and included many MRI sequences, FLAIR, T2w TSE, T1w TFE/TSE and DWI; our algorithm was evaluated on the T1w. All MRI modalities were skull-stripped, re-sampled to an isotropic spacing of 1^3mm and co-registered to the FLAIR modality. Axial slices of a sample of 6 subjects along with their corresponding lesion masks are displayed in Figure 6.1.

The template is a symmetric template acquired from UCLA's International Consortium for Brain Mapping (ICBM) centre. The template is an unbiased non-linear average of images from the MNI152 database and has been generated in a way that combines the attractions of both high spatial resolution and signal-to-noise while not being subject to the specificities of any single brain [20]. The template was also re-sampled to an isotropic spacing of 1^3mm and included T1w, T2w and PDw modalities; we again used the T1w modality. A pair of subject and template axial image slices is displayed in Figure 6.3.

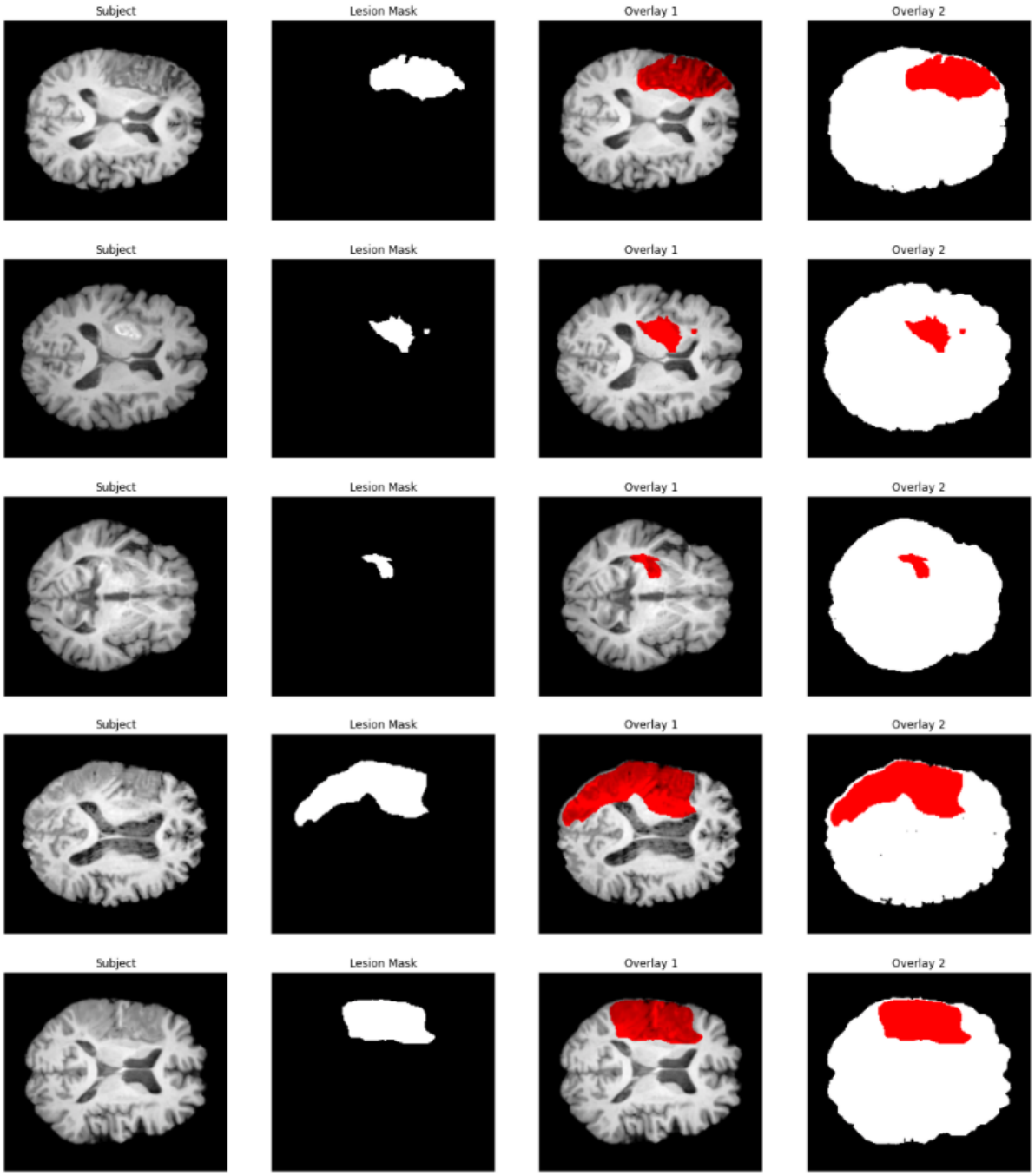
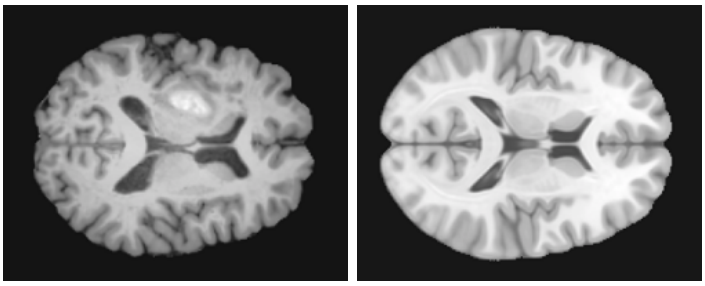


Figure 6.1: Axial slices from 5 subjects, with corresponding lesion masks, from the 2015 Ischemic Stroke Lesion Segmentation (ISLES) dataset [34]. The two columns to the right display brain-lesion overlays with two different translucencies.

Lesion volume	$\mu = 17.59ml \;\; [0.001, 346.064]$
Haemorrhage present	$n = 12 \;\; 0 = yes, 1 = no$
Lesion localization (lobes)	$n_1 = 11, n_2 = 24, n_3 = 42, n_4 = 17, n_5 = 2, n_6 = 6 \;\; 1 = frontal, 2 = temporal, 3 = parietal, 4 = occipital, 5 = midbrain, 6 = cerebellum$
Lesion localization (cortical / subcortical)	$n_1 = 36, n_2 = 49 \;\; 1 = cortical, 2 = subcortical$
Affected artery	$n_1 = 6, n_2 = 45, n_3 = 11, n_4 = 5, n_5 = 0 \;\; 1 = ACA, 2 = ACM, 3 = ACP, 4 = BA, 5 = other$
Midline shift observable	$n_0 = 51, n1 = 5, n2 = 0 \;\; 0 = none, 1 = slight, 2 = strong$
Ventricular enhancement observable	$n_0 = 38, n1 = 15, n2 = 3 \;\; 0 = none, 1 = slight, 2 = strong$

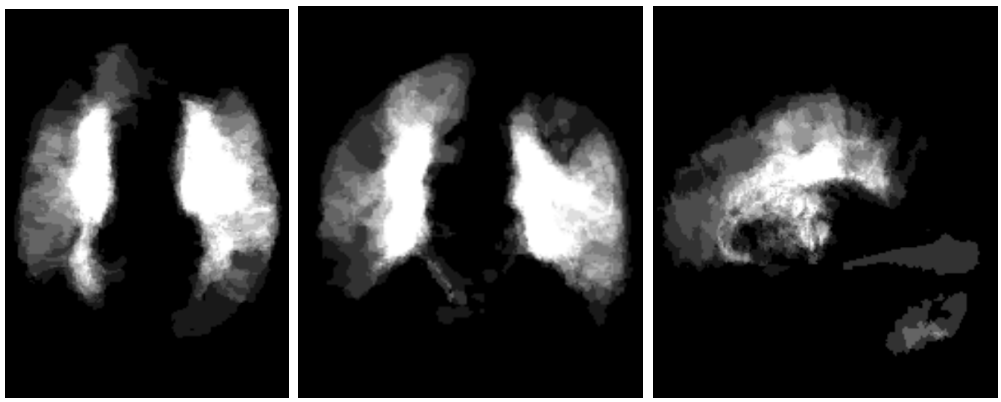
Table 6.1: For SISS, cases with sub-acute ischemic stroke lesions have been collected from clinical routine acquisition. The following table of characteristics reflect the diversity of the SISS dataset. Table taken from [34, Tbl. 3].



(a) Subject (b) Template

Figure 6.3: Subfigure (a) is a subject image from a sub-acute ischemic stroke lesion (SISS) dataset and subfigure (b) is a symmetric template acquired from UCLA's International Consortium for Brain Mapping (ICBM) centre.

We also used a stroke lesion probabilistic spatial mapping that was constructed by averaging over all samples from a dataset of 304 MRI images from 11 cohorts worldwide which were collected from research groups in the ENIGMA Stroke Recovery Working Group consortium [33]; the images consisted of T1w MRIs of individuals post stroke. The spatial prior was normalized to the MNI-152 template and therefore was in template coordinates [32].



(a) Axial view (b) Coronal view (b) Sagittal view

Figure 6.5: A stroke lesion probabilistic spatial mapping that was constructed by averaging over all samples from a dataset of 304 MRI images from 11 cohorts worldwide which were collected from research groups in the ENIGMA Stroke Recovery Working Group consortium [33]. Brighter regions are more probable regions.

6.2 Preprocessing

The template and subject images have very different intensity distributions, as can be seen in Figure 6.3 and 6.4. This is to be expected since MR intensity has no intrinsic anatomical interpretation and instead varies between different MR scanners and scanner protocols. We normalize the intensities in a preprocessing step, by applying a piecewise affine histogram matching algorithm proposed by [46]. To do this, we first calculate a standard histogram that is fitted to the template through intensity histogram percentiles at $1, 10, 20, ..., 90, 99$ percent, where the intensity values below 1% and above 99% are discarded as outliers and the 1% tail-ends are linearly interpolated. We then normalize the subject image by matching its intensity histogram to the standard histogram. To do this, we first calculate the same set of percentiles and use them to segment the image histogram into deciles. The deciles are then piecewise linearly mapped to the corresponding intensities in the standard histogram's scale landmarks. Pre-normalization and post-normalization subject and template intensity histograms are displayed in Figure 6.4.

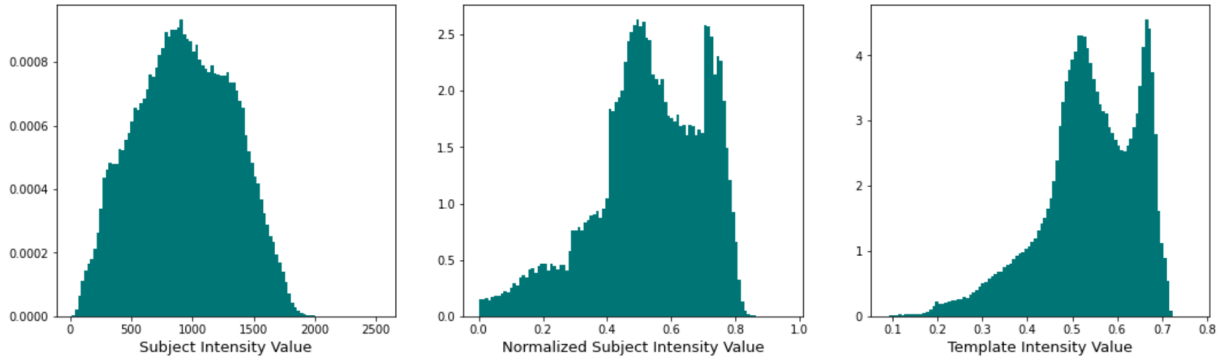


Figure 6.6: Subject intensity value normalization through a piecewise affine histogram matching algorithm [46].

Once the subject image intensities are normalized, we used the dataset as a training set to learn Θ, the parameter that characterizes the lesion intensity distribution introduced in Chapter 5, Section 1, Equation 5.1.20. To estimate this parameter, we modelled the lesion distinctions as a mixture of two Gaussians. Figure 6.5 shows the frequency histogram of the lesion intensity values.

Originally, we intended to learn Θ during the *RegSeg* runtime but due to implementation difficulties and in the interest of time we decided to calculate it directly from the dataset using the ground-truth lesion masks by setting them to the sample mean and covariance of all pixels labelled as lesion in the dataset. Because of this, the segmentation component of our method is not entirely unsupervised, however, the other components of our method may be considered unsupervised since they do not use the the labels to learn the locations of lesions, and the lesion intensity parameters could have been determined from prior knowledge (other datasets) as we did for the lesion probability map prior Γ. We acknowledge that calculating this parameter from the validation set is advised against, however, this is the only parameter learned from this dataset, and we do not believe that we would have gotten very different results if the lesion intensity distribution were determined in another way, and for later publication, we will carefully consider a Training/Validation/Test split.

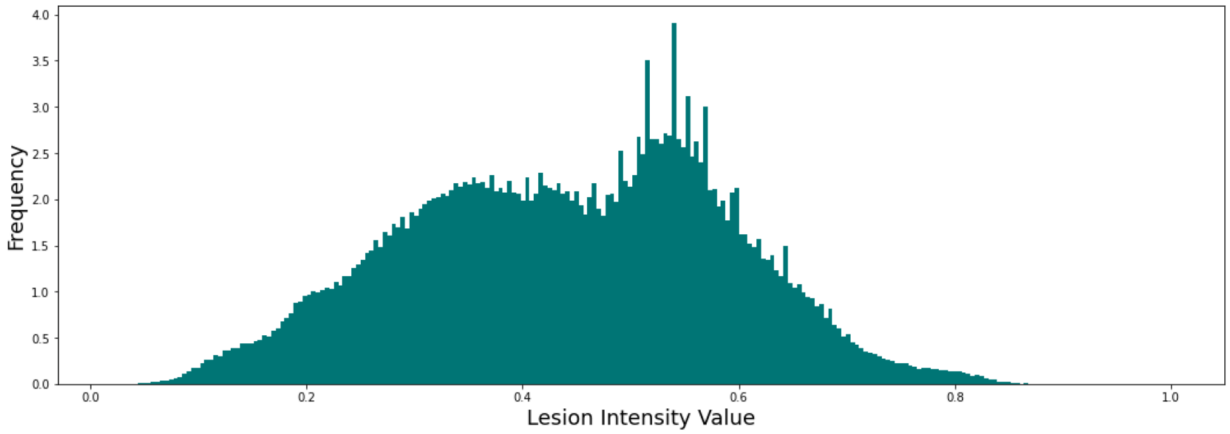
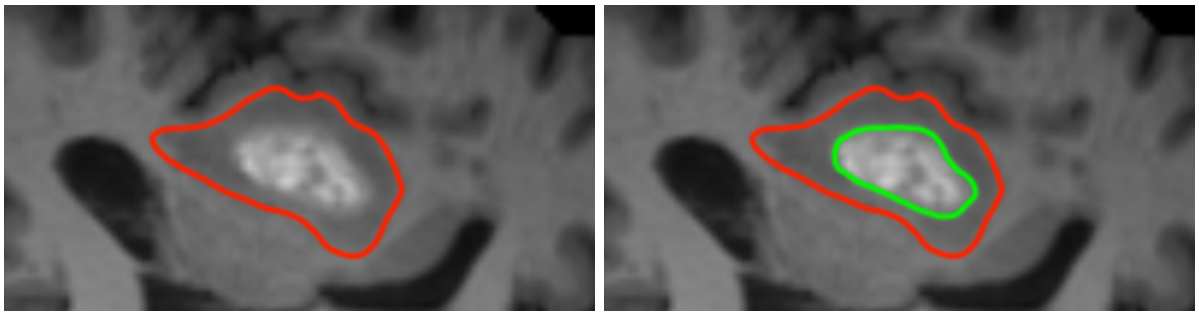


Figure 6.7: The frequency histogram of the lesion intensity values once the dataset was normalized.



(a) Lesion boundary (b) Lesion boundary

Figure 6.9: The lesion has a hyper-intense inner region that is outlined in green and a hypo-intense outer region that is between the red and green outline. Sub-figure (a) displays the entire lesion boundary.

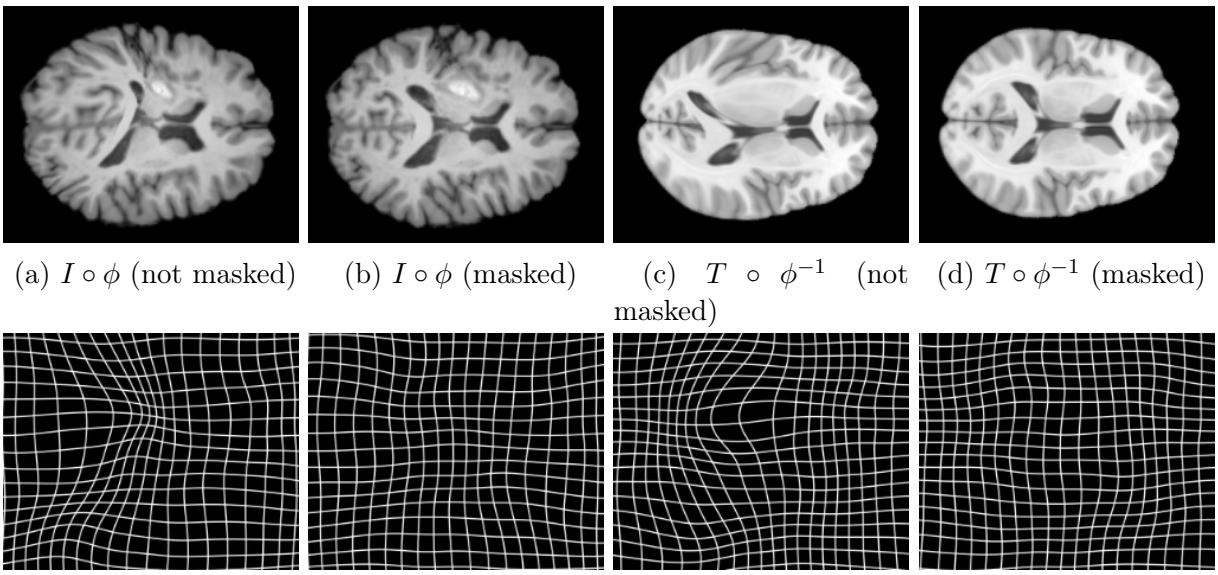
Finally, the subject images were re-sampled to match the size of the template and then affinely registered to the template. Once the subject images were co-registered to the template, a 2D axial slice is taken from all subjects and template to in order to get a 2D dataset to perform our proof-of-concept experiments, the size of each image was 197×220 pixels. We excluded 3 subjects from the dataset whose lesion was in the cerebellum since this location made it hard to retrieve an axial slice that was comparable to the other subjects, leaving 25 subject images.

6.3 Results

Images (a) and (b) in Figure 6.9 demonstrates the shrinkage distortion due to the absence of a lesion mask, just as observed in the geometric shapes experiment presented in Chapter 5, Section 3. The mask effectively behaves as an outlier remover. Images (c) and (d) in Figure 6.9 demonstrates the inverse behaviour, when applying the inverse warp to bring the template into subject coordinates, the shrinkage corresponds to a hyper expansion of the same region.

As seen in Figure 6.13 (c) and (f), the mask estimate retrieved through *RegSeg-Morph* was found to be generally smaller than the true mask and is concentrated in the middle of it. This is due to some lesions having hyper-intense inner regions and hypo-intense outer regions, as displayed in Figure 6.7. This poses a challenge to the segmentation model since the outer region closely resembles white matter and modelling this particular subtlety through intensity alone would require learning a highly precise multi-modal lesion distribution. Moreover, some geometric information could help estimate a larger and more accurate lesion boundary. In addition to the mismatch intensities of lesion tissue, the presence of a lesion geometrically distorts nearby structures. For instance in Figure 6.13 (d) and (e), the presence of the lesion in the left cerebral hemisphere pushed against the left lateral vesicle, making the left frontal horn smaller than the right one.

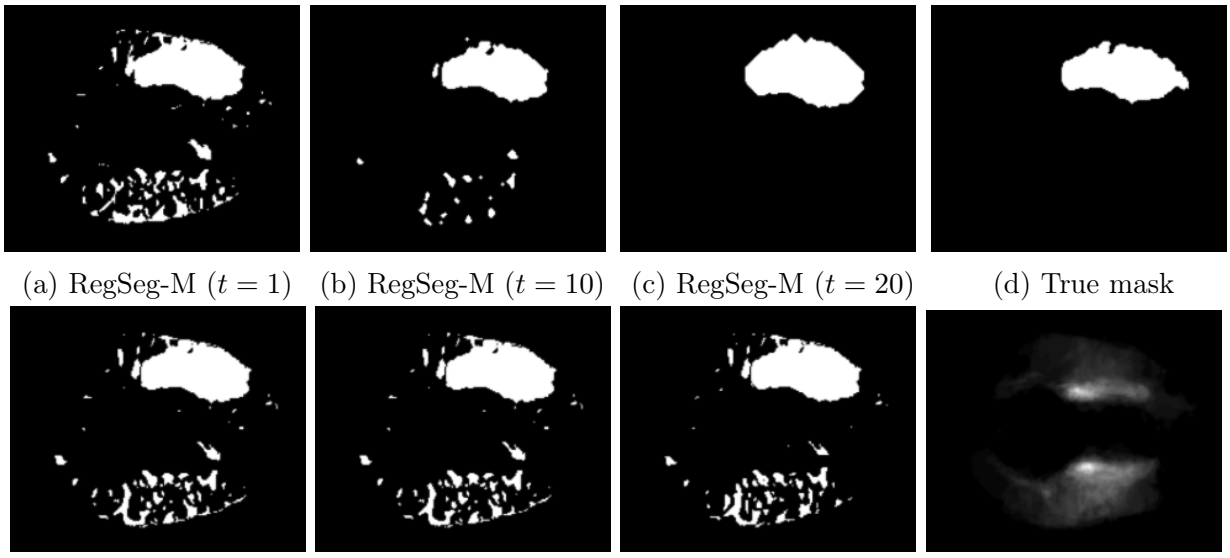
Figures 6.11 (a) - (c) display how the mask estimates evolve over iterations and converges to a region around the lesion. Just as in the second experiment, we can observe that early in the registration, the misaligned grey and white matter corrupts the mask estimate and this is corrected for as the registration goes forward. It's worth noting that the mislabelling due to overlapping grey and white matter creates many disconnected regions that are significantly smaller than the true lesion area. A segmentation regularization that penalizes adjacent pixels that do not share labels would correct for this lack of size and connectivity and such a regularization could be naturally modelled through an MRF, effectively removing false positives.



(a) $I \circ \phi$ (not masked) (b) $I \circ \phi$ (masked) (c) $T \circ \phi^{-1}$ (not masked) (d) $T \circ \phi^{-1}$ (masked)

(e) ϕ (not masked) (f) ϕ (masked) (g) ϕ^{-1} (not masked) (h) ϕ^{-1} (masked)

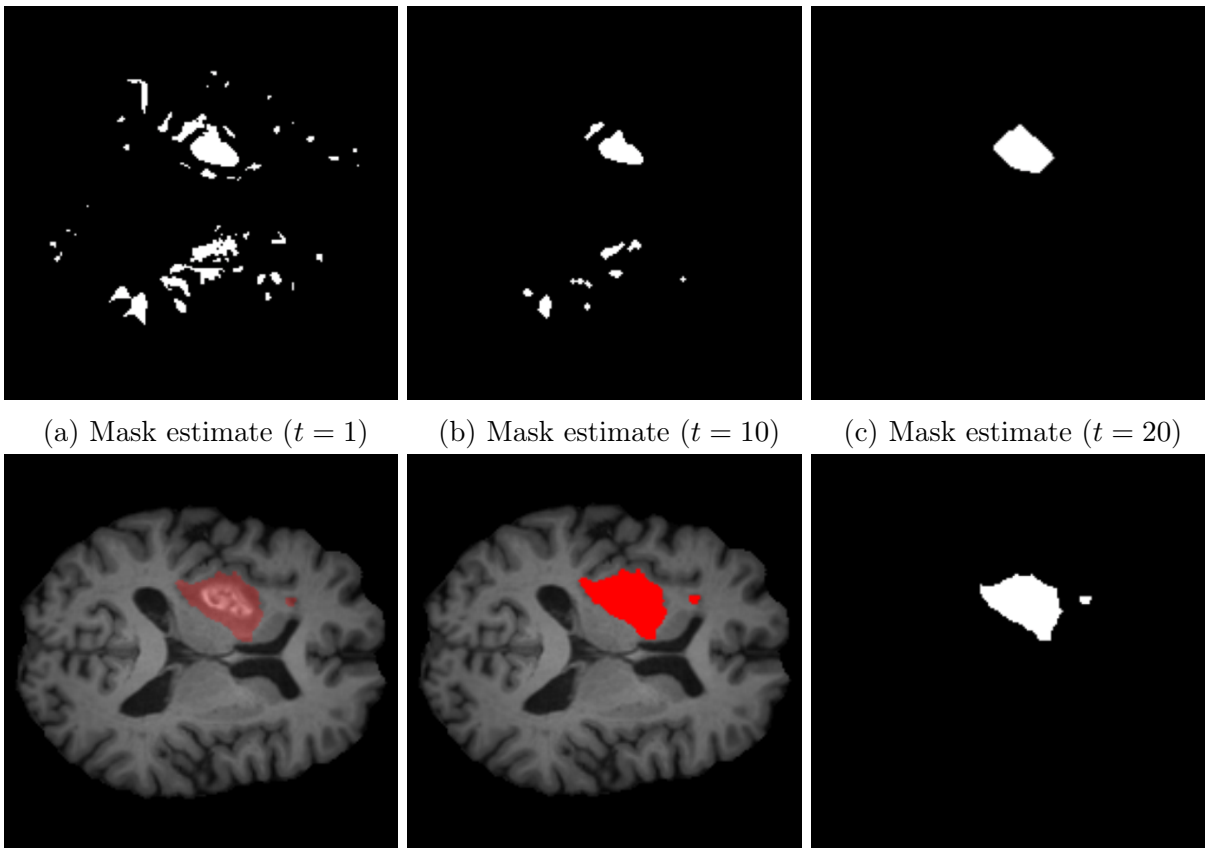
Figure 6.11: The only difference between the warped images in (a) and (b) and the warped images in (c) and (d) is the presents of a mask. Every other parameter was held fixed.



(a) RegSeg-M ($t = 1$) (b) RegSeg-M ($t = 10$) (c) RegSeg-M ($t = 20$) (d) True mask

(e) RegSeg ($t = 1$) (f) RegSeg ($t = 10$) (g) RegSeg ($t = 20$) (h) Lesion prior

Figure 6.13: Subfigures (a), (b) and (c) display mask estimates using the RegSeg-Morph algorithm and at different iterations. Subfigures (e), (f) and (g) display mask estimates using the RegSeg algorithm and at different iterations. The images use subject ID = 1.



(a) Mask estimate ($t = 1$) (b) Mask estimate ($t = 10$) (c) Mask estimate ($t = 20$)

(d) Lesion/brain overlay (e) Lesion/brain overlay (f) True mask

Figure 6.15: Subfigures (a), (b) and (c) display mask estimates using RegSeg-Morph and at different iterations. The subfigures (d) and (e) display the lesion mask overlaid on the corresponding brain with different translucencies. The images use subject ID = 2.

For the remainder of this section we present experimental results. See Tables [6.2, 6.3, 6.4, 6.5] and Figure 6.16. We observed a statistically significant improvement in performance, for both registration and segmentation, when our algorithm is used as opposed to a sequential application of registration and segmentation. In segmentation experiments in particular, *RegSeg-Morph* improved the average Dice score by 200%; the performance between *RegSeg-Morph* and *RegSeg* was not statistically significant. For the registration results, the performance around the region near the lesion boundary was particularly improved, see *"near lesion"* results in Table 6.3.

For the segmentation results (see Table 6.2), the reader should bear in mind that our work considered the performance gained from the addition of registration results to a simple segmentation algorithm rather than a state-of-the-art lesion segmentation method. Our segmentation model is very primitive compared to the current state-of-the-art segmentation models which are typically U-net-like CNNs and yield Dice scores around 0.80-0.90 [22].

Our segmentation could be greatly improved by considering approaches that combine more complex CNN models and MRF-based segmentation. Moreover, we could further consider topological post-processing to construct a more uniform segmentation.

Algorithm	**Mean** *Dice*	**Median** *Dice*	**Minimum** *Dice*	**Maximum** *Dice*	**Average run time**
Seg	0.1603	0.162	0.0147	0.331	28 seconds
RegSeg	0.276	0.264	0.062	0.436	8.5 minutes
RegSeg-Morph	**0.298**	**0.291**	**0.1**	**0.467**	8.9 minutes

Table 6.2: Segmentation results (higher is better). Given are different Dice scores for the *Seg*, *RegSeg* and *RegSeg-Morph* algorithms. The Average run times were calculated for 1 iteration of *Seg* and 25 iterations of *RegSeg* and *RegSeg-Morph*.

Algorithm	**Region**	**Mean** *SSD*	**Median** *SSD*	**Minimum** *SSD*	**Maximum** *SSD*
Reg	*Whole image*	516.4	514.34	388.4	688.9
	Near lesion	222.9	213.5	137.9	311.9
Reg-True-Mask	*Whole image*	381.6	363.7	243.3	512.9
	Near lesion	111.5	119.09	23.55	186.01
RegSeg	*Whole image*	473.8	499.6	319.8	636.8
	Near lesion	129.13	130.66	24.2	216.3

Table 6.3: Registration results (lower is better). Given is the sum of squared differences (SSD) between the source image in subject coordinates, i.e. I, and the template image in subject coordinates, i.e. $T \circ \phi^{-1}$ for the *Reg*, *Reg-True-Mask* and *RegSeg* algorithms. SSD is computed outside the segmentation mask this corresponds to the healthy tissue in the case of lesion segmentation. Moreover, it is computed for two distinct regions, one near the segmentation boundary and one over the entire image. Each algorithm ran for 25 iterations with *Reg* taking an average runtime of 3 minutes and *RegSeg* and *RegSeg-Morph* taking an average runtime of 4 minutes.

Algorithms (Segmentation)	*p-value*
RegSeg vs. Seg	0.015
RegSeg-Morph vs. Seg	0.004
RegSeg-Morph vs. RegSeg	0.616

Table 6.4: We performed a two-sided t-test to compare the segmentation performance of different algorithms. These particular tests were done with the Dice scores of 25 subjects. We excluded 3 subjects from the dataset whose lesion was in the cerebellum since this location made it hard to retrieve an axial slice that was comparable to the other subjects.

Algorithms (Registration)	*Region*	*p-value*
RegSeg vs. Reg	*Whole image*	0.180
	Near lesion	0.0009
Reg-True-Mask vs. RegSeg	*Whole image*	0.005
	Near lesion	0.348

Table 6.5: We performed a two-sided t-test to compare the registration performance of different algorithms. These particular tests were done with the SSD between the template and 25 subjects. We excluded 3 subjects from the dataset whose lesion was in the cerebellum since this location made it hard to retrieve an axial slice that was comparable to the other subjects.

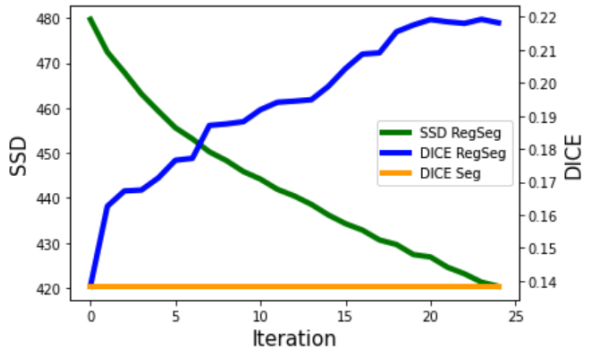


Figure 6.16: Registration and segmentation performance of the *RegSeg* algorithm, showing steady improvement in both the registration metric (SSD) and the segmentation metric (DICE) over 25 iterations. All metrics are mean values over the 25 subjects.

CHAPTER 7

Conclusion

As previously mentioned, the classical approach to solving medical imaging problems that require image segmentation and image registration is to apply the two procedures sequentially as part of two distinct components of a data pipeline without accounting for their co-dependencies. Our work considers the integration of these two fundamental computer vision problems. Such an integrated approach first appeared in [5] where the authors proposed a generative model that combines non-linear image registration, brain tissue segmentation and intensity bias correction. They called this framework *Unified Segmentation* and the approach was implemented in the Statistical Parametric Mapping (`SPM`) software [3], which was and remains widely used. Their segmentation component only considered tissue-type segmentation and their registration component was based on a low-frequency basis function transformation model which is typically outperformed by LDDMM for reasons discussed in Section 3.1.2. This work was successfully extended in [52] to account for lesion segmentation but the code that implemented this extension was not made publicly available and they once again did not consider the LDDMM registration framework. Ashburner's initial work on joint segmentation and registration was extended once more by the authors of [10], in which they proposed a hierarchical model with an LDDMM-based transformation model. However, this work did not consider lesion segmentation and used an older implementation of LDDMM through the `DARTEL` diffeomorphic image registration software [2]. Some contemporary joint registration and segmentation work has considered a modern deep learning integrated approach under a semi-supervised and tissue-type segmentation setting [63]. Our work is the first modern implementation of joint registration and segmentation of lesioned brains in an unsupervised setting, and also the first joint registration and segmentation method based on an efficient open-source implementation of LDDMM.

Our main contributions include a generative model that combines differomorphic registration and brain lesion segmentation, along with an efficient inference algorithm and validation of this model on a real dataset of MR neuroimages with stroke lesions, showing significant improvements in both registration and segmentation over baseline algorithms. Moreover, the registration component of the inference algorithm used the `mermaid` open-source deep-learning package which we modified to incorporate segmentation information during inference. Applying our algorithm to an ischemic stroke lesion dataset, we were able to significantly improve upon a sequential application of registration and segmentation.

There are many opportunities for further work on this topic, including: improvements to our generative model; learning hyper-parameters in supervised and unsupervised settings; learning dataset-specific templates; and validation of these methods in a variety of applications and datasets. In particular, it would be very interesting to add MRFs to our generative model, both for segmentation and registration. For segmentation, an MRF approach would canonically model local dependencies between neighbouring pixels which would penalize segmentations with small disconnected regions and encourage segmenta-

tions with large connected regions. In the case of brain lesion segmentation, this would help decrease both false negatives and false positives. For registration, some MRF-based methods have the unique advantage of finding a global minimum, albeit minimum of an objective function that is less realistic than that of LDDMM; it would be interesting to try to combine these two models. Our work is a step towards a more general model, and a proof of concept for joint registration and segmentation of lesioned neuroimages.

References

[1] Aja-Fernández, S., & Vegas-Sánchez-Ferrero, G. (2016). Statistical noise models for mri. In *Statistical analysis of noise in mri* (pp. 31–71). Springer.

[2] Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*(1), 95–113.

[3] Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K., ... others (1994). Statistical parametric mapping.

[4] Ashburner, J., & Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human brain mapping*, *7*(4), 254–266.

[5] Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage*, *26*(3), 839–851.

[6] Ashburner, J., Friston, K. J., et al. (2003). Spatial normalization using basis functions. *Human brain function*, *2*.

[7] Aydin, O. U., Taha, A. A., Hilbert, A., Khalil, A. A., Galinovic, I., Fiebach, J. B., ... Madai, V. I. (2021). On the usage of average hausdorff distance for segmentation performance assessment: hidden error when used for ranking. *European Radiology Experimental*, *5*(1), 1–7.

[8] Beal, L. D., Hill, D. C., Martin, R. A., & Hedengren, J. D. (2018). Gekko optimization suite. *Processes*, *6*(8), 106.

[9] Blaiotta, C., Cardoso, M. J., & Ashburner, J. (2016). Variational inference for medical image segmentation. *Computer Vision and Image Understanding*, *151*, 14–28.

[10] Blaiotta, C., Freund, P., Cardoso, M. J., & Ashburner, J. (2018). Generative diffeomorphic modelling of large mri data sets for probabilistic template construction. *NeuroImage*, *166*, 117–134.

[11] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), 859–877.

[12] Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., & Cuadra, M. B. (2011). A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine*, *104*(3), e158–e177.

[13] Choi, Y., & Lee, S. (2000). Injectivity conditions of 2d and 3d uniform cubic b-spline

functions. *Graphical models*, *62*(6), 411–427.

[14] Christensen, G. E., Rabbitt, R. D., & Miller, M. I. (1996). Deformable templates using large deformation kinematics. *IEEE transactions on image processing*, *5*(10), 1435–1447.

[15] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, *35*(1), 53–65.

[16] Crinion, J., Ashburner, J., Leff, A., Brett, M., Price, C., & Friston, K. (2007). Spatial normalization of lesioned brains: performance evaluation and impact on fmri analyses. *Neuroimage*, *37*(3), 866–875.

[17] Crum, W. R., Hartkens, T., & Hill, D. (2004). Non-rigid image registration: theory and practice. *The British journal of radiology*, *77*(suppl_2), S140–S153.

[18] Farnoush, R., & ZAR, P. B. (2008). Image segmentation using gaussian mixture model.

[19] Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., ... others (2004). Automatically parcellating the human cerebral cortex. *Cerebral cortex*, *14*(1), 11–22.

[20] Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., ... others (2011). Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, *54*(1), 313–327.

[21] Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets* (pp. 267–285). Springer.

[22] Ghaffari, M., Sowmya, A., & Oliver, R. (2019). Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the brats 2012–2018 challenges. *IEEE reviews in biomedical engineering*, *13*, 156–168.

[23] Glocker, B., Komodakis, N., Tziritas, G., Navab, N., & Paragios, N. (2008). Dense image registration through mrfs and efficient linear programming. *Medical image analysis*, *12*(6), 731–741.

[24] Glocker, B. M. (2011). *Random fields for image registration* (Unpublished doctoral dissertation). Technische Universität München.

[25] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

[26] Grimmett, G. R. (1973). A theorem about random fields. *Bulletin of the London Mathematical Society*, *5*(1), 81–84.

[27] Hinton, G. E. (2007). To recognize shapes, first learn to generate images. *Progress in brain research*, *165*, 535–547.

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

[29] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., ... others (2009). Evaluation of 14 nonlinear deformation algorithms applied to human

brain mri registration. *Neuroimage*, *46*(3), 786–802.

[30] Ko, G. G., Chai, Y., Rutenbar, R. A., Brooks, D., & Wei, G.-Y. (2019). Accelerating bayesian inference on structured graphs using parallel gibbs sampling. In *2019 29th international conference on field programmable logic and applications (fpl)* (pp. 159–165).

[31] Komodakis, N., & Pesquet, J.-C. (2015). Playing with duality: An overview of recent primal? dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, *32*(6), 31–54.

[32] Liew, S.-L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., ... others (2018). A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Scientific data*, *5*(1), 1–11.

[33] Liew, S.-L., Zavaliangos-Petropulu, A., Jahanshad, N., Lang, C. E., Hayward, K. S., Lohse, K. R., ... others (2022). The enigma stroke recovery working group: Big data neuroimaging to study brain–behavior relationships after stroke. *Human brain mapping*, *43*(1), 129–148.

[34] Maier, O., Menze, B. H., von der Gablentz, J., Häni, L., Heinrich, M. P., Liebrand, M., ... others (2017). Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical image analysis*, *35*, 250–269.

[35] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... others (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, *34*(10), 1993–2024.

[36] Miller, M. I., Trouvé, A., & Younes, L. (2002). On the metrics and euler-lagrange equations of computational anatomy. *Annual review of biomedical engineering*, *4*(1), 375–405.

[37] Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*.

[38] Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, *13*(6), 47–60.

[39] Mumford, D. B., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*.

[40] Nesterov, Y. (1983). A method for unconstrained convex minimization problem with the rate of convergence o (1/kˆ 2). In *Doklady an ussr* (Vol. 269, pp. 543–547).

[41] Niethammer, M., Kwitt, R., & François-Xavier. (2019). Metric learning for image registration. *CVPR*.

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

[43] Patil, D. D., & Deore, S. G. (2013). Medical image segmentation: a review. *International Journal of Computer Science and Mobile Computing*, *2*(1), 22–27.

[44] Pit, V., & Schapira, B. (2018). Finiteness of gibbs measures on noncompact manifolds

with pinched negative curvature. In *Annales de l'institut fourier* (Vol. 68, pp. 457–510).

[45] Pock, T., Cremers, D., Bischof, H., & Chambolle, A. (2009). An algorithm for minimizing the mumford-shah functional. In *2009 ieee 12th international conference on computer vision* (pp. 1133–1140).

[46] Reinhold, J. C., Dewey, B. E., Carass, A., & Prince, J. L. (2019). Evaluating the impact of intensity normalization on MR image synthesis. In *Medical imaging 2019: Image processing* (Vol. 10949, p. 109493H).

[47] Risholm, P., Samset, E., & Wells, W. (2010). Bayesian estimation of deformation and elastic parameters in non-rigid registration. In *International workshop on biomedical image registration* (pp. 104–115).

[48] Rojas, R. (1996). The backpropagation algorithm. In *Neural networks* (pp. 149–182). Springer.

[49] Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., & Hammers, A. (2006). Diffeomorphic registration using b-splines. In *International conference on medical image computing and computer-assisted intervention* (pp. 702–709).

[50] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation* (Tech. Rep.). California Univ San Diego La Jolla Inst for Cognitive Science.

[51] Seeger, M. (2004). Gaussian processes for machine learning. *International journal of neural systems*, *14*(02), 69–106.

[52] Seghier, M. L., Ramlackhansingh, A., Crinion, J., Leff, A. P., & Price, C. J. (2008). Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage*, *41*(4), 1253–1266.

[53] Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, *13*(1), 146–165.

[54] Shen, Z., Han, X., Xu, Z., & Niethammer, M. (2019). Networks for joint affine and non-parametric image registration. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4224–4233).

[55] Shen, Z., Vialard, F.-X., & Niethammer, M. (2019). Region-specific diffeomorphic metric mapping. *NeurIPS: Neural Information Processing Systems*.

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

[57] Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, *17*(1), 87–97.

[58] Sotiras, A., Davatzikos, C., & Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, *32*(7), 1153–1190.

[59] Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, *15*(1), 1–28.

[60] Tambe, S. B., Kulhare, D., Nirmal, M., & Prajapati, G. (2013). Image processing

(ip) through erosion and dilation methods.

[61] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, *2018*.

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

[63] Xu, Z., & Niethammer, M. (2019). Deepatlas: Joint semi-supervised learning of image registration and segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 420–429).

[64] Yang, X., Kwitt, R., & Niethammer, M. (2016). Fast predictive image registration. In *Deep learning and data labeling for medical applications* (pp. 48–57). Springer.

[65] Yang, X., Kwitt, R., Styner, M., & Niethammer, M. (2017). Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, *158*, 378–396.

[66] Younes, L., Arrate, F., & Miller, M. I. (2009). Evolutions equations in computational anatomy. *NeuroImage*, *45*(1), S40–S50.

[67] Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, *21*(11), 977–1000.