Trabajo Fin de Máster

LDDMM y GANs: Redes Generativas Antagónicas para Registro Difeomorfico

LDDMM meets GANs: Generative Adversarial Networks for Diffeomorphic registration

Autor/es

Ubaldo Ramón Júlvez

Director/es

Elvira Mayordomo Cámara

Mónica Hernández Giménez

Escuela de ingeniería y Arquitectura

2022

Index

1. Abstracts

1.1 Abstract

Diffeomorphic deformable image registration is a key problem for many Computational Anatomy applications. Traditionally, deformable image registration has been formulated as a variational problem solved by costly numerical optimization problems. In the last decade, contributions to novel traditional formulations are decreasing, while deep learning models are being increasingly developed to learn deformable registration from images. In this work, we contribute to this new trend by proposing a novel adversarial learning LDDMM method for diffeomorphic registration of pairs of 3D images. The method is based on Generative Adversarial Networks, by combining the best performing generator and discriminator architectures in deformable registration with the LDDMM paradigm. We have successfully implemented three models for different parameterizations of diffeomorphisms, which show competitive performance against benchmark state-of-the-art deep learning and traditional methods.

1.2 Resumen

El Registro Difeomorfico de imágenes es un problema clave para muchas aplicaciones de la Anatomía Computacional. Tradicionalmente, el registro deformable de imagen ha sido formulado como un problema variacional, resoluble mediante costosos métodos de optimización numérica. En la última década, contribuciones en la forma de nuevos métodos basados en formulaciones tradicionales están decreciendo, mientras que más modelos basados en Aprendizaje profundo están siendo desarrollados para aprender registros deformables de imágenes. En este trabajo contribuimos a esta nueva corriente proponiendo un novedoso método LDDMM para registro difeomorfico de imágenes 3D, basado en redes generativas antagónicas. Combinamos las arquitecturas de generadores y discriminadores con mejores prestaciones en registro deformable con el paradigma LDDMM. Hemos implementado con éxito tres modelos para distintas parametrizaciones de difeomorfismos, los cuales demuestran resultados competitivos en comparación con métodos del estado del arte tanto tradicionales como basados en aprendizaje profundo.

2. Introduction

Medical imaging refers to a series of techniques aimed at the acquisition and processing of images of the interior of the human body. These techniques have become fundamental in modern clinical practice for the diagnosis of a variety of diseases, along with other tasks such as medical studies, disease prevention, treatment selection, guidance during surgical intervention, etc. Though collectively referred to as imaging, this discipline works with many different image acquisition techniques and technologies, which can derive from the need for different processing or interpretation methods. Among these acquisition techniques, we can find technologies such as magnetic resonance imaging (MRI), ultrasound (US), positron emission tomography (PET), and many others not mentioned or yet to be developed.

It is then of paramount importance that automatic and reliable methods are developed to analyze, process, and interpret the vast and varied imaging data available, either with the purpose of aiding clinical professionals or as part of fully automated systems for diagnosis or other applications, which are on the rise with the recent resurgence of machine learning techniques such as deep learning. To this end, many key problems have been identified, such as segmentation, meaning the identification of different structures of interest within the body, atlas creation, referring to the generation of a significant image template which encompasses a populations anatomy, and image registration, which aims at establishing correspondences between the structures of two images.

Especially interesting is the field of Computational Anatomy, which is focused on quantitative investigation and modeling of anatomical shape variability. It involves the use of mathematical and statistical methods for modeling biological structures. Anatomical information is encoded by the spatial transformation existing between images, which allows for statistical analysis on the spaces of the transformations which can model the anatomical variability within a population.

The problem of deformable image registration arises from the need to find transformations between images. In greater detail, deformable image registration has the aim of matching all or several corresponding anatomical structures in two images through a plausible spatial transformation or mapping. Though this can include linear transformations, usually a non-linear transformation which determines voxel-to-voxel correspondences is desired. The use of diffeomorphisms to model the deformations for Computational Anatomy applications was proposed in the early 2000s as a first step towards ensuring the use of mathematically correct transformations, as diffeomorphisms present many desirable properties such as being differentiable and invertible, which ensures a smooth one-to-one mapping and also conserves the topology of the transformed anatomies.

2.1 Model-based methods

Traditional model-based methods solve registration as a variational problem. Given a deformation model with certain constraints, an iterative optimization is performed so as to minimize a custom energy function, which will ensure the accuracy of the registration and the desired regularity of the transformation. Some early methods include Demons [1] or free-form deformations with b-splines [2].

The Large Deformation Diffeomorphic Metric Mapping (LDDMM) can be found among the most influential approaches for diffeomorphic registration [3]. In LDDMM, diffeomorphic transformations are parametrized by time-varying flows of vector fields, and the solution to the registration problem is calculated by solving the non-stationary transport equation associated with these flows.

The main limitation of the LDDMM model is given by its large computational complexity and slow convergence due to the use of gradient descent optimization. This has motivated many improvements like the stationary field parametrization [4, 5, 6], the momentum conservation constrained (MCC) parameterization under the Euler-Poincare differential (EPDiff) equation [7, 8, 9, 10], second-order optimization approaches [11, 6, 12], and the band-limited parameterization [13, 14].

2.2 Learning-based methods

The introduction of deep learning to medical imaging comes after significant advances for many different computer vision applications boosted by the use of convolutional neural networks (CNN). Early works in CNNs to learn the optical flow (FlowNet [15]) and for biomedical image segmentation (U-net [16]), set the bases for the structure of a deep learning-based method for deformable image registration. The most influential is perhaps U-net. The U-net model approaches the problem of image segmentation with the use of an encoder-decoder architecture, where the image is first downscaled into features which are then reconstructed to form a solution with the size of the original input domain. Based upon this, a vast amount of deep-learning methods have been proposed in recent years to approach the problem of deformable image registration in different clinical applications [17].

Most of the proposals for deep learning based diffeomorphic registration still use the ingredients from traditional model-based methods such as the stationary parameterization [18, 19, 20, 21], the non-stationary parameterization [22] or the parameterization with EPDiff-constrained velocity fields [23, 24]. All the proposals to diffeomorphic registration can be classified into supervised [18, 23, 24], where ground truth from some traditional model-based method is used, or unsupervised learning methods [19, 20, 25] which use traditional image similarity metrics. Unsupervised methods are usually preferred over supervised ones since the transformations can be learned directly from image pairs, avoiding the overhead to compute ground truth transformations for training. Overall, all deep learning methods yield fast inference algorithms for diffeomorphism computation, once the models have been trained, but struggle to considerably improve accuracy over model-based methods.

Among the unsupervised approaches, we find an interesting variant in the use of Generative Adversarial Networks (GANs) for image registration. A GAN combines the interaction of two different networks during training: a generative network and a discrimination network. The generative network itself can be regarded as an unsupervised method that, once included in the GAN system, is trained with the feedback of the discrimination network, which is used to substitute or enhance the information provided by traditional image similarity metrics. Several non-diffeomorphic deformable registration proposals have been made [26] (2D) and [27, 28] (3D). GANs have also been used for diffeomorphic deformable template generation [29], where the registration sub-network is based on an established U-net architecture [30, 31]. However, GANs have still not yet been implemented in any deep learning based diffeomorphic deformable registration methods, which will be the aim of this work.

2.3 Scope of the study and organization

In this master's thesis, we aim to advance the field of deep learning models for diffeomorphic registration, via the design, implementation, training, and evaluation of a GAN-based deep learning model. The network will offer the possibility to work both on 2D and 3D data. Training will be performed on brain MRI images from the Alzheimer's Disease Neuroimaging Initiative (ADNI, adni.loni.usc.edu), while the evaluation will be performed on the independent dataset Non-rigid Image Registration Evaluation Project (NIREP) [32]. For the sake of comparison with the current state of the art, we include in our evaluation some well established model-based methods: diffeomorphic Demons [6], stationary LDDMM (St. LDDMM) [33] and Flash [13], as well as two freely available deep learning models: Voxelmorph II [21] and Quicksilver [23]. This work is a continuation of our previous research in diffeomorphic registration [34].

The structure of this document is as follows: Chapter 2 presents an overview of the deformable registration problem, related methods, and contents of our work. Chapter 3 first describes the methods and techniques relevant to our system and then presents the components of our deep learning architecture and its technical details. Chapter 4 presents our experiments and our obtained results, including their discussion. Finally, Chapter 5 gives an overview of the presented work, proposes future work, and gives conclusions about our results. Additionally, as complementary materials, we present illustrations with annotations of the major brain structures which can be found in appendix 1.

3. Methods

3.1 Computational anatomy

Broadly speaking, the aim of computational anatomy is the study of shape variability from imaged anatomical structures. This study is based upon the use of templates, which are generated images that serve as an anatomical reference, and the transformations that exist between any given real image (i.e. obtained from a real patient) and the aforementioned templates. Human anatomy is thus modeled as a deformable template, being defined via orbits associated with the action of non-rigid transformation groups on the template. The group action allows for the generation of new images, which belong to the same reference anatomy, and which are referred to as deformable templates.

Groups of transformations have a Riemannian manifold structure. This allows for the use of metric distances by means of the group actions, something which is not possible directly over the anatomical images. A correct statistical analysis can be then performed on the transformations that associate images with the anatomical reference template rather than on the images themselves. From this arises the need to find suitable deformation models so as to represent anatomical variance found in the images, and the problem to find these deformations between images referred to as deformable image registration. The preferred representation for these deformations is the diffeomorphisms, in particular those that can be calculated as geodesic paths given the initial conditions.

3.2 Deformable registration

Image registration refers to the problem of finding a warp from one image (source) to another (target) so as to minimize some energy function, which usually incorporates traditional image similarity metrics. Alternatively, this can be interpreted as maximizing the similarity between both images.

The use of rigid or affine transformations is possible although they offer very little representation capability for modeling the complex biological shapes found in human anatomy. The use of non-rigid transformations is better suited for this task, with a large number of degrees of freedom that establish a dense voxel-wise non-linear spatial correspondence between the source and target images.

Let I_0 and I_1 be the source and the target images, and ϕ^{-1} the deformation between I_0 and I_1, then, the deformable image registration problem can be formulated as:

$$\phi^{*-1} = \arg\min_{\phi^{-1}} E_{sim}(I_1, I_0 \circ \phi^{-1}) + E_{reg}(\phi^{-1}), \tag{3.1}$$

where ϕ^{*-1} is the optimal deformation that minimizes a given image similarity metric E_{sim} while maintaining smoothness, governed by some regularization function E_{reg} which ensures desirable properties in the transformation.

Early deformable registration approaches parameterize the deformation model as using a displacement field u, such that:

$$\phi^{-1}(x) = x - u(x), \tag{3.2}$$

for x any point in the image. Though simple, this parameterization does not guarantee the existence of an inverse transformation of the displacement field, especially for large transformations which are commonly found within anatomical variance.

3.3 Diffeomorphic registration with LDDMM

In the search of a more adequate deformation model, and under the assumption that ideally the deformation model responsible for organ growth and capable of a good description of the variation of human anatomy is required to be smooth and invertible, we turn to the use of diffeomorphisms. A diffeomorphism is a bijective map such that itself and its inverse are smooth or differentiable. Diffeomorphisms have other desirable properties, like positive and bounded Jacobian determinants or preserving the topology.

Large Deformation Diffeomorphic Metric Mapping (LDDMM) is widely considered as the reference paradigm for diffeomorphic registration. In this setting, transformations are parametrized as time-varying vector field flows, defined on the tangent space of a Riemannian manifold of diffeomorphisms. In relation to equation 3.1, the sum of squared intensity differences is usually selected as E_{sim}, while E_{reg} is defined as the energy associated with the length of the time-varying path. Classical gradient descent optimization methods and second-order methods like Gauss-Newton's method are most commonly used for numerical optimization.

In greater detail, let $\Omega \subseteq \mathbb{R}^d$ be the image domain. Let $Diff(\Omega)$ be the LDDMM Riemannian manifold of diffeomorphisms and V the tangent space at the identity element. $Diff(\Omega)$ is a Lie group, and V is the corresponding Lie algebra [3]. The Riemannian metric of $Diff(\Omega)$ is defined from the scalar product in V, $\langle v, w\rangle_V = \langle \mathcal{L}v, w\rangle_{L^2}$, where $\mathcal{L}$ is the invertible self-adjoint differential operator associated with the differential structure of $Diff(\Omega)$. In traditional LDDMM methods, $\mathcal{L} = (Id - \alpha\Delta)^s, \alpha > 0, s \in \mathbb{R}$ [3]. We will denote with K the inverse of operator $\mathcal{L}$.

Let I_0 and I_1 be the source and the target images. LDDMM is formulated from the minimization of the variational problem

$$E(v) = \frac{1}{2}\int_0^1 \langle \mathcal{L}v_t, v_t\rangle_{L^2} dt + \frac{1}{\sigma^2}\|I_0 \circ (\phi_1^v)^{-1} - I_1\|_{L^2}^2. \tag{3.3}$$

The LDDMM variational problem was originally posed in the space of time-varying smooth flows of velocity fields, $v \in L^2([0,1], V)$. Given the smooth flow $v : [0,1] \to V$, $v_t : \Omega \to \mathbb{R}^d$, the solution at time $t = 1$ to the evolution equation

$$\partial_t (\phi_t^v)^{-1} = -v_t \circ (\phi_t^v)^{-1} \tag{3.4}$$

with initial condition $(\phi_0^v)^{-1} = id$ is a diffeomorphism, $(\phi_1^v)^{-1} \in Diff(\Omega)$. The transformation $(\phi_1^v)^{-1}$, computed from the minimum of $E(v)$, is the diffeomorphism that solves the LDDMM registration problem between I_0 and I_1.

The most significant limitation of LDDMM is its large computational complexity. In order to circumvent this problem, the original LDDMM variational problem is parameterized on the space of initial velocity fields

$$E(v_0) = \frac{1}{2}\langle \mathcal{L}v_0, v_0 \rangle_{L^2} + \frac{1}{\sigma^2}\| I_0 \circ (\phi_1^v)^{-1} - I_1 \|_{L^2}^2. \tag{3.5}$$

where the time-varying flow of velocity fields v is obtained from the Euler-Poincare differential equation (EPDiff) equation

$$\partial_t v_t + K[(Dv_t)^T \cdot Lv_t + DLv_t \cdot v_t + Lv_t \cdot \nabla \cdot v_t] = 0 \tag{3.6}$$

with initial condition v_0 (geodesic shooting). The diffeomorphism $(\phi_1^v)^{-1}$, computed from the minimum of $E(v_0)$ via Equations 3.6 and 3.4, verifies the momentum conservation constraint (MCC) [8], and, therefore, it belongs to a geodesic path on $Diff(\Omega)$.

Simultaneously to the MCC parameterization, a family of methods was proposed to further circumvent the large computational complexity of original LDDMM [11, 6, 33]. In all these methods, the time-varying flow of velocity fields v is restricted to be steady or stationary [4].

3.4 Generative Adversarial Networks

Generative adversarial networks, GANs for short, are a type of deep learning based generative model usually used for unsupervised learning, which is usually used to generate new samples from a given domain that are similar to those it has been trained on.

GAN-based approaches depart from unsupervised approaches by the definition of two different networks: the Generator network (G) and the Discriminator network (D). The generator model is used to generate samples in the domain, originally from a random seed but informed versions are possible as well. The discriminator model is tasked with distinguishing samples from the target domain and samples obtained from the generator, which is simply a well-understood classification problem between real and generated samples.

The main feature of GANs is that both sub-networks are trained together, in an adversarial fashion. The discriminator is trained to identify the generated images as fake, while the generator tries to fool the discriminator. In this way, both models are competing against each other, which should conclude, in theory, with the generator generating samples that are indistinguishable to the discriminator from those of the target domain.

3.5 GAN-based unsupervised deep-learning networks for diffeomorphic registration

Similarly to model-driven approaches for estimating LDDMM diffeomorphic registration, data-driven approaches for learning LDDMM diffeomorphic registration aim at the inference of a diffeomorphism $(\phi_1^v)^{-1}$ such that the LDDMM energy is minimized for a given (I_0, I_1) pair. In particular, data-driven approaches compute an approximation of the functional

$$\mathcal{S}(\arg\min_{v \in V} E(v, I_0, I_1)), \tag{3.7}$$

where $\mathcal{S}$ represents the operations needed to compute $(\phi_1^v)^{-1}$ from v, and the energy E is either given by Equations 3.3 or 3.5. The functional approximation is obtained via a neural network representation with parameters learned from a representative sample of image pairs. Unsupervised approaches assume that the LDDMM parameterization in combination with the minimization of the energy E considered as a loss function is enough for the inference of suitable diffeomorphic transformations after training. Therefore, there is no need for ground truth deformations.

The generative network in this context is the diffeomorphic registration network. G is aimed at the approximation of the functional given in Equation 3.7 similarly to unsupervised approaches for the inference of $(\phi_1^v)^{-1}$. The discrimination network D outputs the probability $p \in [0, 1]$ that for a pair (I_0^w, I_1) the image I_0^w comes from a warped source *not* being generated by G. The discrimination network D learns to distinguish between a warped source image $I_0 \circ (\phi_1^v)^{-1}$ generated by G and a plausible warped source image. The learnable parameters of the network G are trained to minimize traditional LDDMM cost functions between the warped source image and the target image while trying to fool the discriminator D. In contrast to other unsupervised approaches, the loss function in G is determined from the combination of the LDDMM and the adversarial costs.

In this work, we propose three versions of the network, depending on the parameterization used for computing the transformation. A method with stationary parameterization for the velocity fields which we will refer to as SVF-GAN, a method with EPDiff-constrained parameterization or EPDiff-GAN, and finally a small deformation method mostly for comparison which directly calculates the diffeomorphic transformation Disp-GAN.

3.6 Adversarial training

As stated above, the registration architecture is composed of two neural networks, a generator G and a discriminator D, which are trained alternatively as follows

The discriminator network D is trained using the loss function

$$L_D = \begin{cases} -\log(p) & c \in P^+ \\ -\log(1-p) & c \in P^- \end{cases} \tag{3.8}$$

where c indicates the input case, P^+ and P^- indicate positive or negative cases for the GAN, and p is the probability computed by D for the input case.

In the first place, D is trained on a positive case $c \in P^+$ representing a target image I_1 and a warped source image I_0^w plausibly registered to I_1 with a diffeomorphic transformation. The warped source image is modeled from a strictly convex linear combination of I_0 and I_1

$$I_0^w = \beta I_0 + (1-\beta) I_1. \tag{3.9}$$

It should be noted that, although the warped source image would ideally be I_1, the selection of $I_0^w = I_1$ (e.g. $\beta = 0$) empirically leads to the discriminator rapidly outperforming the generator. The parameter β is the relative mean squared error (MSE) obtained after registration for I_0^w and I_1 since

$$MSE_{rel} = \frac{\|I_0 \circ (\phi_1^v)^{-1} - I_1\|_{L^2}^2}{\|I_0 - I_1\|_{L^2}^2} \quad \text{and} \quad \|I_0^w - I_1\|_{L^2}^2 = \beta \|I_0 - I_1\|_{L^2}^2. \tag{3.10}$$

Therefore, this model for I_0^w can be regarded as a good candidate for warped sources after deformable registration for small βs, and has been successfully used in previous adversarial learning methods for deformable registration [28].

In the second place, D is trained on a negative case $c \in P^-$ representing a target image I_1 and a warped source image I_0^w obtained from the generator network G. In this way, the discriminator network will learn to assign high probabilities to plausibly warped images while giving a low probability to images warped by the registration network.

Finally, the generator network G is trained using the combined loss function

$$L_G = L_{\text{adv}} + \lambda \frac{1}{2} \langle \mathcal{L} v_0, v_0 \rangle_{L^2} + \frac{1}{\sigma^2} \|I_0 \circ (\phi_1^v)^{-1} - I_1\|_{L^2}^2. \tag{3.11}$$

In this loss function, L_{adv} is the adversarial loss function, defined from $L_{\text{adv}} = -\log(p)$ where p is computed from I_0^w with D, and the rest is the LDDMM energy given by Equations 3.3 or 3.5. Finally, λ is the weight for balancing the image similarity and the regularization losses. Note that for Disp-GAN, where velocity fields are calculated, the regularization term $\lambda \frac{1}{2} \langle \mathcal{L} v_0, v_0 \rangle_{L^2}$ can't be calculated, and we instead substitute it by a regularization acting on the displacement fields as proposed in [21].

For each sample pair (I_0^w, I_1), G is fed with the pair of images and updates the network parameters from the back-propagation of the information of the loss function values coming from the LDDMM energy and the discriminator probability of being a pair generated by G.

In the early stages of learning, it is expected that the generator network provides misaligned images and the discriminator penalizes the system with high probabilities for the negative cases. As the learning progresses, the generator is trained to fool the discriminator, so the generated warped sources will be diffeomorphically transformed to resemble I_1 as much as possible according to the convex linear model. The discriminator will eventually hardly distinguish the generated warped sources from the true population, yielding low probabilities for the negative cases, and learning will be considered to converge.

3.7 Proposed GAN architecture

An overview of the whole GAN-based diffeomorphic registration network can be observed in figure 3.1. As usual for GAN architectures, it is composed of two main neural networks: a

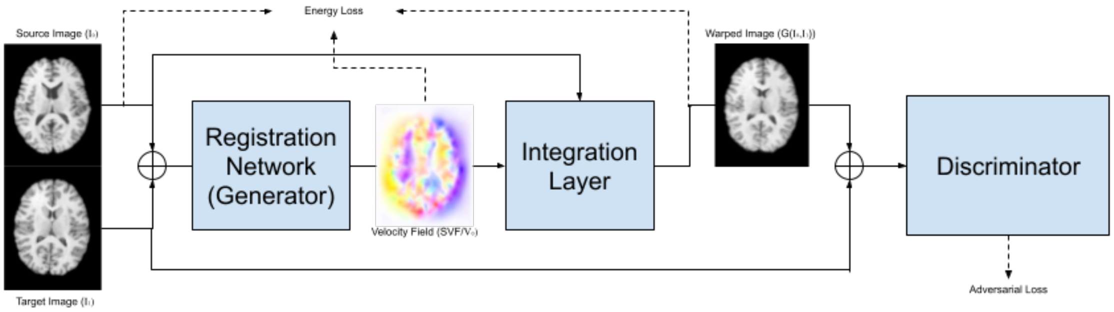

Figure 3.1: Overview of the whole GAN diffeomorphic registration architecture. The Generator takes as input the concatenated source and target images, and outputs a stationary or time dependant velocity field. The velocity field is then feed to the integration layer (which will vary depending on the parameterization) and outputs the final transformation with which the warped image is calculated. Finally, the Discriminator is given the concatenated warped and target images, and the Energy loss is calculated.

Generator and a Discriminator. Additionally, an integration layer connects both networks and calculates the final warped image given the velocity field. This layer will vary depending on the parameterization. The information flow is as follows, the Generator takes as input the concatenated source and target images, and outputs the velocity field or displacement, which is then fed to the integration layer to calculate the final warped image. The discriminator takes as input the original target image, along with the registered warped image, or alternatively a plausible registered image for training as described in section 3.6.

3.7.1 Generator network.

In this work, the diffeomorphic registration network G is intended to learn LDDMM registration parameterized on the space of steady velocity fields, on the space of initial velocity fields subject to the EPDiff equation (Equation 3.6), or directly on the diffeomorphic transformation for the case of Disp-GAN.

A number of different generator network architectures have been proposed in the recent literature, with a predominance of simple fully convolutional (FC) [26] or U-Net like architectures [27, 28]. In this work, we propose to use the architecture by Duan et al. [27] adapted to fit our purposes. The G network can be seen in figure 3.2. The network follows the general U-net design of utilizing an encoder-decoder structure with skip connections between corresponding levels.

The initial level features a regular convolution, followed by an activation function and a max pooling layer. This early max pooling operation is key because of the high dimensionality of the input images. After the initial level, features are fed to two encoding streams. The first one (rightmost on figure 3.2) features two convolutional blocks, composed each of one regular convolution, a max pooling layer, a dilated convolution, and an activation layer. After each block, the output is concatenated with that of the corresponding level of the second stream. This second level also features two convolutional blocks, each composed of two convolutions followed by activation functions, and a third convolution followed by a max pooling layer. For each block, the output after the second convolution is concatenated to that of the corresponding level.

During the decoding phase, the network first features two up-sampling blocks, composed of two convolutions followed by activation functions, and followed by an up-sampling transposed convolution. A third block features a regular convolution followed by an activation function, a transposed convolution, and finally two convolutions to obtain the final results. Note the last convolution is not followed by an activation function.

All regular convolutional layers feature a kernel size of (3x3x3), except for those preceding a max pooling layer which features a kernel size of (2x2x2). Dilated convolutional layers feature a kernel size of (3x3x3) and a dilation rate of (2x2x2). Finally, the transposed convolution layers feature kernel sizes of (2x2x2) and a stride of (2x2x2).

The purpose of the two encoding streams is to work at two scale levels, so as to obtain fine-scale features from the regular stream and coarser but broader features from the dilated. This is achieved by the dilated convolutions which feature a larger receptive field more suitable to learn large deformations.

The up-sampling is performed with a deconvolutional operation based on transposed convolutional layers [35]. We have empirically noticed that the learnable parameters of these layers help reduce typical checkerboard GAN artifacts in the decoding [36].

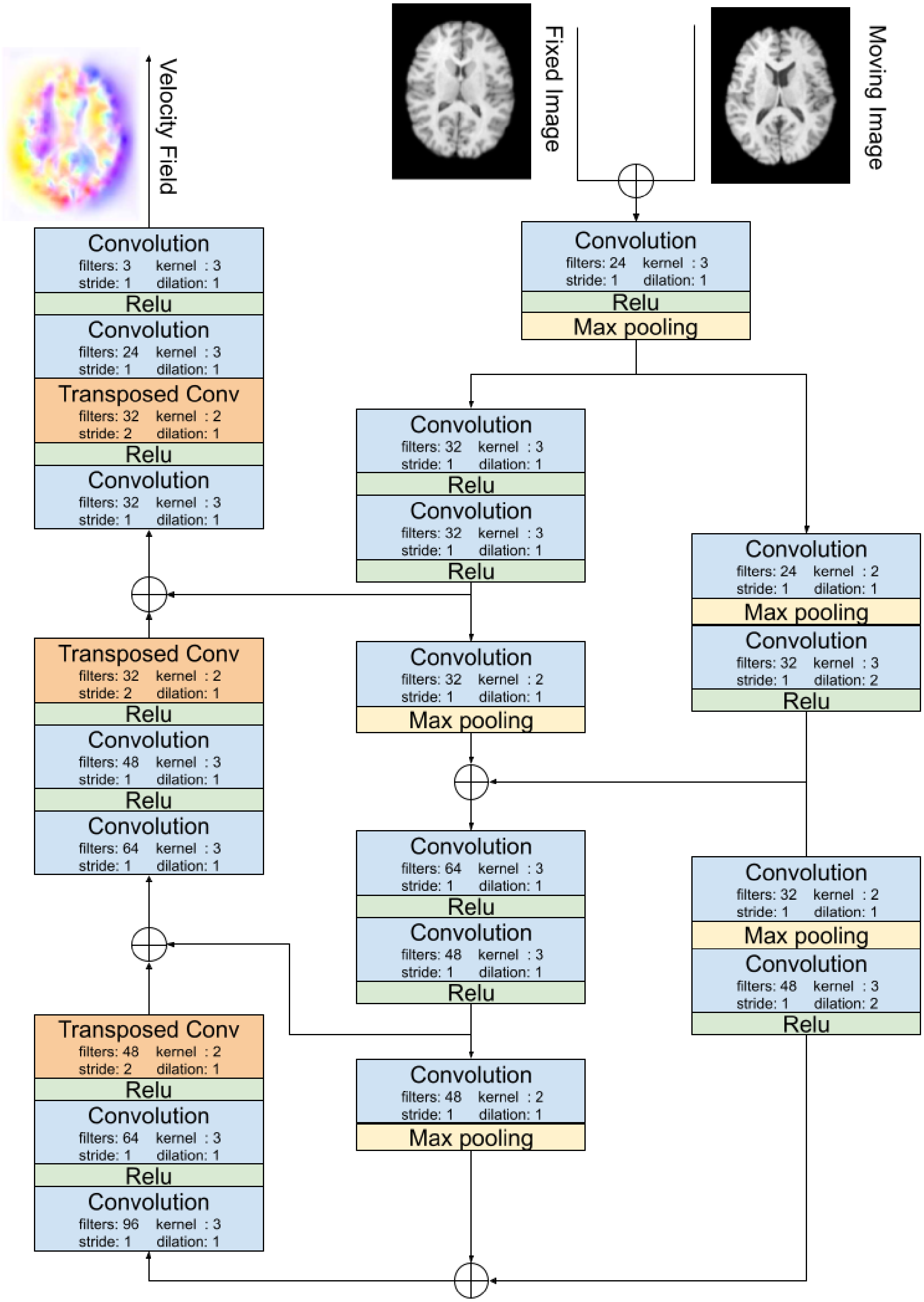

Figure 3.2: Illustration of the proposed generator network architecture.

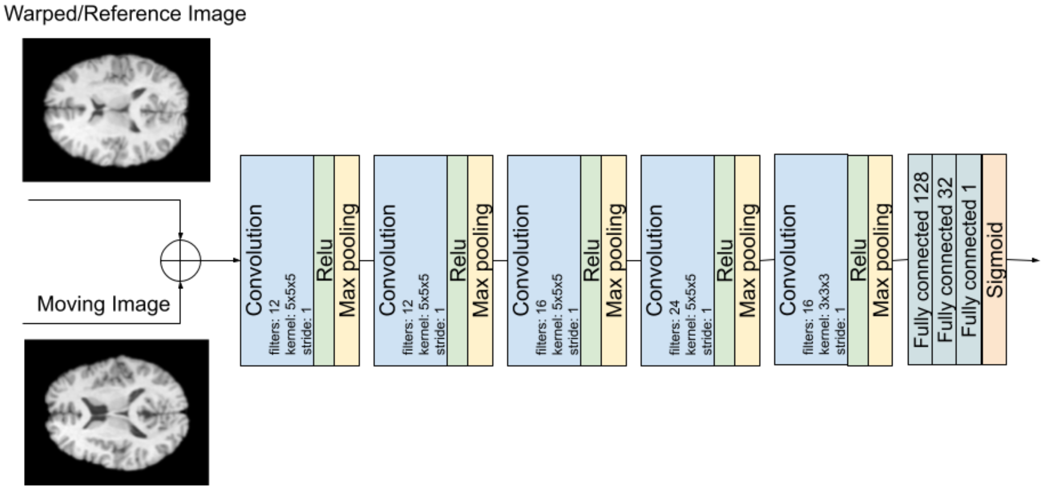

Figure 3.3: Illustration of the proposed discriminator network architecture.

3.7.2 Discriminator network.

The discriminator network D follows a very traditional CNN architecture. The two input images are concatenated and passed through five convolutional blocks. Each block includes a convolutional layer, a RELU activation function, and a size-two max-pooling layer. The size of all convolutional filters is (5x5x5), except for the last block that is (3x3x3). After the convolutions, the 4D volume is flattened and passed through three fully connected layers, ending in a sigmoid activation function. The output of the last layer is the probability of the input images to come from a registered pair not generated by G.

Throughout our experiments, we empirically encountered that the Discriminators task is more easily achieved than the generators and that a simple architecture can keep up with the generators learning and thusly aimed for efficiency in its design.

3.7.3 Generative-Discriminative integration layer.

The generator and the discriminator networks G and D are connected through an integration layer followed by a spatial transformation layer. This integration layer allows calculating the diffeomorphism $(\phi_1^v)^{-1}$ that warps the source image I_0 into I_0^w. The selected integration layer depends on the velocity parameterization: stationary or EPDiff-constrained time-dependent. For the stationary velocity fields, we employ the scaling and squaring method for diffeomorphisms [4]. For the EPDiff-constrained time dependent velocity fields, the solution of the deformation state equation [3] is found using geodesic shooting via Euler integration. For the case of Disp-GAN, the integration layer is omitted since G produces $(\phi_1^v)^{-1}$ directly.

The computed diffeomorphisms are applied to the source image via a 3D spatial transformation layer [37], using cubic interpolation. Neither the integration nor the spatial transformation layers feature any learnable parameters, but they must be included within the network since the obtained warped image is necessary for the training of D and thus the gradients of the integration layer have to be back-propagated from D to train G.

3.7.4 Parameter selection

We selected the parameters $\lambda = 500$ for SVF-GAN, $\lambda = 100$ for EPDiff-GAN and Disp-GAN, $\sigma^2 = 0.03$, $\alpha = 0.0025$, and $s = 2$ and a unit-domain discretization of the image domain Ω [3]. The values for λ were selected via hyperparameter tuning, as will be commented later on chapter 4, while the values for the rest are standard for LDDMM.

Scaling and squaring and Euler integration were performed in 8 and 10 time samples respectively as extensively used in traditional and deep learning based LDDMM literature [4, 13].

The parameter β for the convex linear modeling of warped images was selected equal to 0.2. This means that the discriminator is trained to learn deformable image registration results with a 20% of MSE_{rel} level of accuracy. We empirically found that while lower values of β did result in a decrease in MSE_{rel}, this did not translate to the other evaluation metrics which are arguably more important, and thus settled for the higher value which helps maintain training loss equilibrium between the Generator and Discriminator during training.

Both the generator network and the discriminator network were trained with Adam's optimizer with default parameters and learning rates of $5e^{-5}$ for G and $1e^{-6}$ for D, respectively.

4. Results

The main target domain of our method is the registration of 3D brain MRI datasets, but although not implicitly multi-model our architecture could easily be adapted to work on other domains. For our purposes of tuning and demonstrating the effectiveness of our non-supervised proposed method, we employ 2D simulated and brain MRI datasets for hyperparameter searching as well as for proof of concepts, while we employ 3D brain MRI datasets for evaluating the effectiveness of the methods on the target domain. This is necessary since training several models on 3D datasets can prove prohibitive because of the very high dimensionality of the data, while 2D datasets prove much less resource-consuming but present a similar enough trend with the 3D case so as to use them for parameter selection.

4.1 Metrics

The ideal ground-truth for a diffeomorphic registration is not well-defined, though usually results from traditional model-based models are used as a point of comparison, they can't be used as proper evaluation and thus the use of other metrics is required.

Dice Similarity Coefficient (DSC). The most commonly used metric for measuring how similar a warped image is to the target one is the Dice Similarity Coefficient (DSC), also known as Sørensen–Dice coefficient. DSC is a measure of the similarity between two sets. In the context of registration and anatomical study, it is used to estimate the spatial overlap of anatomical segmentation maps. These anatomical maps assign a label to each voxel of an image, corresponding to the anatomical structure present in that voxel. It can be calculated as :

$$\frac{2 \cdot |X \cap Y|}{(|X| + |Y|)} \tag{4.1}$$

Where X and Y are two sets, and $\cap$ is the intersection operation. For anatomical segmentation maps, the sets are those voxels in the target image which are assigned the same label as in the warped source image. Though automatic tools for image segmentation have been developed, for the purpose of evaluation very precise maps are preferred and thus usually those created by experts are used as the gold standard. This process is evidently very time-consuming and thus datasets with anatomical segmentations available are always of very few samples.

Jacobian Matrix Determinant. The jacobian matrix of a vector valued function is the matrix of all of its first-order partial derivatives. For a deformation field ϕ^{-1}, its jacobian matrix is the first-order derivative of each of its three spatial directions (x, y, z) with respect to the others. At any given point p, its definition is :

$$J_{\phi^{-1}}(p) = \begin{pmatrix} \frac{\partial \phi_x^{-1}(p)}{\partial x} & \frac{\partial \phi_x^{-1}(p)}{\partial y} & \frac{\partial \phi_x^{-1}(p)}{\partial z} \\ \frac{\partial \phi_y^{-1}(p)}{\partial x} & \frac{\partial \phi_y^{-1}(p)}{\partial y} & \frac{\partial \phi_y^{-1}(p)}{\partial z} \\ \frac{\partial \phi_z^{-1}(p)}{\partial x} & \frac{\partial \phi_z^{-1}(p)}{\partial y} & \frac{\partial \phi_z^{-1}(p)}{\partial z} \end{pmatrix} \tag{4.2}$$

For a deformation to be diffeomorphic locally, it must guarantee that the determinant of its jacobian matrix is positive (i.e. $|J_{\phi^{-1}}(p)| > 0$). Thus, we can check if the deformation is globally diffeomorphic by checking that all of its Jacobian determinants are positive. High values of the jacobian determinant are also undesirable as they indicate that the deformation is not sufficiently locally smooth, and could result in negative Jacobian determinants in the inverse. Though the value for being considered high is not well defined, for our purpose we will consider smooth deformations those with $|J_{\phi^{-1}}(p)| < 10$.

4.2 Datasets

4.2.1 3D brain MRI datasets.

As a dataset for training, we used a total of 2113 T1-weighted brain MRI 3D images from the Alzheimer's Disease Neuroimaging Initiative (ADNI, adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The images were acquired at the baseline visit and belong to all the available ADNI projects (1, 2, Go, and 3). The images were preprocessed with N3 bias field correction, affinely registered to the MNI152 atlas, skull-stripped, and affinely registered to the skull-stripped MNI152 atlas. Sagittal, axial, and coronal views of a sample from ADNI and from the MNI152 atlas can be seen in figures 4.2 and 4.1 respectively.

Image pairs were selected randomly during training from among the total of 2113, one image was assigned as the source and the other as the target before being fed to the network. This constitutes a total of 1056 samples per training epoch.

The evaluation of our generated GAN models in the task of diffeomorphic registration was performed in NIREP dataset [32]. This dataset was released for the evaluation of non-rigid registration. The geometry of the segmentations in NIREP provides a specially challenging framework for deformable registration evaluation. The images were acquired from 8 males and 8 females with a mean age of 32.5 ± 8.4 and 29.8 ± 5.8 years, respectively. The substantial age differences between train and evaluation subjects are intended to demonstrate the generalization capability of our non-supervised models.

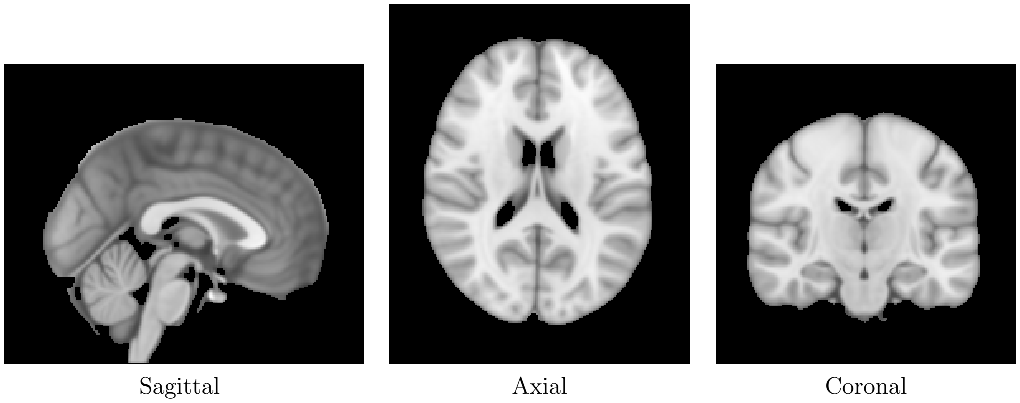
Sagittal Axial Coronal

Figure 4.1: Example of the 3D MNI152 atlas. Sagittal, axial and coronal views from the MNI152 atlas..

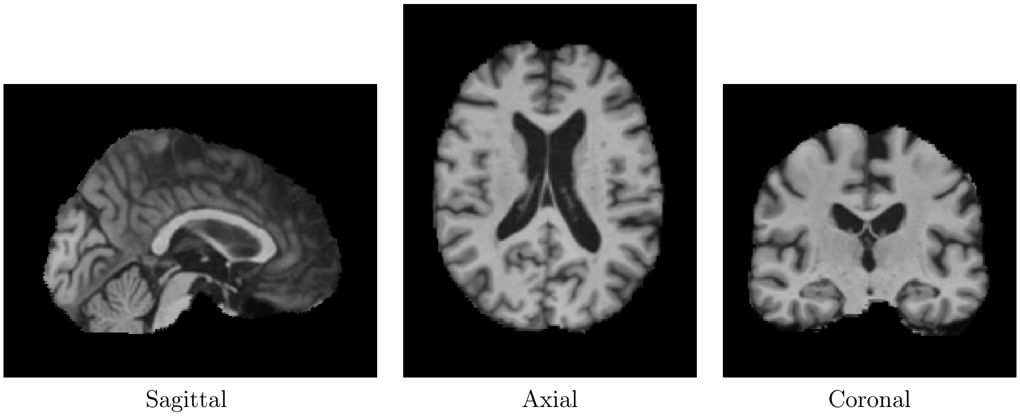
Sagittal Axial Coronal

Figure 4.2: Example of a 3D ADNI MRI image. Sagittal, axial and coronal views from a sample image from the ADNI dataset.

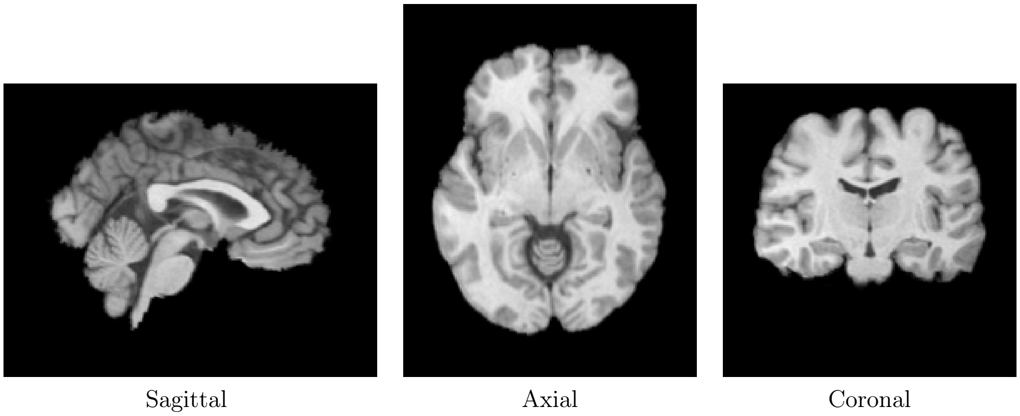
Sagittal Axial Coronal

Figure 4.3: Example of a 3D NIREP MRI image. Sagittal, axial and coronal views from a sample image from the NIREP dataset.

All images were scaled to a size of 176 × 224 × 176. Note that since the independent NIREP dataset was used for testing, no images from the ADNI dataset were reserved for this purpose. Additionally, no evaluation set was segmented from the ADNI dataset either since hyperparameter searching was performed on the 2D datasets and thus was not necessary for 3D. This allows us to use all of the 2113 images, which are still pretty limited for deep learning standards, exclusively for training.

4.2.2 2D Datasets

2D datasets were generated from the brain MRI 3D images by taking slices along the middle of the second component (axial view), resulting in images of size 176 × 176. This was done for both the NIREP and ADNI datasets. Experiments on 2D image slices make for a good approximation of full 3D brain MRI images, while requiring a small portion of resources and time to execute, making them very useful for hyperparameter searching which we can then extrapolate to 3D experiments.

A second 2D dataset of 2560 simulated torus images was also used. The torus images were generated by varying the parameters of two ellipse equations, similarly to [24]. The parameters were drawn from two Gaussian distributions: $\mathcal{N}(4,2)$ for the inner ellipse and $\mathcal{N}(12,4)$ for the outer ellipse. The simulated images were of size 64 × 64. This constitutes a much simpler to register dataset but is still useful for testing the properties of our resulting transformations. Furthermore, registration results for brain MRI images are usually very difficult to interpret visually, while for these simpler shaped datasets transformations can be identified easily.

4.3 Results on the 2D datasets

The first batch of experiments was done on the 2D simulated dataset. The network was trained on the 2560 simulated torus images for 1000 epochs, with a batch size of 64 samples. All other parameters were as previously described. The objective of this experiment is first to serve as proof of concept of the proposed model by observing its performance on a geometrically easy dataset that still features large deformations. Also, because of the simplicity of the images, we can compare our transformations with those from other learning-based or model-based approaches, and search for similarities in the calculated deformation and velocity fields.

Figure 4.4 shows the deformed images and the velocity fields obtained in the 2D simulated dataset by diffeomorphic Demons [6], a stationary version of LDDMM (St. LDDMM) [33], the spatial version of Flash [13], and our proposed SVF and EPDiff GANs. The 5 pairs of torus shown for testing are randomly generated with the same parameters as stated for the training set. Apart from diffeomorphic Demons that use Gaussian smoothing for regularization, all the considered methods use the same parameters for operator L. Therefore, St. LDDMM and SVF-GAN can be seen as a model-based and a data-based approach for the minimization of the same variational problem. The same happens with Flash [13] and EPDiff-GAN.

From the figure, it can be appreciated that our proposed GANs are able to obtain accurate warps of the source to the target images, similarly to model-based approaches. For SVF-

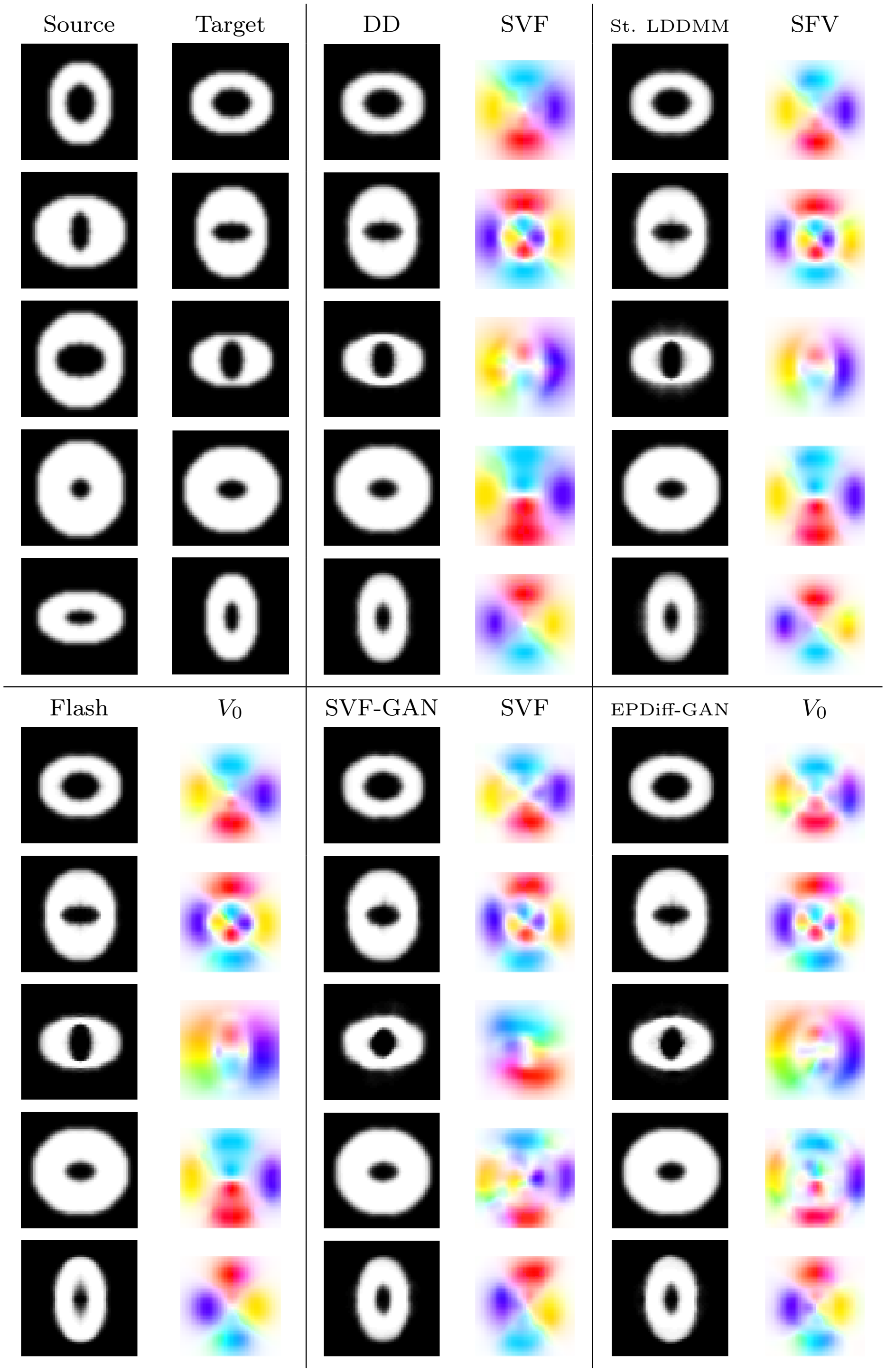

Figure 4.4: Example of simulated 2D registration results. First row left to right: source and target images of five selected experiments, deformed images and velocity fields computed from diffeomorphic Demons (DD) and stationary LDDMM (St. LDDMM). Second row left to right: deformed images and velocity fields computed from Flash, and our proposed SVF-GAN and EPDiff-GAN. SVF stands for a stationary velocity field and V_0 for the initial velocity field of a geodesic shooting approach, respectively.

Method	Reg λ	MSE_{rel}	$\#\lvert J_{\phi^{-1}}\rvert < 0$	$\#\lvert J_{\phi^{-1}}\rvert > 10$
Disp-GAN	$\lambda = 0.01$	0.087	1067.7	20.2
	$\lambda = 0.04$	0.085	1071.0	13.1
	$\lambda = 0.1$	0.083	752.4	14.7
	$\lambda = 0.5$	0.079	935.6	15.5
	$\lambda = 1$	0.077	856.0	16.1
	$\lambda = 5$	0.075	591.73	15.2
	$\lambda = 10$	0.087	438.8	7.6
	$\lambda = 50$	0.104	606.0	4.2
	$\lambda = 100$	0.155	145.4	0
	$\lambda = 500$	0.300	0	0
	$\lambda = 1000$	0.490	0	0
SVF-GAN	$\lambda = 50$	0.108	0	33.5
	$\lambda = 100$	0.134	0	10.9
	$\lambda = 500$	0.188	0	0
	$\lambda = 1000$	0.196	0	0
	$\lambda = 5000$	0.321	0	0
	$\lambda = 10000$	0.414	0	0
	$\lambda = 50000$	0.703	0	0
	$\lambda = 100000$	0.825	0	0
EPDiff-GAN	$\lambda = 50$	0.161	0	0
	$\lambda = 100$	0.140	0	0
	$\lambda = 500$	0.154	0	0
	$\lambda = 1000$	0.178	0	0
	$\lambda = 5000$	0.278	0	0
	$\lambda = 10000$	0.394	0	0
	$\lambda = 50000$	0.713	0	0
	$\lambda = 100000$	0.840	0	0

Table 4.1: Results for Lambda parameter searching on 2D dataset. From top to bottom, results for Disp-GAN, SVF-GAN and EPDiff-GAN with varying λ values. From left to Right, model type, regularization λ used, mean MSE_{rel} obtained across test set, mean negative Jacobian determinant across test set and mean high Jacobian determinant across test set.

GAN, the inferred velocity fields are visually similar to model-based approaches in three of five experiments. For EPDiff-GAN, the inferred initial velocity fields are visually similar to model-based approaches in four of five experiments.

Overall, this experiment shows our model is able to represent large deformations while maintaining local detail (i.e. the shape of the inner holes on all our results is maintained) and to produce results with globally minimum energy (those similar to the model-based ones) in most of the cases.

The second batch of experiments was performed on the 2D brain slices dataset. The network was trained for 500 epochs, with a batch size of 32 samples for all runs. The aim of this experiment was to estimate the value for lambda that would allow the registration network to generate the best possible deformations while ensuring these remained diffeomorphic. For each of the 3 proposed models, Disp-GAN, SVF-GAN, and EPDiff-GAN several runs were performed while varying the λ parameter from equation 3.11.

Table 4.1 shows the MSE_{rel} and Jacobian determinant results obtained from the trained models on the NIREP test dataset. For the Disp-GAN model, results show many negative Jacobian determinants appear for low regularization values, implying the model doesn't generate properly diffeomorphic deformations. Only for very high $\lambda \geq 500$ does it obtain proper diffeomorphic deformations, but with a serious performance downgrade as it obtains a MSE_{rel} of 0.3. Since we are aiming at a value of 0.2 for our adversarial training, we deem this model not to be appropriate to obtain diffeomorphic deformations, though for the sake of comparison moving forwards trials with this model will be performed with $\lambda = 100$, as it is the most regularised model with MSE_{rel} under 0.2.

The results obtained for SVF-GAN and EPDiff-GAN on the other hand show that these two models generate globally diffeomorphic deformations much more easily, as we don't encounter a negative Jacobian determinant across any of the trials, though for the case of SVF-GAN some Jacobian determinants with value over 10 are present. We deem the most appropriate λ parameters for the SVF-GAN model to be $\lambda = 500$ since it features smooth diffeomorphic deformations while staying under our MSE_{rel} objective, though a model with $\lambda = 100$ was also trained to check if 2D trends were maintained in 3D. For the EPDiff-GAN model a $\lambda = 100$ was selected, since it obtained the lowest MSE_{rel} (though these were very similar for $\lambda < 1000$).

4.4 Results in the 3D NIREP dataset

Method	MSE_{rel}	DSC	$\#\lvert J_{\phi^{-1}}\rvert < 0$	$\#\lvert J_{\phi^{-1}}\rvert > 10$	$min(\lvert J_{\phi^{-1}}\rvert)$	$max(\lvert J_{\phi^{-1}}\rvert)$
Disp-GAN	0.085 ± 0.006	0.604 ± 0.095	746,966.75 ± 14,221.93	22,259.60 ± 793.83	-109.04 ± 12.19	187.59 ± 19.38
SVF-GAN-100	0.150 ± 0.012	0.597± 0.086	0.00 ± 0.00	1,494.00 ± 427.03	0.01 ± 0.01	53.37 ± 20.25
SVF-GAN-500	0.224 ± 0.015	0.576 ± 0.096	0.00 ± 0.00	5.20 ± 11.31	0.07 ± 0.02	9.99 ± 1.35
EPDiff-GAN	0.235 ± 0.017	0.557 ± 0.097	0.00 ± 0.00	0.00±0.00	0.09±0.02	5.98±0.66
QS	0.190 ± 0.013	0.575 ± 0.097	0.00 ± 0.00	0.00 ± 0.00	0.31 ± 0.03	9.54 ± 2.44
VM-Disp	0.144 ± 0.010	0.583 ± 0.102	80,292.47 ± 9,181.35	282.67 ± 88.49	-24.82 ± 7.52	23.80 ± 2.85
VM-SVF	0.227 ± 0.009	0.555 ± 0.102	0.00 ± 0.00	804.27 ± 225.85	0.00 ± 0.00	102.31 ± 50.35

Table 4.2: Evaluation in NIREP, all measures show the mean and standard deviation across the 15 registrations. From left to right, : Deep learning method used for registration, relative mean squared error, DICE score, negative jacobian determinants, high jacobian determinants, minimum jacobian determinant and maximum jacobian determinant. Deep learning Methods, from top to bottom: Disp-GAN, SVF-GAN($\lambda = 100$), SVF-GAN($\lambda = 500$), EPDIff-GAN, Quicksilver(QS), Voxelmorph II (Deformation) and Voxelmorph II (Stationary parametrization)

Training of the 3D models was performed on the 2113 ADNI MRI samples introduced before. In total we trained 6 models, one for Disp-GAN with $\lambda = 100$, two for SVF-GAN, with $\lambda = 100$ and $\lambda = 500$, and lastly EPDiff-GAN with $\lambda = 100$. Additionally, as an ablation test, two models were trained in where we substitute our generator architecture for a simpler U-net architecture, one for SVF-GAN and another for EPDiff-GAN. The rest of the network's parameters were kept in all models as described in subsection 3.7.4.

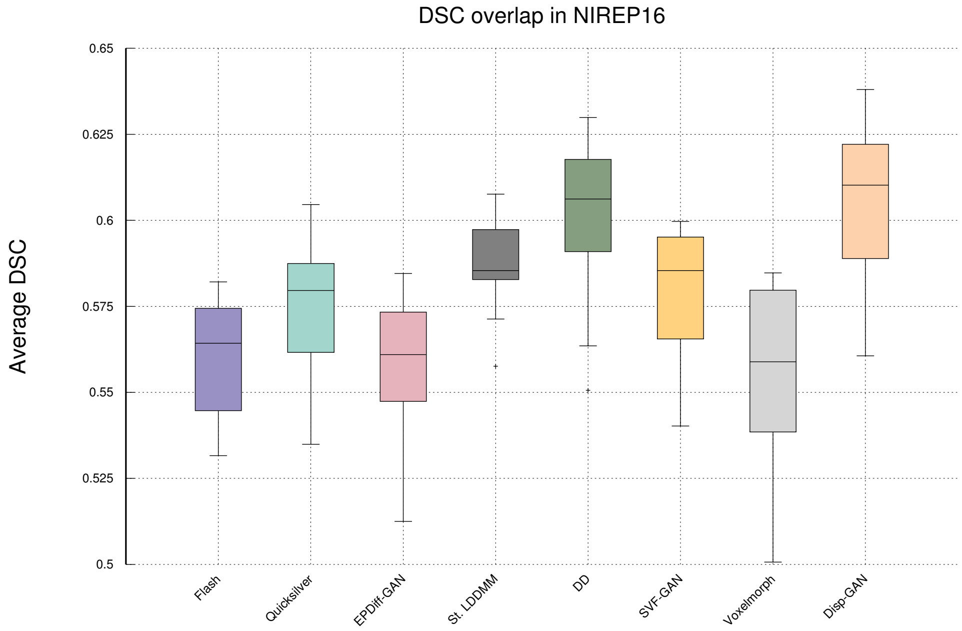

Figure 4.5: Evaluation in NIREP. Dice scores obtained by propagating the diffeomorphisms to the segmentation labels on the 16 NIREP brain structures, box plot showing averages across all structures. Methods: Flash, Quicksilver, EPDiff-GAN(ours), stationary LDDMM (St. LDDMM), diffeomorphic Demons(DD), SVF-GAN(ours), Voxelmorph and Disp-GAN(ours).

All GANs were trained during 50 epochs with a batch size of 1 sample. This selection of batch sampling was performed due to VRAM memory issues since models working with 3D data are very memory intensive.

Testing of the trained models was performed on the NIREP dataset, which was registered with all our model variations and evaluated using the metrics presented. We also performed registration on the same NIREP dataset with two state-of-the-art deep learning based registration methods, the unsupervised Voxelmorph II [21] (two versions, one with stationary velocity fields parametrization and one with direct deformation generation, similar to our Disp-GAN), and the supervised Quicksilver [23]. Registration was also performed with some well-known and well-performing model-based methods, diffeomorphic Demons [6] and St. LDDMM [33] for stationary velocity fields, and the spatial version of Flash [13] for time dependent velocity fields.

4.4.1 Quantitative assessment

In figure 4.6 we can observe box plots showing the Dice similarity coefficients obtained with the stationary methods, from the 16 brain structures (initials on the x axis) across the 16 NIREP samples. Figure 4.7 shows the same results for the geodesic shooting methods. Figure

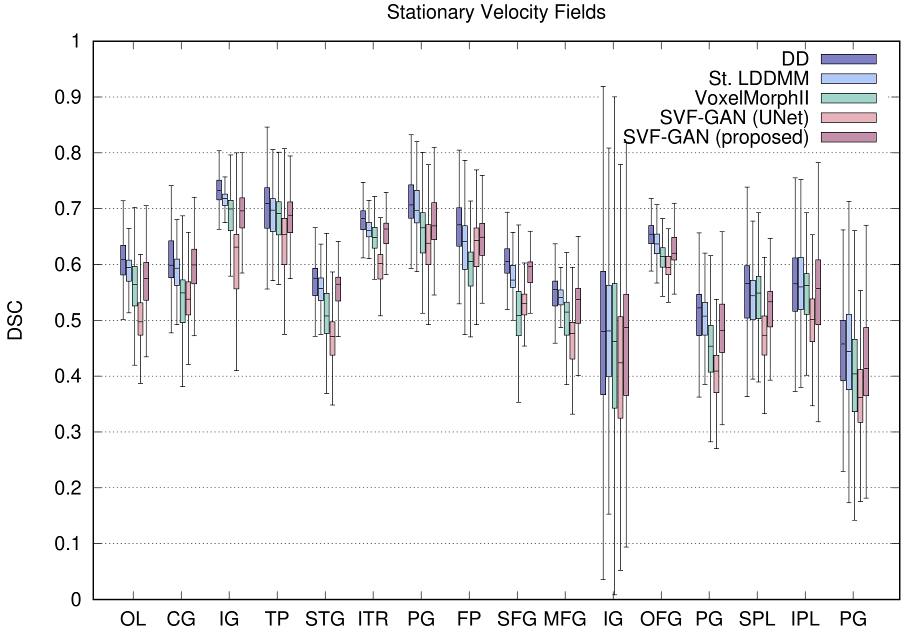

Figure 4.6: Evaluation in NIREP. Dice scores obtained by propagating the diffeomorphisms to the segmentation labels on the 16 NIREP brain structures, one box plot for each structure. Methods parameterized with stationary velocity fields: diffeomorphic Demons (DD), stationary LDDMM (St. LDDMM), Voxelmorph II, our SVF-GAN with U-Net architecture, and our proposed SFV-GAN with the two-stream architecture.

4.5 shows box plots of the DSC score average across all NIREP brain structures, for all relevant methods.

Results on all brain structures show how our proposed two-stream architecture greatly improves the performance obtained by a basic U-Net, probably because as theorized the two-different streams with different scales help the network with the prediction of large deformations, crucial for obtaining good results on a difficult dataset such as NIREP.

With regard to the stationary methods, SVF-GAN shows an accuracy similar to St. LDDMM and is competitive with diffeomorphic Demons. Our proposed method tends to overpass Voxelmorph II in the great majority of the structures.

On the other hand, EPDiff-GAN shows an accuracy similar to Flash and Quicksilver in the great majority of regions, with the exception of the temporal pole (TP) and the orbital frontal gyrus (OFG), two small localized and difficult to register regions. It drives our attention that Flash under-performed in the superior frontal gyrus (SFG). Finally, on table 4.2 we can see metric comparisons for all the deep learning methods tested, our 4 main trained models, two variants of Voxelmorph II, and Quicksilver.

Regarding the metrics showcasing the accuracy of the registrations, MSE_{rel} and DSC, we can observe how the models with simpler parametrizations, which directly calculate the deformation, Disp-GAN and VM-Disp, obtain the lowest MSE_{rel} and highest DSC scores.

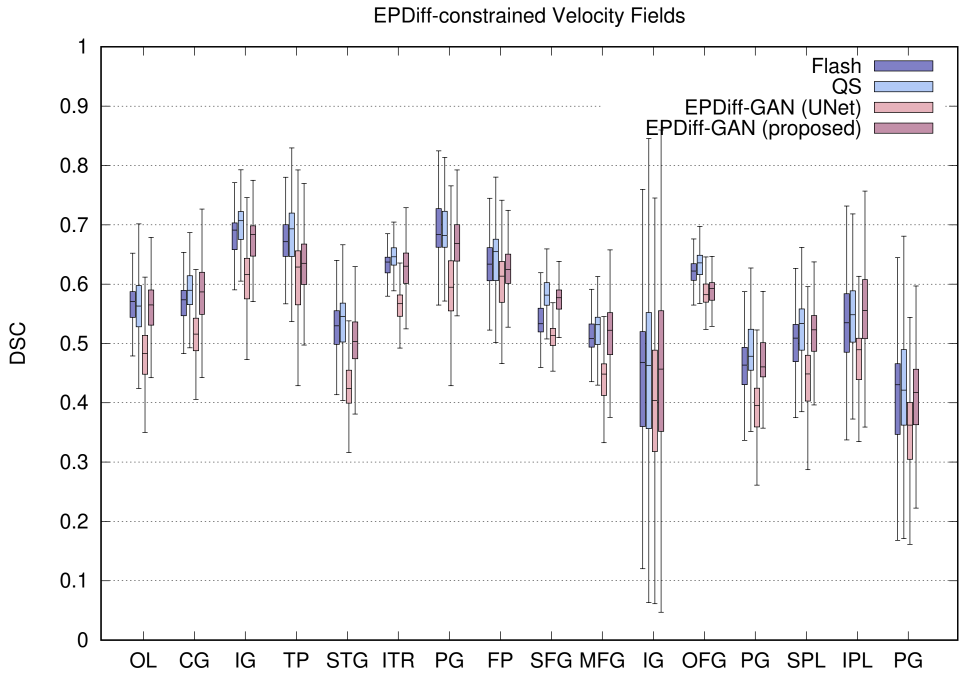

Figure 4.7: Evaluation in NIREP. Dice scores obtained by propagating the diffeomorphisms to the segmentation labels on the 16 NIREP brain structures, one box plot for each structure. Geodesic shooting methods: Flash, Quicksilver (QS), our EPDiff-GAN with U-Net architecture, and our proposed EPDiff-GAN.

This is likely because this less complex architectures allow for easier training of the network. However both these methods generate many negative Jacobian determinant values, trend maintained from the results obtained on 2D, which means the resulting deformations cannot be considered diffeomorphic. We believe this property to be fundamental to the registration problem, and thus we don't consider fair their comparison with the results from the other models.

Among the methods which generate no negative determinant values, SVF-GAN-100 presents the best performance, in fact very closely behind that of Disp-GAN. SVF-GAN-500 produces however considerably lower DSC but is still ahead of both Quicksilver and Voxelmorph II. Our network with time dependent velocity fields EPDiff-GAN underperforms with respect to the also time dependent QS, whose better performance is likely due to being supervised and thus using ground truth initial velocities as training. We believe this underperformance to be likely due to a lack of training resources. In fact, EPDiff-GAN showed very similar results to the other architectures in all our 2D experiments. However, the number of epochs and batch size used in 2D is simply not viable for 3D, and since EPDiff-GAN is the most complex model its performance was degraded the most.

Lastly, concerning the smoothness of the deformations, only Quicksilver and EPDiff-GAN present no Jacobian determinants over the value of 10, showing how the time dependant parametrization is the easiest to regularize so as to obtain smooth deformations. SVF-GAN-100 obtains large Jacobian determinants on only a few samples, while SVF-GAN-500 presents them on all samples but in small quantities, again following the trend from our 2D experiments. Whether this is acceptable or not would depend on the application and remains a trade-off between better smoothness and better registration accuracy. In conclusion, from all the test data we can affirm that we have developed a GAN-based learning registration architecture, capable to compete with model-based and learning based state-of-the-art registration methods. Additionally, we show how trends from simplified experiments in 2D translate accurately to 3D and thus facilitate tests for architecture testing and hyperparameter search. Lastly, we observe how simpler parametrizations of the deformations generally obtain more accurate but less smooth results, being the stationary parametrization of the velocity fields the optimal middle ground with current architecture designs and computational capacity.

4.4.2 Qualitative assessment of 3D results

For a qualitative assessment of the quality of the registration results, Figures 4.8 and 4.9 show the sagittal and axial views of one selected NIREP registration result. In the figure, we can appreciate a high matching between the target and the warped ventricles, and more difficult to register regions like the cingulate gyrus (observable in the sagittal view) or the insular cortex (observable in the axial view). For those nonfamiliar with brain anatomical regions, these regions are easily identified as the garnet and orange regions as seen in the supplementary material.

4.4.3 Implementation details and Computational complexity

The experiments were run on a machine equipped with one NVidia Titan RTX with 24 GBS of video memory and an Intel Core i7 with 64 GBs of DDR3 RAM. All training and testing procedures were run on the GPU. Code for the networks was developed in Python, making use primarily of the Deep learning libraries Keras and TensorFlow.

Our 2D GAN models were trained for 1 hour and 30 minutes, while our 3D GAN models were trained for 2 days and 22 hours. The VRAM memory load was equal to the whole GPU capacity (24GB). Once trained, for the 3D models, the inference time for the Disp-GAN model was of 0.6 seconds, while for the EPDiff-GAN and SVF-GAN models of 1.3 seconds. For comparison, the fastest version of Flash [13] has a computation time of around 229.4 seconds, though only using 1.3 GB of VRAM.

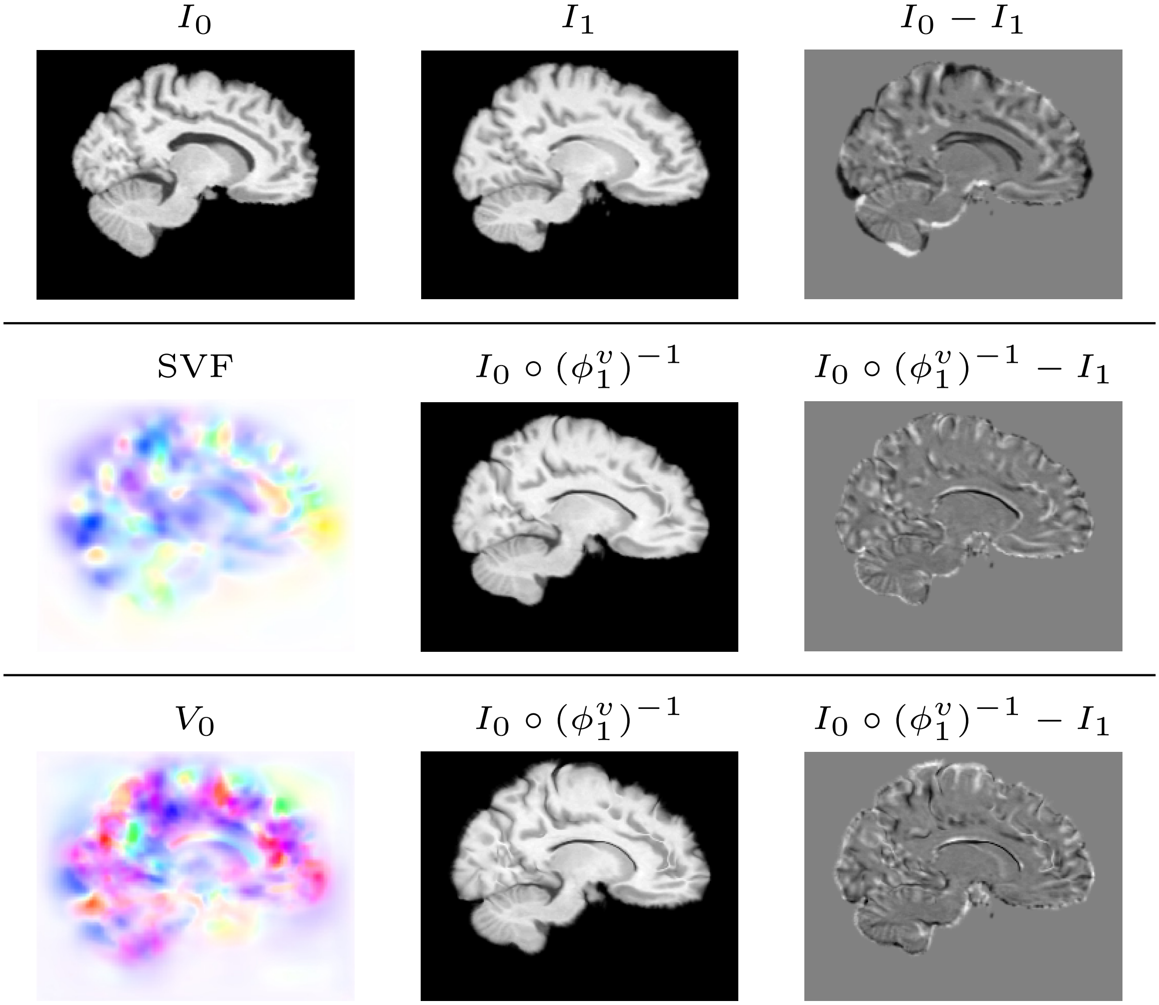
I_0
I_1
$I_0 - I_1$
SVF
$I_0 \circ (\phi_1^v)^{-1}$
$I_0 \circ (\phi_1^v)^{-1} - I_1$
V_0
$I_0 \circ (\phi_1^v)^{-1}$
$I_0 \circ (\phi_1^v)^{-1} - I_1$

Figure 4.8: Example of 3D registration results, sagittal view. First row, sagittal views of the source and the target images and the differences before registration. Second row, inferred stationary velocity field, warped image, and differences after registration for SVF-GAN. Third row, inferred initial velocity field, warped image, and differences after registration for EPDiff-GAN.

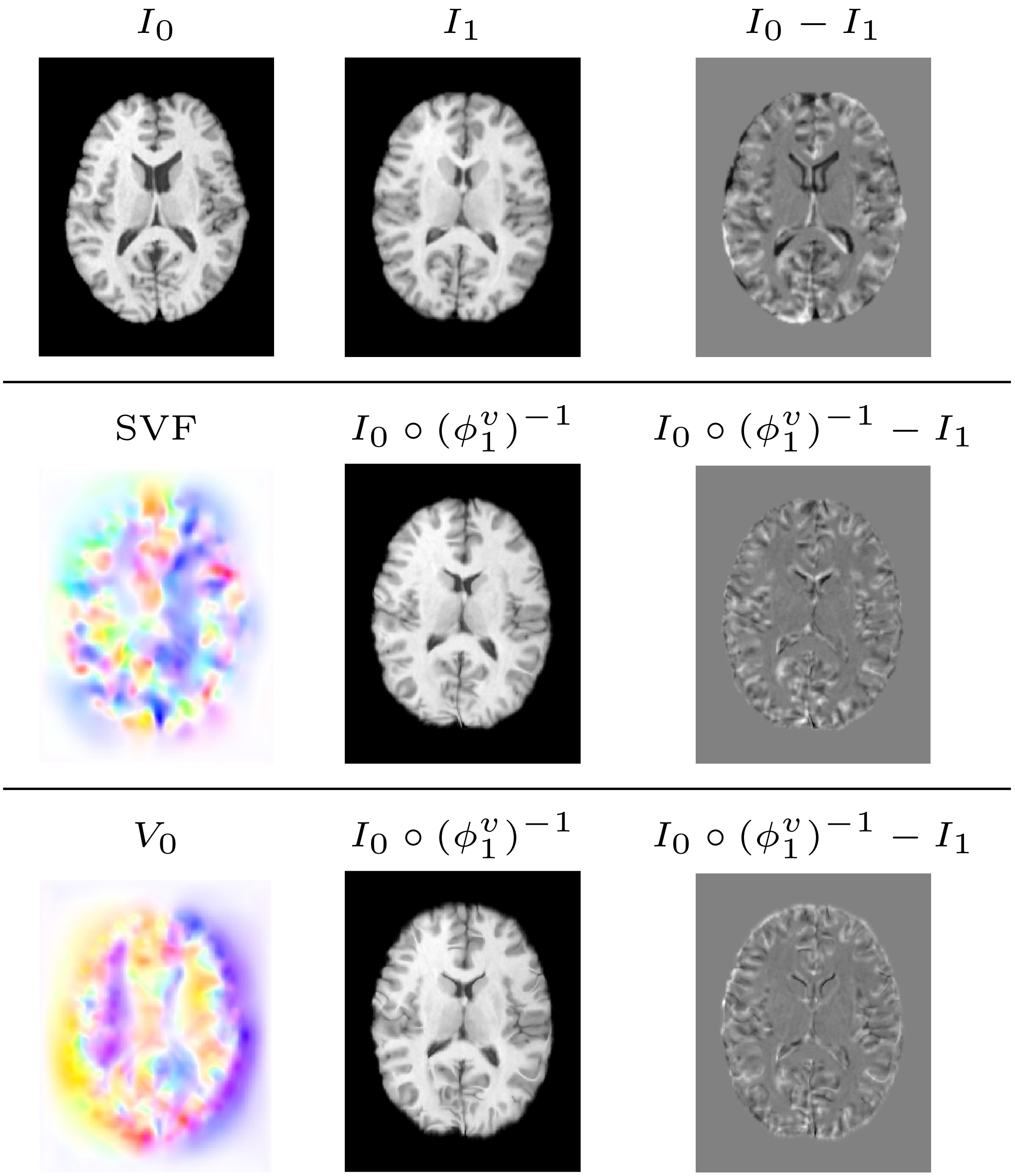
I_0
I_1
$I_0 - I_1$
SVF
$I_0 \circ (\phi_1^v)^{-1}$
$I_0 \circ (\phi_1^v)^{-1} - I_1$
V_0
$I_0 \circ (\phi_1^v)^{-1}$
$I_0 \circ (\phi_1^v)^{-1} - I_1$

Figure 4.9: Example of 3D registration results, axial view. First row, axial views of the source and the target images and the differences before registration. Second row, inferred stationary velocity field, warped image, and differences after registration for SVF-GAN. Third row, inferred initial velocity field, warped image, and differences after registration for EPDiff-GAN.

5. Conclusions

In this work, we have proposed a novel adversarial learning LDDMM method for diffeomorphic registration of 3D MRI brain images, inspired by recent literature on deep learning methods for diffeomorphic registration and combined with the LDDMM paradigm. We successfully implement three variations with different parameterizations for the deformations obtained, train the models on both 2D and 3D datasets and perform testing on an independent dataset. We additionally perform our testing on two state-of-the-art deep learning based diffeomorphic registration methods, and three model-based methods so as to compare our results.

Our experiments in 2D show that the stationary velocity fields and time dependent velocity fields parameterizations are easier to regularize in order to ensure the resulting deformations are diffeomorphic while maintaining competitive performance when trained enough. They also show how the velocity fields inferred by our network are similar to those of model-based methods. Our 3D experiments show that our models obtain competitive results with state-of-the-art model and deep learning methods, and can generate smooth diffeomorphic deformations with both the stationary and time dependent velocity fields parameterizations. In detail, our SVF-GAN achieves DSC superior to Voxelmorph II and very similar to Quicksilver, with the added benefit ours is an unsupervised method while Quicksilver is a supervised method that requires ground-truth deformations. While EPDiff-GAN achieves worse overall results, this was mostly because of its poor registration in two regions that are located in challenging locations. Though EPDiff-GAN showed worse results than SVF-GAN, this is likely because the time varying parameterization of the velocity fields is more restrictive. However, this parameterization is usually preferred for many computational anatomy applications, as they belong to geodesic paths which are the most appropriate for applications such as Principal Geodesic Analysis [38] o Geodesic Regression [39]. Finally, our proposed architecture is over 100 times faster than state-of-the-art traditional model-based methods, making it a good candidate for processing large datasets.

We believe future work in deep learning for diffeomorphic registration should focus on trying to improve registration accuracy to that of model-based methods. For this, we believe unsupervised models are the most appropriate since the only source of ground data is model-based methods. We also believe using more complex and resource-hungry neural networks is not the solution, but rather efforts should be made to incorporate more well-studied elements from model-based methods shown to improve performance, like the different parameterizations developed in this work. Another approach could be to end the paradigm of end-to-end deep learning methods and opt for hybridization with model-based methods, where deep learning could instead substitute sensitive elements like regularizers, image similarity metrics, or constraints. Finally, though most of the deep learning methods mentioned in this work are aimed solely at deformable registration, they could be adapted to be used directly in Computational Anatomy studies. One such application which is of great interest at the moment and in which brain MRIs are involved is Alzheimer's Disease diagnosis or anatomical characterization [40, 41].

6. Supplementary Material

6.1 Appendix 1

In table 6.1 a list of the cortical structures segmented in the NIREP database can be seen, along with their acronyms as used in the results presented in this work. In figures 6.1, 6.2 and 6.3 a brain MRI with color annotations for the major structures can be found. These images were obtained from `https://doi.org/10.53347/rID-61691`, where an interactive version can be found.

Occipital lobe	OL	Cingulate gyrus	CG
Insula gyrus	InsG	Temporal pole	TP
Superior temporal gyrus	STG	Infero temporal gyrus	ITG
Parahippocampal gyrus	PG	Frontal pole	FP
Superior frontal gyrus	SFG	Middle frontal gyrus	MFG
Inferior gyrus	InfG	Orbital frontal gyrus	OFG
Precentral gyrus	PreG	Superior parietal lobe	SPL
Inferior parietal lobe	IPL	Postcentral gyrus	PostG

Table 6.1: List of cortical structures manually segmented in the NIREP database and acronyms.

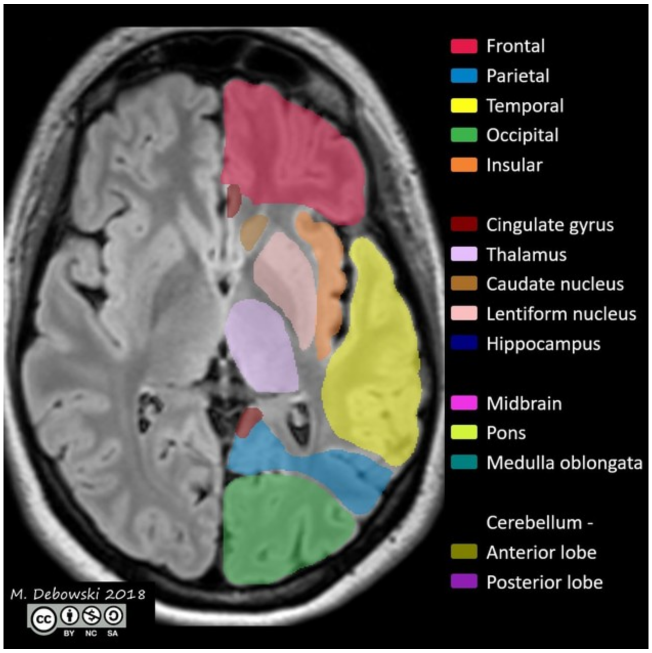

Figure 6.1: Axial view of a brain MRI with annotations of major structures, obtained from: `https://radiopaedia.org/cases/brain-lobes-annotated-mri-1`.

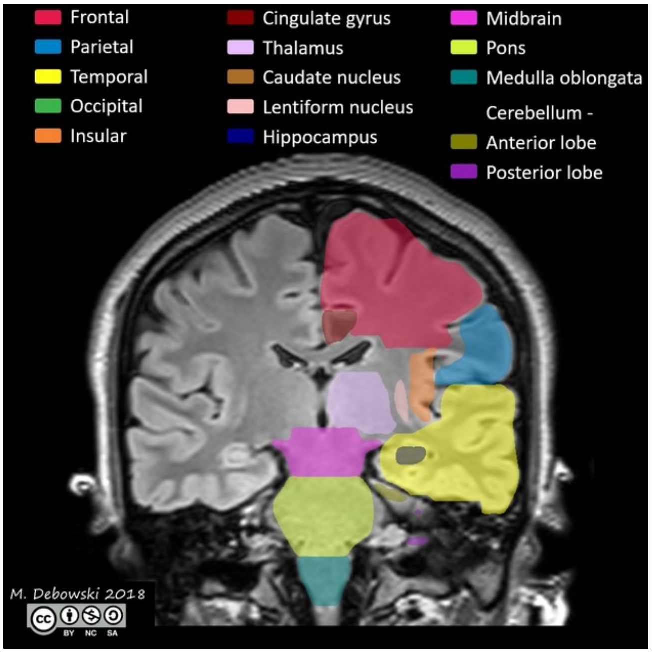

Figure 6.2: Coronal view of a brain MRI with annotations of major structures, obtained from: `https://radiopaedia.org/cases/brain-lobes-annotated-mri-1`.

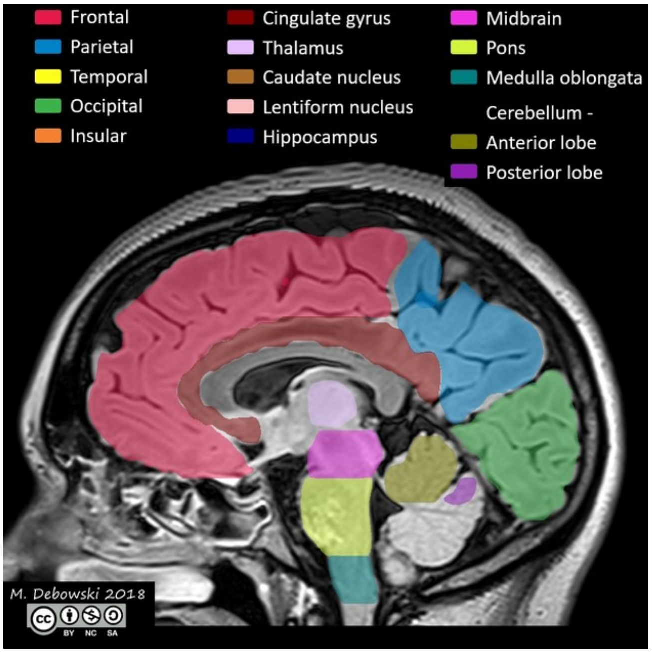

Figure 6.3: Sagittal view of a brain MRI with annotations of major structures, obtained from: `https://radiopaedia.org/cases/brain-lobes-annotated-mri-1`.

Bibliography

[1] Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Med. Image Anal. **2(3)** (1998) 243 – 260

[2] Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, J.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Med. Imaging **18(8)** (1999) 712 – 721

[3] Beg, M.F., Miller, M.I., Trouve, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vision **61 (2)** (2005) 139–157

[4] Arsigny, V., Commonwick, O., Pennec, X., Ayache, N.: Statistics on diffeomorphisms in a Log-Euclidean framework. Proc. of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), Lecture Notes in Computer Science **4190** (2006) 924 – 931

[5] Hernandez, M., Bossa, M.N., Olmos, S.: Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. Int. J. Comput. Vision **85 (3)** (2009) 291–306

[6] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic Demons: Efficient non-parametric image registration. Neuroimage **45(1)** (2009) S61–S72

[7] Miller, M.I., Trouve, A., Younes, L.: Geodesic shooting for computational anatomy. J. Math. Imaging Vis. **24** (2006) 209–228

[8] Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. Q. Appl. Math. **65** (2007) 113 – 134

[9] Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. Int. J. Comput. Vision **97(2)** (2011) 229 – 241

[10] Hernandez, M.: PDE-constrained LDDMM via geodesic shooting and inexact Gauss-Newton-Krylov optimization using the incremental adjoint Jacobi equations. Phys. in Med. and Biol. **64(2)** (2019)

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

[12] Mang, A., Biros, G.: Constrained H1 regularization schemes for diffeomorphic image registration. SIAM J. Imaging Sciences **9(3)** (2016) 1154–1194

[13] Zhang, M., Fletcher, T.: Fast diffeomorphic image registration via Fourier-Approximated Lie algebras. Int. J. Comput. Vision (2018)

[14] Hernandez, M.: Band-limited stokes large deformation diffeomorphic metric mapping. IEEE J. of Biom. and Health Inf. **23(1)** (2019)

[15] Dosovitskiy, A., Fischere, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V.: Flownet: Learning optical flow with convolutional networks. Proc. of the 16th IEEE International Conference on Computer Vision (ICCV'15) (2015) 2758 – 2766

[16] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., eds.: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham, Springer International Publishing (2015) 234–241

[17] Boveiri, H., Khayami, R., Javidan, R., Mehdizadeh, A.: Medical image registration using deep neural networks: A comprehensive review. Computers and Electrical engineering **87** (2020) 106767

[18] Rohe, M.M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: Learning deformable image registration using shape matching. Proc. of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'17), Lecture Notes in Computer Science (2017) 266 – 274

[19] Dalca, A.V., Blakrishnan, G., Guttag, J., Sabuncu, M.: Unsupervised learning for fast probabilistic diffeomorphic registration. Proc. of the 21th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'18), Lecture Notes in Computer Science (2018) 729 – 738

[20] Krebs, J., Delingetter, H., Mailhe, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE Trans. Med. Imaging (2019)

[21] Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: A learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38(8)** (2019) 1788–1800

[22] Liu, R., Li, Z., Zhang, Y., Zhao, C., Huang, H., Luo, Z., Fan, X.: A multi-scale optimization learning framework for diffeomorphic deformable registration. ArXiv (2020)

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

[24] Wang, J., Zhang, M.: DeepFLASH: an efficient network for learning-based medical image registration. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'20) (2020)

[25] Fan, J., Cao, X., Yap, P., Shen, D.: BIRNet: brain image registration using dual-supervised fully convolutional networks. Med. Image Anal. **54** (2019) 193 – 206

[26] Mahapatra, D., Antony, B., Sedai, S., Garvani, R.: Deformable medical image registration using generative adversarial networks. IEEE International Symposium on Biomedical Imaging (ISBI'18) (2018)

[27] Duan, L., Yuan, G., Gong, L., Fu, T., yang, X., Chen, X., Zheng, J.: Adversarial learning for deformable registration of brain MR image using a multi-scale fully convolutional network. Biomed. Signal Procc. Control **53** (2018) 101562

[28] Fan, J., Cao, X., Wang, Q., Yap, P., Shen, D.: Adversarial learning for mono- or multi-modal registration. Med. Image Anal. **58** (2019) 1015 – 1045

[29] Dey, N., Ren, M., Dalca, A.V., Gerig, G.: Generative adversarial registration for improved conditional deformable templates. Proc. of the 18th IEEE International Conference on Computer Vision (ICCV'21) (2021)

[30] Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Medical Image Analysis **57** (2019) 226–236

[31] Dalca, A.V., Rakic, M., Guttag, J.V., Sabuncu, M.R.: Learning conditional deformable templates with convolutional networks. In: NeurIPS. (2019)

[32] Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H.: Introduction to the non-rigid image registration evaluation project (NIREP). Proc. of 3rd International Workshop on Biomedical Image Registration (WBIR'06) **4057** (2006) 128 – 135

[33] Hernandez, M.: Gauss-Newton inspired preconditioned optimization in large deformation diffeomorphic metric mapping. Phys. in Med. and Biol. **59(20)** (2014)

[34] Ramon, U., Hernandez, M., Mayordomo, E.: Lddmm meets gans: Generative adversarial networks for diffeomorphic registration (2021)

[35] Zeiler, M.D., Taylor, G.W., Fergus, R.: Adaptive deconvolutional networks for mid and high level feature learning. ICCV 2011 (2011) 2018–2025

[36] Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts. Distill (2016)

[37] Jaderberg, M., Simonyan, K., Zissermann, A., Kavukcuoglu, K.: Spatial transformer networks. Proc. of Conference on Neural Information Processing Systems (NeurIPS'15) (2015)

[38] Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. (2013)

[39] Hong, Y., Joshi, S., Sanchez, M., Styner, M., Niethammer, M.: Metamorphic geodesic regression. Proc. of the 15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'12), Lecture Notes in Computer Science **15** (2012) 197 – 205

[40] Spasov, S.E., Passamonti, L., Duggento, A., Lio, P., Toschi, N., ADNI: A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer's disease. Neuroimage **189** (2019) 276 – 287

[41] Ramon-Julvez, U., Hernandez, M., Mayordomo, E., ADNI: Analysis of the influence of diffeomorphic normalization in the prediction of stable vs progressive MCI conversion with convolutional neural networks. IEEE International Symposium on Biomedical Imaging (ISBI'20) (2020)