PH.D. DISSERTATION

Deep Image Analysis Based on Deformable Shapes and Its Applications to Neuroimaging

Jian Wang

Department of Computer Science
School of Engineering and Applied Science
University of Virginia

April 24, 2023

APPROVAL SHEET

This
dissertation
is submitted in partial fulfillment
of the requirements for the degree
Doctor of Philosophy

This dissertation has been read and approved by the examining committee:

Advisor: Miaomiao Zhang
Committee Chair: Tom Fletcher
Committee Member: Frederick Epstein
Committee Member: Miaomiao Zhang
Committee Member: Yangfeng Ji
Committee Member: William Wells III

Accepted for the School of Engineering and Applied Science
Jennifer L. West, Dean of the School of Engineering and Applied Science
May 2023

Abstract

Deformable shapes play a pivotal role in numerous image analysis tasks, such as image registration for real-time image-guided navigation systems in tumor removal surgery, image classification to detect neuro-degenerative diseases, and template-based image segmentation for object tracking. In recent years, the advancements in deep learning-based image analysis have resulted in remarkable performance by providing a universal mechanism for extracting image features in the context of texture, intensity, or simple geometry features that may be hidden behind raw data. Nevertheless, existing deep learning methods fail to recognize geometric deformable shape features that can capture complex and detailed geometric information in images, resulting in a considerable limitation of the image analysis models when the analysis and quantification of geometric shapes are crucial.

Learning and modeling deformable shapes is particularly challenging due to their high-dimensional and non-linear nature of data, which inevitably cause expensive network training and inference with high computational complexity. In addition, current related algorithms suffer from time and labor-consuming parameter tuning and appearance change (i.e., caused by missing data values, corrupted signals, or occurrence of objects). To address these issues, the ultimate goal of my dissertation is to develop a deep learning framework that learns and analyzes efficient and robust representations of shapes central to image analysis tasks. This research naturally merges low-dimensional deformable geometric shape features in various deep learning based image analysis tasks, including but not limited to image registration, image classification, image segmentation, uncertainty quantification, atlas building, and parameter estimation. The developed framework has great impact on a variety of real-world clinical applications. For example, it allows neurosurgeons to identify brain shifts (deformations) caused by multiple factors during surgery (i.e., gravity, fluid drainage, or changes in intracranial pressure and swelling of brain tissue), and modify surgical plans in real-time. Such a framework helps clinicians better understand, interpret, and analyze the image registration results and further enables precise diagnosis according to the patient's clinical condition. This research also facil-

itates more robust clinical diagnostic routines in neuro-degenerative disease prediction, such as Alzheimer's disease detection, and post-treatment for patient care.

In particular, I first develop efficient image analysis models based on low-dimensional shape representations. The outcome of my research speeds up Bayesian uncertainty quantification for image registration, enable fast atlas building with automatic parameter selection, and allow rapid training data generation for learning-based registration regularization parameter estimation. I then develop deep neural networks to learn low-dimensional shape representations through image registration with much lower computational complexity in training. To further enhance the efficiency of the image analysis models, I further utilize learned deformable shape representations for population-based image studies by developing joint image classification models with group mean estimation, also known as atlas building, which improves model accuracy, robustness, and efficiency. Additionally, I introduce a deep metamorphic neural network that effectively controls geometric shape deformation and appearance changes (i.e., caused by tumor resection), leading to precise image registration with lower model error. My dissertation has significant potential to impact clinical applications, such as automated diagnosis for neuro-degenerative diseases and image-guided navigation systems for neurosurgery. Overall, my research provides a theoretical foundation in machine learning and computer vision, facilitating better image analysis in unmet clinical needs and neuroscience studies.

Acknowledgements

Throughout my doctoral studies, I have been fortunate to receive an abundance of valuable knowledge and support from numerous individuals, and I would like to express my heartfelt gratitude to all of them. Although it is impossible to mention everyone by name, I want to acknowledge the significant impact that so many people have had on my journey. My first two years of study at Lehigh University and Washington University in St. Louis were a busy but fulfilling time. I gained a deeper understanding of the American educational system and developed a solid foundation for my subsequent research projects at the University of Virginia (UVa). I have fond memories of carefree nights around the campfire at Lehigh and the breathtaking scenery of Mark Twain National Park in the Midwest.

First and foremost, I would like to express my sincere gratitude to my dissertation committee members, Dr. Tom Fletcher, Dr. Frederick Epstein, Dr. Yangfeng Ji, and Dr. Sandy Wells. Since 2020, they have generously shared their invaluable insights and suggestions, expertly guiding me towards the optimal research direction. Their dedication to advancing our research has been truly inspiring. I deeply thank my coauthors, Dr. Polina Golland, and Dr. Michael Kirby for their collaborations. Their professional suggestions and expertise have been integral in shaping the direction of my work and I feel fortunate to have had the opportunity to collaborate with such brilliant minds. I would like to thank my junior Jiarui Xing, my internship mentor Dr. Adam Harrison, and Dr. Jason Druzgal from the Radiology and Medical Imaging Department at UVa for their guidance and support with data, programming, and research ideas. I also would like to extend a special thanks to my dear friends Dr. Shaoxia Zhang, Dr. Jie Wang, Hang Shao, Jiangtao Lang, Wencan Guan, Peiyun Zhao, Ge Fang, Josie Ho, Wendy Qi, Pierce Corson and Jun Li. Their unwavering support and encouragement have motivated me to persevere through the challenges of my doctoral studies. The friendship has been a source of inspiration, and I am grateful for the love and care they have shown me. Of course, I must express my deepest appreciation for my advisor, Dr. Zhang Miaomiao. Her patience, expertise, and guidance have been invaluable to me throughout my entire doctoral research. She has always been willing to answer my questions and shed light when I encountered research bottlenecks. Under

her tutelage, I have grown as a scholar and researcher, and I am forever grateful for her mentorship.

Lastly, I want to thank my beloved family, my father Bao Wang, my mother Lufang Liang, my sister Jia Wang, and three wonderful children Yelei Liu, Yinuo Liu, and Yihao Liu. Your unconditional love and support have been a constant source of comfort and inspiration. Your cheerfulness and optimism have helped me to remain focused and motivated even during the most challenging times. I am incredibly fortunate to have such a loving family, and I love you all.

Acknowledgement in Chinese

在博士学习和完成论文答辩期间，我学到了许多宝贵的知识，也得到了很多人的帮助。因此，在此我想要特别感谢所有出现在我博士学习历程中的人们，是你们让我变得更勇敢、坚韧、乐观和从容。尽管有很多人帮助了我，但由于篇幅所限，我无法一一列举所有人，对此我深表歉意。

我的博士学习始于理海大学和圣路易斯华盛顿大学。两年的课程学习很充实，完成课程作业的过程让我更加了解美国的教学体系，为之后着手科研项目打下了坚实的基础。感谢我的同学和朋友们在生活上对我的支持和帮助。我会时常回忆起在理海大学篝火之夜自由自在的时光，以及在中西部的马克吐温国家公园欣赏壮丽景色的美好时刻。

2019年，我搬到弗吉尼亚大学继续完成博士学位的学习。在此期间，我想首先感谢我的博士论文委员会的教授们: Tom Fletcher, Frederick Epstein, Yangfeng Ji和Sandy Wells。自2020年以来，他们慷慨地分享了他们的宝贵见解和建议，专业地指导我走向最佳的研究方向。 此外，感谢我的合作者们: Polina Golland教授和Michael Kirby教授。他们专业的建议对我的研究方向的塑造起到了至关重要的作用，能与这样杰出的研究者合作我感觉非常幸运。我要感谢我的师弟邢家瑞、我的实习主管Adam Harrison、以及弗吉尼亚大学医学影像中心的Jason Druzgal教授，他们对我在科研项目中的数据、编程和思路上给予了很大的帮助。特别感谢我的挚爱朋友们张少霞、王婕、邵航、朗江涛、关文灿、赵佩芸、方戈、何艳娟、祁婉悦、柯品圣和李君。谢谢你们一直以来的陪伴和支持，帮助我认识到生命的不同层面的意义。是你们的真诚和关心让我感到世界上还有无血缘关系的亲情存在，你们无时无刻的鼓励让我更有动力完成学业。能够完成整个博士期间的研究，我想特别感谢我的导师张妙妙教授，她耐心地教导我，回答我在科研中的每一个疑问，培养我从一无所知的学生成长为一个愿意不断学习和专心研究的人。最后，我要感谢我的家人们：爸爸、妈妈、姐姐，以及我们家族的三个可爱的孩子。你们的开朗和乐观时时刻刻感染着我，让我永远感到无比轻松、快乐和自在。谢谢你们无私的关心和爱，我感到非常幸运能够出生在这样一个温暖的家庭里。我爱你们。

To myself and my family.

Table of Contents

Chapter I

Introduction

Shape information is important to the field of image analysis, with applications in various scientific research fields, such as machine learning, medical imaging, computer vision, and biology [200, 20, 72, 103, 114, 197]. The representation of geometric shapes is one of the fundamental features in images and play a pivotal role in a wide range of image analysis models, including but not limited to image segmentation, recognition [164, 167], classification [60, 11], registration [180, 177, 203, 206], and retrieval [128, 192, 29]. Various geometrical shape representations, including landmarks [16, 23, 40], point clouds [1], binary segmentation [25, 163], Fourier descriptors [131] and medial axes [132], have been studied in bountiful literatures. For instance, recent works of learning skeleton-based point clouds shape representations [82, 196] dynamically animate and reconstruct 3D human shapes from raw image data. Geometric deep learning methods [26, 139] render and reconstruct images via learning shape representations in the form of analytic graphs or points. Shape-guided image generation techniques successfully synthesize realistic images while combining shape and appearance representations with different input queries [53].

The aforementioned shape representations capture simple geometric structure of objects and have limited ability to characterize fine-detailed geometric shape information presented in images. Therefore, many important geometric shape features are oftentimes neglected, or underutilized as an implicit knowledge in image analysis tasks. In contrast, deformation-based shape descriptors (based on elastic deformations or diffeomorphic flows) focus on highly detailed shape information from images [36, 149]; hence offering more flexibility in describing changes and variability of shapes with complex structures. Such deformable shape representations are valuable asset in a wide range of healthcare systems, such as pathological disease progression monitoring [18, 60], efficient management of clinical interventions [106, 179, 198], intraoperative quantitative functional anal-

ysis [144, 122], and radiation therapy planning [87, 141, 151]. The use of deformable shapes has also been successful in numerous neuroimaging tasks, for example, a real-time image-guided neurosurgery navigation to evaluate brain shift [106, 178, 57], and abnormality detection systems that analyze shape variations of anatomical structures (i.e., brain shrinkage or expansion caused by neurodegenerative disease [21, 77, 180, 204]). Deformation-based shape descriptors can also be applied in other clinical scenarios, including irregular placental shape detection [190], cardiac motion estimation via shape variability analysis [184], and automated cardiac image segmentation with shape priors [30]. Injecting pre-learned ventricular shape priors into deep neural networks significantly improves the performance of cardiac segmentation, providing more accurate segmentation labels to assess cardiac morphology [30].

Deformable shape representations, despite their significant value, are not fully exploited in current machine/deep learning paradigms for image analysis. On the other hand, commonly used deep learning models often struggle to effectively capture complex and intrinsic shape features. The existence of this gap motivates us to explore how incorporating deformable shape learning into deep learning paradigms can impact performance. Existing deep learning-based models typically extract and learn image features hidden behind the raw data through deep neural networks [41, 47, 55], and achieves great performance in various image analysis tasks, e.g., image classification [71, 13], segmentation [90] and reconstruction [82]. However, existing approaches lack the ability to learn and quantify deformable shapes that are central to image analysis tasks [1, 82, 90]. An explicit deep learning approach for deformable shape learning, which captures most of the detailed and complex image features, has been missing. This greatly limits the power of deep learning based models in image analysis where analyzing shapes is important. Learning sufficient deformable geometric representations enables one to properly estimate the transformations between images and extract shape variability from population-based data, which also holds a key to extracting critical diagnostic information and investigating effective treatments in neuroimaging. In this dissertation, I will focus on merging deformable shape representation learning into different deep learning-based tasks, including image classification, registration, and atlas construction. Particularly, I use diffeomorphic transformations as desired geometric shape descriptors due to the nice properties of transformations, including differentiable and bijective mappings with differentiable inverses, and the invertibility of deformation fields [15, 116]. This guarantees the correctness of biological information of objects, and no artifacts in deformed images.

1 Problems and Challenges

A major challenge of learning deformable shape representations is the high-dimensional and non-linear nature of medical image data (e.g., the dimension of a 3D magnetic resonance imaging (MRI) scan of a brain is 256^3). This results in high computational complexity of deformable shape modeling [15, 161] and inference algorithms [195]. Image registration is one of the basic concepts to obtain geometric shape descriptors from one image to another. To improve the efficiency of image registration, state-of-the-arts typically adopt convolutional neural networks (CNN) based models to extract associated spatial transformations as mappings for pairwise images or population groups in training dataset [194, 145, 14, 35, 28]. While the aforementioned deep learning approaches can rapidly predict the deformation parameters in testing, the training process is slow and memory intensive due to the high dimensionality of deformation parameters in the image space.

The second challenge is that existing methods often need labor-consuming manual parameter tuning since the selection of registration parameters (e.g., smoothness level of transformations) dramatically affects the quality of the deformable shape descriptor. Optimal parameter values can differ significantly across image anatomy of different subjects, and even small changes can have a large impact on accuracy. Choosing appropriate hyperparameter values is therefore a crucial step and difficult task in developing, evaluating, and deploying image registration-based models. Current models either exhaustively search for an optimal regularization in the parameter space [81, 170, 150], or treat them as unknown variables to be estimated from Bayesian models that sample in a high-dimensional transformation space [159, 158, 207]. It inevitably results in expensive computational cost and leads to a long execution time with high memory consumption.

The third challenge is that traditional image registration algorithms may fail due to the appearance change caused by missing data issues, i.e., corrupted signals, or brain tumor resection. This is because the similarity metric of diffeomorphic image registration is often measured either by comparing image intensities directly or using intensity-based measures like mutual information or cross-correlation [124]. However, for images with pathologies (as geometric metamorphosis), assumptions of structural and intensity similarity may fail. Previous methods of handling missing data issues are either utilizing metamorphic image registration [124, 57, 110, 22] or adopting segmentation as indicators to mask out the appearance changes. However, the transformed images still contain artifacts, which hardly maintain the structural anatomical details.

2 Dissertation Statements and Contributions

This dissertation outlines a comprehensive framework that utilizes low-dimensional deformable geometric shape analysis and learning to improve medical image analysis across various tasks. The framework has been designed to be highly efficient, robust, and stable, making it suitable for clinical interventions where tracking deformable shape changes is crucial for accurate diagnoses and effective treatments. The framework's ability to effectively analyze medical images has the potential to significantly enhance the diagnosis and treatment of medical conditions, ultimately improving patient outcomes.. To achieve this goal, I developed the following algorithms: i) fast predictive deformable image registration, reliable registration uncertainty quantification and predictive parameter estimation with much lower computational complexity; ii) efficient and robust population-based study algorithms such as atlas building and image classification with a better model interpretability; iii) precise metamorphic image registration where images have appearance change (caused by missing data values, pathologies and geometric object occurrence). To test this thesis statement, I have made the following contributions:

1. A learning-based predictive image registration model, names as DeepFLASH, is presented. It substantially speeds up the training, testing and inference of current deep learning based registration algorithms via low-dimensional Fourier representations. The learned low-dimensional deformable shape representations enable various post-registration algorithms, such as fast uncertainty quantification, rapid data generation for regularization parameter estimation, and hierarchical Bayesian atlas building;
2. A classification model, named as Geo-SIC, which incorporates deformable shapes learning is presented for population study. Geo-SIC simultaneously learns deformable shapes from a deformation space and features from image space for an improved performance of image classification. Accessing underlying geometric features also gains increased model interpretability;
3. A predictive model, named as MetaMorph, for metamorphic registration of images with appearance changes (i.e., caused by brain tumors) is presented. I develop a piecewise regularization on the tangent space of diffeomorphic transformations via learned segmentation maps of abnormal regions. The geometric transformation and appearance changes are treated as joint tasks that are mutually beneficial. MetaMorph has great potential in various image-guided clinical interventions, e.g., real-time image-guided navigation systems for tumor removal surgery.

3 Outline

The remainder of this dissertation is organized as follows:

Chapter II introduces the background of deformable shape representations. The first section briefly overviews common models to derive geometric shapes, including basic shape representations (skeleton-based, point cloud, medial axis, Fourier basis and segmentation/landmarks) and deformation-based methods, e.g., deriving geometric features from elastic/diffeomorphic image registration. The second section describes the concepts for statistical shape modeling such as atlas-based geometric shape representation.

Chapter III presents a comprehensive analysis of low-dimensional shape representations in multiple medical image analysis tasks, such as image registration, uncertainty quantification, parameter estimation and atlas building. The first section introduces a low-dimensional reduced-order image registration algorithm. The second section presents a low-dimensional Bayesian model for registration uncertainty quantification. The third section shows a hierarchical Bayesian model with automatic best parameter selection via low-dimensional shape characterizations in atlas building.

Chapter IV presents the deep neural networks that I developed by low-dimensional geometric deformable shape representations. The first section presents an efficient deep neural network, DeepFLASH, which are characterized by compact deformable shape representations for diffeomorphic image registration. It fundamentally lowers the model computational complexity of neural networks, thus resulting in fast model training, testing inference. The second section presents a classification network, Geo-SIC, which incorporates deformable geometric shape learning into deep image classifiers with boosted performance guaranteed. The third section presents an efficient predictive approach for regularization parameter estimation that automatically produces regularization parameters via deep neural networks. The developed low-dimensional Bayesian framework speeds up the training data generation plus the parameter estimation while maintaining comparable registration results.

Chapter V presents a accurate and robust predictive image registration model, MetaMorph, for metamorphic registration of images with appearance changes (i.e., caused by brain tumors). A newly-derived piecewise regularization on the tangent space of diffeomorphic transformations is developed via learned segmentation maps of abnormal regions. The geometric transformation and appearance changes are treated as joint learning tasks that are mutually beneficial. MetaMorph has great potential in various image-guided

clinical interventions, e.g., real-time image-guided navigation systems for tumor removal surgery.

Chapter VI concludes the dissertation with a general discussion and future work.

Chapter II

Background

1 Geometric Shape Representations

In this chapter, I briefly review different types of geometric shape representations. I first introduce commonly-used basic geometric shape representations, including medial axis, segmentation and landmark based shape representations, point cloud and Fourier shape descriptors. I then introduce deformable shape representations, including shape representation modeling based on elastic, diffeomorphic deformations. I further review the concepts of statistical shape representation modeling.

Medial axis. The medial axis (also referred to as the ”topological skeleton”) is a mathematical concept, proposed by Blum [19], used to describe the center line or skeleton of a shape. It is a set of points in a plane or 3D space that are equidistant to two or more boundaries or edges of a shape. For instance, in a two-dimensional geometric shape, such as a circle or a square, the medial axis is characterized as a single line that runs down the middle of the shape. However, tiny change to an object‘‘s boundary can cause a large change in the original medial axis algorithm. To improve the stability of medial axis transform, Katz et. al [88] further propose a more stable and robust method for instrumenting the Blum medial axis transform via performing perceptual part-decomposition of objects. This method considers objects to be a collection of solid parts and a set of connections among those parts. The medial axis based shape representations are considered as basic shape descriptors that can be used for path planning, fundamental shape analysis, single object recognition, and other areas where understanding the geometry and structure of shapes is important [132].

Segmentation and landmarks. Segmentations and landmarks based shape representation are methods for describing the shape of an object or structure by identifying and charac-

terizing specific points, known as landmarks, on its surface or boundaries [25, 163]. These landmarks are typically chosen based on their anatomical or functional significance and are used as reference points for describing the shape and spatial relationships of the object. Landmark-based shape representation has many applications in various fields, including medical imaging, biology, anthropology, and computer graphics. For example, In anthropology, landmarks are important to study the morphology of a certain structure, such as leaves, bones and fossils [171]. In medical imaging, landmarks oftentimes are used to analyze the shape and structure of organs and tissues and to track changes over time in response to disease or treatment, e.g., Alzheimer‘‘s Disease diagnosis using landmark-based features from longitudinal structural MRIs [99].

Point clouds. A point cloud is a collection of points in a 3D coordinate system that represent the surface of an object or a scene. Point clouds are used as a way to represent the shape of an object without explicitly modeling its surface geometry [97]. One way to represent a shape using point clouds is to sample points uniformly on its surface, which is called surface reconstruction. Another way is to use depth sensors or Light Detection and Ranging (LiDAR) to capture the 3D positions of points on the object's surface [48]. These points can then be organized into a point cloud, which can be used for shape analysis and processing. Point cloud shape representation has shown to be a promising approach for medical image analysis, especially in segmentation tasks. In medical imaging, segmentation is crucial for identifying different anatomical structures in an image, such as organs or tumors, which can assist in diagnosis, treatment planning, and image-guided interventions. Point cloud segmentation involves the labeling of each point in the point cloud as belonging to a particular anatomical structure or background. One common approach is to use clustering techniques [155], which group together similar points based on their spatial and geometric properties. Another approach is graph-cut [152], which assigns labels to the points by minimizing an energy function that models the cost of assigning each point to a particular structure. Deep learning-based methods, such as convolutional neural networks (CNNs), have also been applied to point cloud segmentation tasks with impressive results. However, point clouds can be noisy and sparse, which can make shape analysis and processing challenging. Various techniques such as point cloud denoising, upsampling, and feature extraction have been developed to address these challenges [52].

Fourier descriptors. Fourier descriptors are a popular set of shape descriptors obtained from the Fourier transformation of a shape's boundary or contour [174]. They represent the shape globally, capturing its geometric properties such as curvature, size, and orientation. Unlike other local descriptions of the contour such as polylines or splines, Fourier descriptors provide a holistic representation where each descriptor component corresponds

to a specific shape feature. Thus, modifying any component of the descriptor will affect the entire shape, making Fourier descriptors particularly useful for capturing coarse shape properties with just a few numeric values. One advantage of Fourier descriptors is their ability to capture shape information at multiple scales [93]. Adding or removing descriptor components can increase or decrease the level of detail captured in the representation, making Fourier descriptors adaptable to various tasks. By using Fourier coefficients instead of raw curvature values [51], the descriptors become invariant to rotation, translation, and scale. This is because rotation, translation, and scale operations do not change the frequency content of the shape, but only its position and size in space. Since the Fourier coefficients capture the frequency content of the shape, they remain the same regardless of the object's orientation, position, or size. Fourier descriptors can be used to compare and identify different shapes based on their Fourier coefficients, enabling applications such as shape recognition, classification, object retrieval and tracking, as well as robotics and automation for object manipulation and grasping.

1.1 Deformable Shape Representations via Image Registration

As addressed in Chapter I, aforementioned shape representations have limit ability to capture complex shape structures, therefore we introduce deformable shape modeling in this section. Deformable shapes can be derived from deformable image registration, which is a computational technique to align or match two or more images. The goal of deformable image registration is to find a deformation field (deformable shape descriptor) that maps each voxel in one image to its corresponding voxel in the other image, taking into account any differences in position, orientation, or shape between the images.

Let S be the source image and T be the target image defined on a d-dimensional torus domain $\Gamma = \mathbb{R}^d/\mathbb{Z}^d$ ($S(x), T(x) : \Gamma \to \mathbb{R}$). The optimization function for deformable image registration is,

$$E(\phi) = \mathrm{Dist}[S \circ \phi^{opt}, T] + \mathrm{Reg}[\phi], \tag{1}$$

where ϕ^{opt} denotes the optimal deformation field. The distance function $\mathrm{Dist}(\cdot,\cdot)$ measures the image dissimilarity between the source and the deformed image. Commonly used distance functions include sum-of-squared difference of image intensities [15], normalized cross correlation [9], and mutual information [185]. The regularization term $\mathrm{Reg}(\cdot)$ is a constraint that enforces the spatial smoothness of transformations. Various algorithms with different regularizations exist for calculating the deformation field, including optical flow, Demons algorithm, and B-spline-based registration.

1.1.1 Deformable Shape Representations via Elastic Deformation Elastic image registration formula adopts a regularity term that directly constrains on deformation fields (or displacement field) to enforce the smoothness of transformations,

$$\mathrm{Reg}(\phi) = \int_{\Omega} \|\phi\|^2 \, d_{\Omega}. \tag{2}$$

The gradient of optical flow is typically computed using partial derivatives of the image intensity with respect to different directions. The regularization results in Laplacian terms, which has the smoothing effect while searching for the optimal transformation field[95].

1.1.2 Deformable Shape Representations with Diffeomorphisms In many applications, it is natural to require the deformation field to be a diffeomorphism, i.e., a differentiable, bijective mapping with a differentiable inverse. Shape representations in the space of diffeomorphic transformations highlight a set of desirable features: (i) they capture large geometric variations within image groups; (ii) the topology of objects in the image remains intact; and (iii) no non-differentiable artifacts, such as creases or sharp corners, are introduced. Moreover, a theoretical framework of Large Deformation Diffeomorphic Metric Mapping (LDDMM) defines a metric in the space of diffeomorphic transformations that in turn induces a distance metric on the shape space [15].

Given an open and bounded d-dimensional domain $\Omega \subset \mathbb{R}^d$, we use $\mathrm{Diff}(\Omega)$ to denote a space of diffeomorphisms (i.e., a one-to-one smooth and invertible smooth transformation) and its tangent space $V = T\mathrm{Diff}(\Omega)$. The LDDMM algorithm [15] provides a distance metric in the deformation-based shape space, which is used as a regularization of atlas building in Eq (1). Such a distance metric is formulated as an integral of the Sobolev norm of the time-dependent velocity field $v_n(t) \in V (t \in [0, 1])$ in the tangent space, i.e.,

$$\mathrm{Reg}(\phi_n) = \int_0^1 (\mathcal{L}v_n(t), v_n(t))\, dt, \quad \text{with} \quad \frac{d\phi_n^{-1}(t)}{dt} = -D\phi_n^{-1}(t) \cdot v_n(t), \tag{3}$$

where $\mathcal{L} : V \to V^*$ is a symmetric, positive-definite differential operator that maps a tangent vector $v(t) \in V$ into its dual space as a momentum vector $m(t) \in V^*$. We typically write $m(t) = \mathcal{L}v(t)$, or $v(t) = \mathcal{K}m(t)$, with $\mathcal{K}$ being an inverse operator of $\mathcal{L}$. In this paper, we adopt a commonly used Laplacian operator $\mathcal{L} = (-\alpha\Delta + \mathrm{Id})^3$, where α is a weighting parameter that controls the smoothness of transformation fields and Id is an identity matrix. The $(\cdot, \cdot)$ is a dual pairing, which is similar to an inner product between vectors. The operator D denotes a Jacobian matrix and $\cdot$ represents an element-wise matrix multiplication.

According to a well known geodesic shooting algorithm [173], the minimum of Eq (3) is uniquely determined by solving a Euler-Poincaré differential equation (EPDiff) [5, 116]

with a given initial condition. This nicely proves that the deformation-based shape descriptor ϕ_n can be fully characterized by an initial velocity field $v_n(0)$, which lies in the tangent space of diffeomorphisms.

$$\frac{\partial v_t}{\partial t} = -K\left[(Dv_t)^T \cdot m_t + Dm_t \cdot v_t + m_t \cdot \operatorname{div} v_t\right], \tag{4}$$

A recent work has shown that the velocity fields generated by EPDiff Eq (4) can be efficiently captured via a low-dimensional representation in Fourier space with bandlimited signals [203, 206, 204]. The key idea behind is that the velocity fields do not develop high frequencies and only a small amount of low frequencies contributes to the computation of generating deformation fields (as shown in Fig. 1). Therefore, we are able to capture the deformations in a bandlimited space as accurately as the original space.

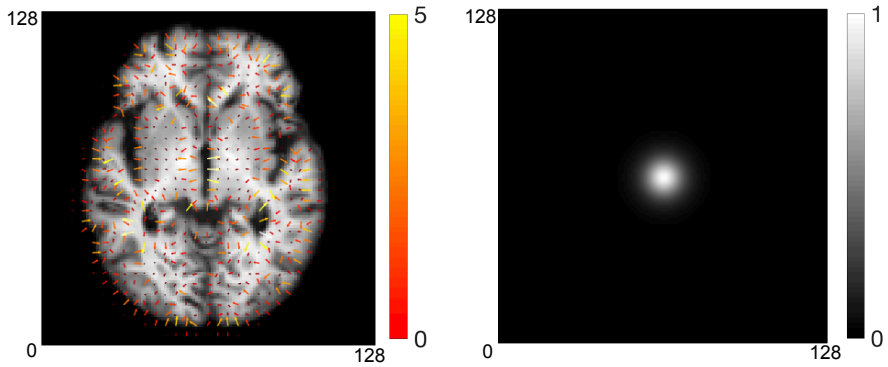


Fig. 1: Left to right: an example of 2D velocity field visualized in image space vs. Fourier space. The white dot centered in middle displays centered low frequencies.

$$\frac{\partial \widetilde{v}_n(t)}{\partial t} = -\widetilde{\mathcal{K}}\left[(\widetilde{\mathcal{D}}\widetilde{v}_n(t))^T \star \widetilde{\mathcal{L}}\widetilde{v}_n(t) + \widetilde{\nabla} \cdot (\widetilde{\mathcal{L}}\widetilde{v}_n(t) \otimes \widetilde{v}_n(t))\right], \tag{5}$$

where $\star$ is the truncated matrix-vector field auto-correlation. Here $\widetilde{\mathcal{K}}$ is an inverse operator of $\widetilde{\mathcal{L}}$, which is the Fourier transform of a Laplacian operator $\mathcal{L}$. The operator $\widetilde{\mathcal{D}}$ represents the Fourier frequencies of a Jacobian matrix with central difference approximation, and $*$ is a circular convolution with zero padding to avoid aliasing [1]. The operator $\widetilde{\nabla}\cdot$ is the discrete divergence of a vector field, and $\otimes$ represents a tensor product between Fourier frequencies.

[1] To prevent the domain from growing infinity, we truncate the output of the convolution in each dimension to a suitable finite set.

2 Statistics of Diffeomorphic Shapes

This section briefly reviews the concept of atlas building [85], which is commonly used to derive deformation-based shape representations from images. With the underlying assumption that the geometric information in the deformations conveys a shape, descriptors in this class arise naturally by matching a template to an input image with smoothness constraints on the deformation field.

Given a number of N images $\{I_1, \cdots, I_N\}$, the problem of atlas building is to find a mean or template image I and deformation fields $\phi_1, \cdots, \phi_N$ that minimize the energy function

$$E(I, \phi_n) = \sum_{n=1}^{N} \frac{1}{\sigma^2} \text{Dist}[I \circ \phi_n^{-1}, I_n] + \text{Reg}(\phi_n), \tag{6}$$

where σ^2 is a noise variance and $\circ$ denotes an interpolation operator that deforms image I with an estimated transformation ϕ_n. The $\text{Dist}(\cdot, \cdot)$ is a distance function that measures the dissimilarity between images, i.e., sum-of-squared differences [15], normalized cross correlation [9], and mutual information [185]. The $\text{Reg}(\cdot)$ is a regularization that guarantees the smoothness of transformations.

2.1 Atlas-based Geometric Shape Representations

We are now ready to optimize the problem of atlas building (Eq (6)) with reduced computational complexity in a low-dimensional bandlimited space as

$$E(I, \phi_n) = \sum_{n=1}^{N} \frac{1}{\sigma^2} \text{Dist}[I \circ \phi_n^{-1}, I_n] + (\mathcal{L}\widetilde{v}_n(0), \widetilde{v}_n(0)), \quad \text{s.t. } Eq\ (3) \& \ Eq\ (5). \tag{7}$$

The deformation ϕ_n^{-1} corresponds to $\widetilde{\phi}_n^{-1}$ in Fourier space via the Fourier transform $\mathcal{F}(\phi_n^{-1}) = \widetilde{\phi}_n^{-1}$, or its inverse $\phi_n^{-1} = \mathcal{F}^{-1}(\widetilde{\phi}_n^{-1})$. Note that we will drop the time index t, i.e., $\widetilde{v}_n(0) \triangleq \widetilde{v}_n$, for simplified notations in next chapters.

Chapter III

Low-dimensional Geometric Deformable Shape Analysis

In this chapter, an extensive analysis of low-dimensional shape representations is provided on different medical image analysis tasks, including image registration, uncertainty quantification, regularization parameter estimation and atlas building. The first section Sec. 1 of this chapter presents a reduced-order image registration algorithm that operates in low-dimensional space that are parameterized by orthogonal basis. The second section Sec. 2 presents a low-dimensional Bayesian model that can be used for registration uncertainty quantification, which efficiently offers a probabilistic confidence map on top of registration solution. The third section Sec. 3 demonstrates the use of hierarchical Bayesian models for atlas building, where low-dimensional shape characterizations are employed to automatically select the best regularization parameters.

1 Data-Driven Model Order Reduction for Diffeomorphic Image Registration

1.1 Related Work

Diffeomorphic image registration has been successfully applied in the field of medical image analysis, as it maximally maintains the biological correctness of deformation fields in terms of object topology preservation. Examples of applications include alignment of functional data to a reference coordinate system [36, 207], anatomical comparison across individuals [169, 137], and atlas-based image segmentation [7, 86]. The problem of image registration is often formulated as constrained optimization over the transformation

that well aligns a source image and a target image. A plethora of transformation models to today fit into various categories of parameterizations, e.g., stationary velocity fields that remain constant over time [6], and time-varying velocity fields in the framework of LDDMM [15]. We focus on the latter as it supports a distance metric in the space of diffeomorphisms that is critical for deformation-based statistical shape analysis, such as least squares mean estimator [86], geodesic regression [125, 162], anatomical shape variability quantification [137], and groupwise geometrical shape comparison [169].

Despite the aforementioned advantages, one major challenge that hinders the widespread use of LDDMM is its high computational cost and large memory footprint [37, 15, 173]. The algorithm inference typically requires costly optimization, particularly on solving a full-scale regularization term defined on a dense image grid (e.g., a brain MRI with size of 256^3). Prior knowledge in the form of regularization is used to enforce the smoothness of transformation fields, also known as geodesic constraints in the space of diffeomorphisms, by solving a set of high-dimensional PDEs [116, 173]. This makes iterative optimization approaches, such as gradient descent [15], BFGS [134], or the Gauss-Newton method [8], computationally challenging. While improved computational capabilities have led to a substantial run-time reduction, such solution of a single pairwise image registration still takes tens of minutes to finish on dense 3D image grids [162].

We aim at an approximate inference method that significantly lowers the computational complexity with little to no impact on the alignment accuracy. Our solution is motivated by the observation that smooth vector fields in the tangent space of diffeomorphisms can be characterized by a low-dimensional geometric descriptor, including a finite set of control points [50], or Fourier basis functions representing low frequencies [203, 206]. As a consequence, we hypothesize that the solution to high-dimensional PDE systems can be effectively approximated in a subspace with much lower dimensions. We develop a data-driven model reduction algorithm that constructs a low-dimensional subspace to approximate the original high-dimensional dynamic system for diffeomorphic image registration. We employ proper orthogonal decomposition (POD), a widely used technique for PDE systems, in which the approximating subspace is obtained from a discretized full-order model at selected time instances. A reduced-order model can then be constructed through Galerkin projection methods [39], where the PDEs are projected onto a compact set of POD eigen-functions.

To the best of our knowledge, this method has not yet been applied to large systems of PDEs such as the one employed in diffeomorphic registration. While we focus on the context of LDDMM, the theoretical tools developed in this work are broadly applicable to all PDE-constrained diffeomorphic registration models. To evaluate the proposed method,

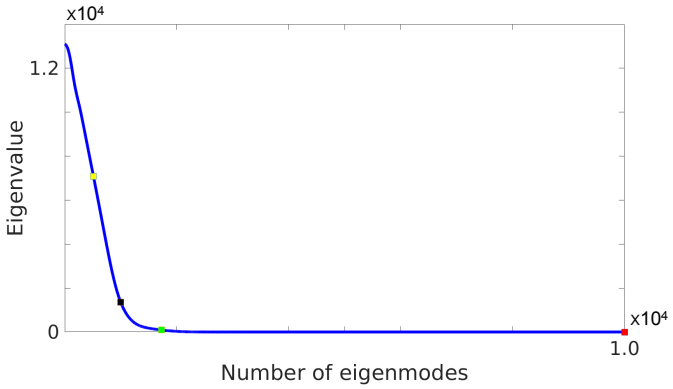


Fig. 2: An example eigenvalue plot of velocity fields generated from 2D synthetic data.

we perform image registration of real 3D MR images and show that the accuracy of our estimated results is comparable to the state of the art, while with drastically lower runtime and memory consumption. We also demonstrate this method in the context of brain atlas building (mean template estimation) for efficient population studies.

1.2 Method: Model Order Reduction (ROM) For Diffeomorphic Image Registration

We develop a POD-based model order reduction algorithm, particularly for the registration regularization term governed by high-dimensional PDEs (EPDiff), to approximate a subspace via a given set of velocity fields in an optimal least-square sense. We then derive a Galerkin projection (orthogonal projection) of EPDiff equations onto the subspace to obtain a reduced-order model.

1.2.1 Low-dimensional Subspace of Velocity Fields Given a set of full-dimensional velocity fields $\{\mathbf{v}_t\} \in \mathcal{V}^q$, e.g., $q = 3 \times 128^3$ for a 3D discretized image grid with the size of 128^3, we are seeking an approximated subspace $\mathcal{U}^r = \text{span}\{\mathbf{u}_1, \cdots, \mathbf{u}_r\} \subset \mathcal{V}^q \, (r \ll q)$, where $\mathbf{u}_i$ is the basis, to best characterize our data. A projection from such low-dimensional subspace to the original space can be effectively performed by $\mathbf{v}_t = \mathbf{U}\boldsymbol{\alpha}_t$, where $\mathbf{U}^{q\times r} = [\mathbf{u}_1, \cdots, \mathbf{u}_r]$ and α_t is a r-dimensional vector representing factor coefficients. Here, we require the basis vectors to be orthonormal, i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. The inverse projection can thus be written as $\boldsymbol{\alpha}_t = \mathbf{U}^T\mathbf{v}_t$. Our objective is to minimize the projection error defined in the tangent space of diffeomorphisms

$$\arg\min_{\mathbf{U}} \int (\mathbf{L}(\mathbf{v}_t - \mathbf{U}\boldsymbol{\alpha}_t), \mathbf{v}_t - \mathbf{U}\boldsymbol{\alpha}_t) \; dt. \tag{8}$$

where $\mathbf{L}$ is the discrete operator of L defined in Eq (3). The minimization problem of Eq (8), is the classic problem known as Karhunen-Loève decomposition or principal component analysis, and holds an equivalent solution to the following eigen decomposition problem of a covariance matrix $\mathbf{C}^{q\times q}$ [123, 154],

$$\mathbf{C}\mathbf{u}_i = \lambda_i \mathbf{u}_i, \text{ with } \mathbf{C} = \int \mathbf{L}(\mathbf{v}_t - \bar{\mathbf{v}})(\mathbf{v}_t - \bar{\mathbf{v}})^T dt,$$

where $\bar{\mathbf{v}} = \int \mathbf{v}_t dt$ is the mean field, and the basis $\mathbf{u}_i, i \in \{1, \cdots, r\}$ corresponds to the i-th eigenvector of $\mathbf{C}$ with associated eigenvalue λ_i. In practice, the covariance is empirically computed by a finite set of M observations (snapshots) over the full-scale dynamic system of $\mathbf{v}_t$, i.e., $\mathbf{C} \approx \frac{1}{M}\sum_{t=1}^{M} \mathbf{L}(\mathbf{v}_t - \bar{\mathbf{v}})(\mathbf{v}_t - \bar{\mathbf{v}})^T$.

Due to a key fact that the spectrum of eigenvalues decays incredibly fast (as shown in Fig. 2), we propose to use an optimal set of eigen-functions to form the projected subspace of velocity fields. An explicit way thus to formulate the projection error in Eq (8) is

$$\sum_{i=r+1}^{q} \lambda_i, \text{ with } \lambda_1 > \cdots > \lambda_r \cdots > \lambda_q.$$

This closed-form solution provides an elegant way to measure the projection loss $e = 1 - (\sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{q} \lambda_i)$, where r is typically chosen such as $e < 0.01$ [66, 123].

1.2.2 Reduced-order Regularization via Galerkin Projection As introduced in the previous section, we developed a method to estimate a low-dimensional subspace of velocity fields that uniquely determines the geodesics of diffeomorphisms. We are now ready to construct a reduced-order model of image registration, subject to complex regularizations governed by high-dimensional PDEs (i.e., EPDiff). This procedure is known as *Galerkin projection* and has been widely used to reduce the high computational complexity of PDEs, or ODEs [75, 66, 123].

Consider the EPDiff in Eq (4), we characterize a velocity field $\mathbf{v}_t$ by projecting it onto a finite-dimensional subspace $\mathcal{U}^r$ with much compact basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$. To simplify the notation, we drop the time index t of velocity fields in remaining sections. A discretized formulation of EPDiff equation in terms of matrix multiplication is

$$\begin{aligned}
\frac{\partial \mathbf{v}}{\partial t} &= -\mathbf{K}\left(\operatorname{diag}(\mathbf{L}\mathbf{v})\mathbf{D}^T\mathbf{v} + \operatorname{diag}(\mathbf{v})\mathbf{D}(\mathbf{L}\mathbf{v}) + \operatorname{diag}(\mathbf{L}\mathbf{v})\,\mathbf{div}\,\mathbf{v}\right), \\
&= -\mathbf{K}\sum_{i=1}^{q}\left(\operatorname{diag}(\mathbf{l}_i)v_i\mathbf{D}^T\mathbf{v} + v_i\mathbf{D}\mathbf{L}\mathbf{v} + \operatorname{diag}(\mathbf{l}_i)v_i\,\mathbf{div}\,\mathbf{v}\right), \\
&= -\mathbf{K}\sum_{i=1}^{q}\left(\operatorname{diag}(\mathbf{l}_i)\mathbf{D}^T + \mathbf{D}\mathbf{L} + \operatorname{diag}(\mathbf{l}_i)\,\mathbf{div}\right)v_i\mathbf{v}, \qquad (9)
\end{aligned}$$

where $\mathbf{v}$ is a q-dimensional vector, and $\mathrm{diag}(\cdot)$ converts a vector to a diagonal matrix. The matrices $\mathbf{L}^{q\times q}$, $\mathbf{K}^{q\times q}$, $\mathbf{D}^{q\times q}$, and $\mathbf{div}^{q\times q}$ denote discrete analogs of the differential operator L with its inverse K, Jacobian matrix D, and divergence div obtained by finite difference schemes respectively. Here, $\mathbf{l}_i$ is the i-th column of the matrix $\mathbf{L}$ and v_i is the i-th element of vector $\mathbf{v}$.

By defining a composite operator $\mathbf{A}_i^{q\times q} \triangleq \mathbf{K}(\mathrm{diag}(\mathbf{l}_i)\mathbf{D}^T + \mathbf{DL} + \mathrm{diag}(\mathbf{l}_i)\,\mathbf{div})$, we write Eq (9) as

$$\frac{\partial \mathbf{v}}{\partial t} = \sum_{i=1}^{q} \mathbf{A}_i v_i \mathbf{v}. \tag{10}$$

Next, we derive a reduced-order model of EPDiff via Galerkin projection by plugging $\mathbf{v} = \mathbf{U}^{q\times r}\boldsymbol{\alpha}$ into Eq (10). We then have

$$\begin{aligned}
\frac{\partial \mathbf{U}\boldsymbol{\alpha}}{\partial t} &= \sum_{i=1}^{q} \mathbf{A}_i(\mathbf{U}\boldsymbol{\alpha})_i \mathbf{U}\boldsymbol{\alpha}, \\
\frac{\partial \boldsymbol{\alpha}}{\partial t} &= \mathbf{U}^T \sum_{i=1}^{q} \mathbf{A}_i (\sum_{j=1}^{r} \mathrm{U}_{ij}\alpha_j)\mathbf{U}\boldsymbol{\alpha} = \sum_{j=1}^{r}\sum_{i=1}^{q} \mathbf{U}^T\mathbf{A}_i\mathbf{U}\mathrm{U}_{ij}\alpha_j\boldsymbol{\alpha} \triangleq \sum_{j=1}^{r} \widetilde{\mathbf{A}}_j\alpha_j\boldsymbol{\alpha},
\end{aligned} \tag{11}$$

where U_{ij} the element of $\mathbf{U}$ in the i-th row and j-th column. Here, we define $\widetilde{\mathbf{A}}_j^{r\times r} = \sum_{i=1}^{q} \mathbf{U}^T\mathbf{A}_i\mathbf{U}\mathrm{U}_{ij}$ as a reduced-order model operator of A_j. It is worthy to mention that the computation of $\widetilde{\mathbf{A}}_i$ is a one-time cost accomplished offline. No further update is needed once a proper subspace is sought. Solution to this reduced-order model can be found by employing commonly used temporal differential schemes, e.g., Euler or Runge-Kutta Method, with an initial condition $\boldsymbol{\alpha}_0 = \mathbf{U}^T\mathbf{v}_0$.

1.3 ROM for Diffeomorphic Image Registration

In this section, we present a reduced-order model of LDDMM algorithm with geodesic shooting for diffeomorphic image registration. We run gradient descent on a projected initial velocity, represented by the loading coefficient $\boldsymbol{\alpha}_0$, entirely in a low-dimensional subspace. A geodesic path consequently generates a flow of diffeomorphisms by Eq (3) after constructing the time-dependent velocity fields back in its original space using $\mathbf{v}_t = \mathbf{U}\boldsymbol{\alpha}_t$.

The redefined energy function of LDDMM in Eq (3) with sum-of-squared dissimilarity between images is

$$\mathbb{E}(\boldsymbol{\alpha}_0) = \frac{1}{2\sigma^2}\|S\circ\psi_1 - T\|_2^2 + (\mathbf{L}\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_0),\ \ s.t.\ \ \text{Eq (11)}. \tag{12}$$

Here, we adopt a commonly used Laplacian operator $\mathbf{L} = (-\beta\Delta + \mathbf{I})^c$, where β is a positive weight parameter, c controls the level of smoothness, and $\mathbf{I}$ is an identity matrix.

Analogous to solving the optimal control problems in [203], we compute the gradient term by using a forward-backward sweep scheme. Below are the general steps for gradient computation (please refer to Alg. 1 for more details):

(i) Compute the gradient $\nabla_{\boldsymbol{\alpha}_1}\mathbb{E}$ of the energy Eq (12) at $t = 1$ by integrating both the diffeomorphism ψ_t and the projected velocity field α_t forward in time, i.e.,

$$\nabla_{\boldsymbol{\alpha}_1}\mathbb{E} = \mathbf{K}\left(\frac{1}{\sigma^2}(S \circ \psi_1 - T) \cdot \nabla(S \circ \psi_1)\right). \tag{13}$$

(ii) Bring the gradient $\nabla_{\boldsymbol{\alpha}_1}\mathbb{E}$ back to $t = 0$. We obtain $\nabla_{\boldsymbol{\alpha}_0}\mathbb{E}$ by integrating reduced adjoint Jacobi field equations [27] backward in time as

$$\frac{d\widehat{\alpha}}{dt} = -\mathrm{ad}^{\dagger}_{\alpha}\widehat{h}, \quad \frac{d\widehat{h}}{dt} = -\widehat{\alpha} - \mathrm{ad}_{\alpha}\widehat{h} + \mathrm{ad}^{\dagger}_{\widehat{h}}\alpha, \tag{14}$$

where $\mathrm{ad}^{\dagger}$ is an adjoint operator and $\widehat{h}, \widehat{\alpha} \in V$ are introduced adjoint variables with an initial condition $\widehat{h} = 0, \widehat{\alpha} = \nabla_{\boldsymbol{\alpha}_1}\mathbb{E}$ at $t = 1$.

Algorithm 1 Optimization of Reduced order model for diffeomorphic image registration

Input: source image S, target image T

`/* Online optimization */`

Initialize $\boldsymbol{\alpha}_0 = 0$ **repeat**

`/* Forward shooting of` $\boldsymbol{\alpha}_0$ `*/`

forward integrate the reduced-order model of EPDiff equation Eq (11) to generate $\{\boldsymbol{\alpha}_t\}$ at discrete time points

`/* Compute the diffeomorphic transformations` ψ_t `*/`

integrate the transformation fields ψ by using Eq (3) after constructing velocity fields back to the original space via $\mathbf{v}_t = \mathbf{U}^T\boldsymbol{\alpha}_t$ `/* Compute gradient at time point` $t = 1$ `*/`

compute the gradient $\nabla_{\boldsymbol{\alpha}_1}E$ by Eq (13) `/* Propagate gradient back to time point` $t = 0$ `*/`

integrate the reduced adjoint Jacobi field equations Eq (14) backward in time to obtain $\nabla_{\boldsymbol{\alpha}_0}E$.

Update $\boldsymbol{\alpha}_0 \leftarrow \boldsymbol{\alpha}_0 - \varepsilon\nabla_{\boldsymbol{\alpha}_0}E$, where ε is the step size

until *convergence*;

1.4 Experimental Evaluation

To demonstrate the effectiveness of the proposed model, we compare its performance with the state-of-the-art vector momentum LDDMM [162] in applications of pairwise

image registration and atlas building. For fair comparison, we use $\beta = 3, c = 6$ for the L operator, and $\sigma = 0.01$ with 10 time-steps for Euler integration across all baseline algorithms.

Data. We applied the algorithm to 3D brain MRI scans from a public released resource Open Access Series of Imaging Studies (OASIS) for Alzheimer's disease [56]. The data includes fifty healthy subjects as well as disease, aged 60 to 96. To better evaluate the estimated transformations, we employed another fifty 3D brain MRI scans with manual segmentations from Alzheimer's Disease Neuroimaging Initiative(ADNI) [80]. All MRIs are of dimension $128 \times 128 \times 128$ with the voxel size of $1.25mm^3$. The images underwent downsampling, skull-stripping, intensity normalization, bias field correction and co-registration with affine transformation.

Experiments. We first tested our algorithm for pairwise image registration at different levels of projected dimension $r = 4^3, 8^3, 12^3, 20^3$ and compared the total energy formulated in Eq (12). In order to find an optimal basis, we ran parallel programs of the full-scale EPDiff equation in Eq (4) and generated a collection of snapshots to perform POD effectively. Since the learning process of basis functions were conducted offline with one-time cost for all experiments, we only focused on the exact runtime, memory consumption, and convergence rate of our model after the fact.

We validated registration results by examining the accuracy of propagated delineations for cortex (Cor), caudate (Caud), and corpus collusum (CC). After aligning all volumes to a reference image, we transformed the manual segmentation from the reference to other volumes by using the estimated deformations. We evaluated dice similarity coefficient (volume overlap) between the propagated segmentation and the manual segmentation for each structure.

We also ran both our method and the baseline algorithm to build an atlas from a set of 3D brain MRIs. We initialized the template image as an average of image intensities, and set the projected dimension as $r = 20^3$ as that was shown to be optimal in our eigen plots. In this experiment, we used a message passing interface (MPI) parallel programming implementation for all methods, and distributed data on four processors in total.

Results. Fig. 3 reports the total energy in formulation Eq (12) averaged over 10 random selected pairs of test images for different values of projected dimensions. Our method arrives at the same solution at $r = 12^3$ and higher, which indicates that the estimated subspace has fairly recovered the result of full-scale registration algorithms. Fig. 3 also provides runtime and memory consumption across all three methods, including the baseline algorithm vector momentum LDDMM. Our algorithm has substantially lower com-

putational cost than vector momentum LDDMM performed in a full-dimensional space.

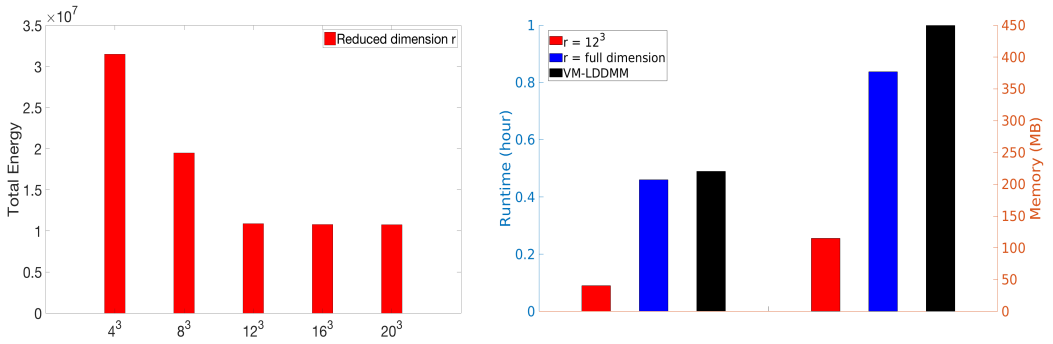


Fig. 3: Left: average total energy for different values of projected dimensions $r = 4^3, 8^3, 12^3, 16^3, 20^3$. Right: exact runtime and memory consumption for all methods.

Fig. 20 reports segmentation volume overlap on different brain structures, estimated from both our method and the baseline algorithm. It show that our algorithm is able to achieve comparable results, while offering significant improvements in computational efficiency. The right panel of Fig. 20 illustrates results for an example case from the study. We observe that the delineations achieved by transferring manual segmentations from the reference frame to the coordinate system of the target frame align fairly well with the manual segmentations. The left panel of Fig. 5 shows the axial and coronal slices from 12 of the

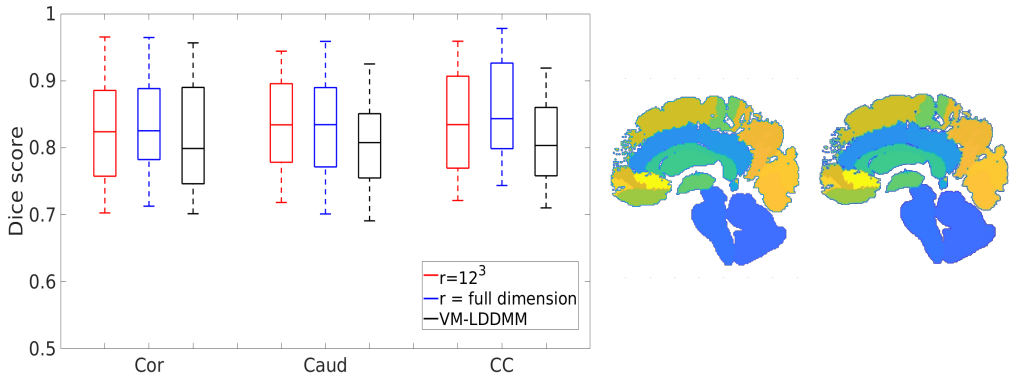


Fig. 4: Left: volume overlap between manual segmentations and propagated segmentations of three important regions cortex (Cor), caudate (Caud), and corpus collusum (CC); Middle: example ground truth segmentation; Right: propagated segmentation with three structures obtained by our method. 2D slices are shown for visualization only, all computations are carried out fully in 3D.

selected 3D MRI dataset. The right panel demonstrates the atlas image estimated by our algorithm, followed by the atlas estimated by vector momenta LDDMM. The difference image between the two atlas results shows that our algorithm generated a very similar atlas to vector momenta LDDMM, but at a fraction of the time and memory cost (as illustrated on the right bottom panel of Fig. 5).

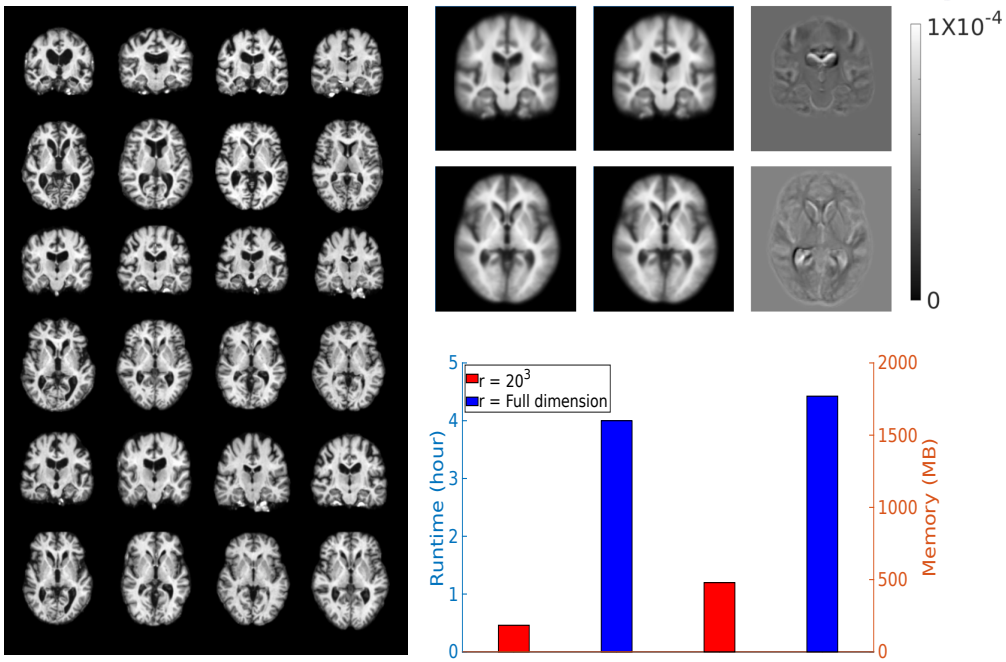


Fig. 5: Left: axial view and coronal view of twelve example brain MRIs selected from dataset. Right top: atlas images estimated by our method and vector momentum LDDMM with difference map shown by side. Right bottom: a comparison of exact runtime and memory consumption.

Fig. 6 shows the eigenvalue spectrum and convergence plot for both image registration (top) and atlas building (bottom). It is clear to see that our method conducted in a low-dimensional space is able to arrive at the same solution as the full dimensional scenario. We outperform the baseline algorithm vector momentum LDDMM, i.e., lower energy at the optimal solution.

1.5 Conclusion

In this section, I presented a data-driven model reduction algorithm for diffeomorphic image registration in the context of LDDMM with geodesic shooting. Our method is

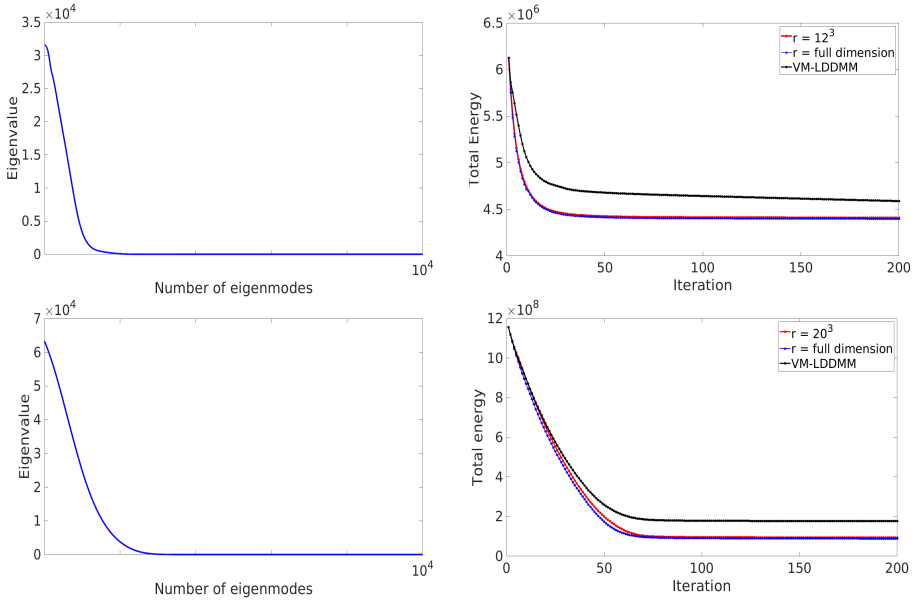


Fig. 6: Top to bottom: results of pairwise image registration vs. atlas building. Left to right: eigenvalue spectrum of velocity fields vs. total energy with an optimal projected dimension, a full dimension of our method, and vector momemtum LDDMM.

the first to simulate the high-dimensional dynamic system of diffeomorphisms in an approximated subspace via proper orthogonal decomposition and Galerkin projection. This approach substantially reduces the computational cost of diffeomorphic registration algorithms governed by high-dimensional PDEs, while preserving comparative accuracy. The theoretical tools developed in this work are broadly applicable to all PDE-constrained diffeomorphic registration models with gradient-based optimization.

2 Low-dimensional Bayesian Registration Uncertainty Quantification

2.1 Related Work

As addressed in Chapter I, diffeomorphic Image registration is a fundamental tool in image analysis as it provides one-to-one smooth and invertible smooth spatial correspondences (also known as diffeomorphisms) between pairwise images. Examples of applications include alignment of functional data to a reference coordinate system [37, 36], anatomical comparisons across individuals [201, 202], and atlas-based image segmenta-

tion [7, 86]. In general, image registration is an ill-posed problem since the image data are usually contaminated by unknown noise. Therefore, developing efficient measures to quantify the registration uncertainty or error is critical to fair assessments on estimated transformations, as well as subsequent improvement on the accuracy of predictive models. This also forms the basis for model-assisted decision making, for example, image-guided navigation system for neurosurgery [142, 106]. An efficient uncertainty quantification can provide important information on brain shifts (deformations) for neurosurgeons to identify residual tumor tissues during surgery; hence leading to a significant increase in the extent of tumor removal while lowering the risk of collateral tissue damage.

Existing methods have been investigated to estimate the uncertainty by having a probability distribution over the latent space of transformation parameters either in a small or large deformation setting [143, 183, 105]. These approaches formulate Bayesian image registration as an image matching likelihood term regularized by a prior that guarantees the smoothness of deformation fields. A posterior distribution is then generated as a measure of the registration uncertainty estimation. Due to the fact that such posterior distributions do not have closed-form formulations, stochastic and sampling methods typically have been employed [94, 96, 142]. This makes the entire inference extremely challenging on a dense image grid (e.g., a 3D brain MRI with the size of 128^3). Large computational resources and efforts are required to sample over a high-dimensional space with extremely slow convergence. Also, none of the aforementioned algorithms is able to compute a full posterior covariance. These disadvantages dramatically decrease the model usability in important applications that require computational efficiency.

In this section, I propose an efficient registration uncertainty quantification model that employs a low-dimensional representation of diffeomorphic transformations [203, 206]. Based on the key fact that the tangent space of diffeomorphisms do not develop high frequencies, we develop a Bayesian registration framework entirely in a bandlimited space with low frequencies well preserved. A posterior distribution of transformation fields is carefully designed in a truncated frequency domain, where all diffeomorphic properties remain valid. Similar to [195], we derive a Laplace approximation of the log-posterior distribution at an optimal solution to further avoid intensive computation cost of sampling methods. More specifically, we assume a complex Gaussian distribution at the mode of the posterior, with its covariance estimated by second-order information to describe uncertainties in the registration model. This is effectively done by computing an inverse Hessian of the log-posterior specifically defined in a low-dimensional bandlimited space. Our model dramatically reduces the computational complexity of approximating posterior marginals in the space of diffeomorphisms, which makes registration uncertainty analysis

tractable in time. In practical, the efficiency of our method strengthens its feasibility in prospective clinical applications under a tight time constraint. This seection includes (i) in-depth derivations of the Hessian computation and inference procedure provided; and (ii) comprehensive experiments to thoroughly validate the method included. We run tests on both 2D synthetic data and real 3D brain MRI scans from OASIS dataset [113]. We then compare estimated results with the state-of-the-art registration uncertainty quantification method in the full image space [195]. Experimental results show that our method is significantly faster than the baseline algorithm, while producing comparable results.

The rest of this section is structured as follows. Sec. 2.2 introduces our developed Bayesian model with Sec. 2.3 presenting the inference. Sec. 2.4 presents a detailed step-by-step derivation for first and second order derivative computation. Sec. 2.5 demonstrates our experiments with detailed discussions.

2.2 Method: Uncertainty Quantification Via Low-dimensional Geometric Deformations

We introduce a low-dimensional Bayesian model of diffeomorphic image registration represented in the bandlimited velocity space $\widetilde{V}$, with registration uncertainty explicitly encoded as latent variables of the model.

Assuming independent and identically distributed (i.i.d.) Gaussian noise on image intensities, we obtain the likelihood

$$p(T \mid S, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma^2)^M} \exp\left(-\frac{1}{2\sigma^2}||S \circ \psi_1 - T||_2^2\right), \tag{15}$$

where σ^2 is the noise variance and M is the number of image voxels. The deformation ψ_1 corresponds to $\widetilde{\psi}_1$ in Fourier space via the Fourier transform $\mathcal{F}(\psi_1) = \widetilde{\psi}_1$, or its inverse $\psi_1 = \mathcal{F}(\widetilde{\psi}_1)$.

We define a prior on the initial velocity field $\widetilde{v}_0$ to be a complex multivariate Gaussian distribution that ensures the smoothness of the geodesic path, i.e.,

$$p(\widetilde{v}_0) = \frac{1}{(2\pi)^{\frac{Md}{2}}|\widetilde{\mathcal{L}}^{-1}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\widetilde{\mathcal{L}}\widetilde{v}_0, \widetilde{v}_0)\right), \tag{16}$$

where $|\cdot|$ is matrix determinant. We define $\widetilde{\mathcal{L}}$ as the Fourier coefficients of a commonly used Laplacian operator $\mathcal{L} = (-\alpha\Delta + I)^c$ with a positive weight parameter α and a

smoothness parameter c. The Fourier transform of the Laplacian operator is given by

$$\widetilde{\mathcal{L}}(\xi) = (-2\alpha \sum_{j=1}^{d} (\cos(2\pi\xi_j) - 1) + 1)^c,$$

with $\widetilde{\mathcal{K}} = 1/\widetilde{\mathcal{L}}$ being an inverse operator of $\widetilde{\mathcal{L}}$.

Combining the likelihood Eq (15) and prior Eq (16) together, we obtain the negative log posterior distribution on the deformation parameter parameterized by $\widetilde{v}_0$ as

$$-\ln\, p(\widetilde{v}_0 \,|\, S, T, \sigma^2) = \frac{1}{2}(\widetilde{\mathcal{L}}\widetilde{v}_0, \widetilde{v}_0) + \frac{\|S \circ \psi_1 - T\|_2^2}{2\sigma^2} + M \ln \sigma + \text{const.} \qquad (17)$$

In most probabilistic formulations of image-based registration, the likelihood Eq (15), as a function of the transformation parameters, is highly non-Gaussian because of the complex spatial structure of the images. This brings difficulties in the inference of such a non-Gaussian posterior. While sampling methods have been investigated to empirically approximate the distribution [207, 160], the computation tends to be extremely time-consuming and expensive. To address this issue, we next introduce Laplace's method [186] to approximate the posterior distribution in a much more efficient manner.

2.2.1 Laplace approximation The basic idea is to find the mode of the posterior as a MAP solution, and then apply a second-order Taylor series approximation for the log-posterior function that corresponds to a complex Gaussian distribution. We first minimize the negative log posterior in Eq (17) to the optimum $\widetilde{v}_0^{opt}$, which is considered to be the mean of the Gaussian distribution (details of inference will be introduced in Sec. 2.3). Estimation of covariance is not straightforward, as we will need to derive a quadratic function of log-posterior by second-order approximation.

To simplify the notation, we use $f(\widetilde{v}_0) \triangleq -\ln\, p(\widetilde{v}_0 \,|\, S, T, \sigma^2)$ to represent the log-posterior. The function $f(\widetilde{v}_0)$ is approximated to quadratic order by using second order Taylor series expansion at the optimal solution $\widetilde{v}_0$ as

$$\begin{aligned} f(\widetilde{v}_0) &= f(\widetilde{v}_0^{opt} + (\widetilde{v}_0 - \widetilde{v}_0^{opt})) \\ &\approx f(\widetilde{v}_0^{opt}) + \nabla f^T(\widetilde{v}_0^{opt})(\widetilde{v}_0 - \widetilde{v}_0^{opt}) + \frac{1}{2}(\widetilde{v}_0 - \widetilde{v}_0^{opt})^T \mathcal{H} f(\widetilde{v}_0^{opt})(\widetilde{v}_0 - \widetilde{v}_0^{opt}), \end{aligned}$$

where ∇ denotes the first derivative and $\mathcal{H}$ is a second-order Hessian. Since the first derivative of f vanishes at the optimal solution $\widetilde{v}_0^{opt}$, we have

$$f(\widetilde{v}_0) \approx f(\widetilde{v}_0^{opt}) + \frac{1}{2}(\widetilde{v}_0 - \widetilde{v}_0^{opt})^T \mathcal{H} f(\widetilde{v}_0^{opt})(\widetilde{v}_0 - \widetilde{v}_0^{opt}). \qquad (18)$$

This indicates that the posterior is approximately a multivariate Gaussian distribution $\mathcal{N}(\widetilde{v}_0^{opt}, \mathcal{H}^{-1} f(\widetilde{v}_0^{opt}))$. The optimal solution $\widetilde{v}^{opt}$ and the inverse Hessian corresponds to the mean velocity and covariance matrix of the registration parameters respectively.

2.3 Inference

The algorithmic inference includes two major components: seeking for mean and covariance of the approximated posterior distribution. For mean estimation, we develop a gradient descent algorithm to minimize the negative log posterior distribution Eq (17) w.r.t. the initial velocity $\widetilde{v}_0$ and the image noise variance σ^2. For covariance estimation, we derive a second variation of Eq (17) to compute the Hessian-vector product via a linearized forward-backward sweep. All computational steps will be introduced in the following sections.

2.3.1 Mean Estimation Following optimal control theory [173], we add Lagrange multipliers to constrain the diffeomorphism $\widetilde{\psi}_t$ to be a geodesic path in the frequency domain. This is done by introducing time-dependent adjoint variables, $\widehat{v}_t$ and $\widehat{u}_t$, and writing the augmented energy [2],

$$E(\widetilde{v}_0) = -\ln\, p(\widetilde{v}_0 \,|\, S, T, \sigma^2) + \int_0^1 \langle \widehat{v}_t, \dot{\widetilde{v}}_t + \mathrm{ad}^{\dagger}_{\widetilde{v}_t} \widetilde{v}_t \rangle + \langle \widehat{u}_t, \dot{u}_t + \widetilde{v}_t + \widetilde{\mathcal{D}}\widetilde{u}_t * \widetilde{v}_t \rangle \, dt, \quad (19)$$

where the last two terms correspond to Lagrange multipliers enforcing the geodesic constraint Eq (5) and the deformation transport equation Eq (3).

The optimality conditions for the adjoints $\widehat{v}_t, \widehat{u}_t$ are given by the following time-dependent system of ordinary differential equations, termed the adjoint equations (equivalent to error-back propagation):

$$\begin{aligned} -\dot{\widehat{v}}_t + \mathrm{ad}_{\widetilde{v}_t} \widehat{v}_t - \mathrm{ad}^{\dagger}_{\widehat{v}_t} \widetilde{v}_t + \widehat{u}_t + (\widetilde{\mathcal{D}}\widetilde{u}_t)^T \star \widehat{u}_t = 0, \\ -\dot{\widehat{u}}_t - \widetilde{\mathcal{D}}\delta\widehat{u}_t * \widetilde{v}_t - \delta\widehat{u}_t * (\widetilde{\nabla} \cdot \widetilde{v}_t) = 0, \end{aligned} \quad (20)$$

subject to initial conditions $\widehat{v}_1 = 0$ and $\widehat{u}_1 = -\frac{1}{\sigma^2}\mathcal{F}[\langle \nabla S(1), S(1) - T \rangle]$, where $S(1) = S \circ \psi_1$. Please refer to Appendix A for more details.

After integrating the geodesic equations (state equations) Eq (3)& Eq (5) forward in time to $t = 1$ and then backward integrating the adjoint equations Eq (20) in time to $t = 0$, the gradient of E w.r.t. $\widetilde{v}_0$ is $\nabla_{\widetilde{v}_0} E = \widetilde{v}_0 - \widehat{v}_0$.

Setting the gradient w.r.t. σ^2 to zero, we have a closed form update

$$\sigma^2 = \frac{1}{M} \| S(1) \circ \psi_1 - T \|_2^2.$$

[2] For notation simplification, we define the time derivative $\dot{\widetilde{v}}_t \triangleq d\widetilde{v}_t/dt$.

2.3.2 Covariance Estimation To estimate the full covariance matrix as an inverse Hessian, we develop a similar forward-backward approach to compute the Hessian-vector products that involves the second variation of the augmented energy function Eq (19). More specifically, we read off the Hessian-vector product after deriving the second variation in the direction $\delta\widetilde{v}_0$ as

$$\frac{\partial^2}{\partial\epsilon^2}E(\widetilde{v}_0+\epsilon\cdot\delta\widetilde{v}_0)|_{\epsilon=0}=\langle\delta\widetilde{v}_0,\mathcal{H}E\,\delta\widetilde{v}_0\rangle.$$

Given an initial condition $\delta\widetilde{v}_0$, we compute the Hessian-vector product as

$$\mathcal{H}E\,\delta\widetilde{v}_0=\delta\widetilde{v}_0-\delta\widehat{v}_0,$$

where $\delta\widehat{v}_0$ is the adjoint variable of $\delta\widetilde{v}_0$.

The second variation can be accomplished by forward-sweeping the linearized geodesic constraint around the optimal solution, followed by a backward sweep of the linearized adjoint system. Introducing time-dependent adjoint variables $\delta\widehat{v}_t$ and $\delta\widehat{u}_0$, the forward linearized geodesic equations are

$$\delta\dot{\widetilde{v}}_t=-\mathrm{ad}^\dagger_{\delta\widetilde{v}_t}\widetilde{v}_t-\mathrm{ad}^\dagger_{\widetilde{v}_t}\delta\widetilde{v}_t,\qquad\delta\dot{\widetilde{u}}_t=-\widetilde{\mathcal{D}}\delta\widetilde{u}_t*\widetilde{v}_t-\widetilde{\mathcal{D}}\widetilde{u}_t*\delta\widetilde{v}_t-\delta\widetilde{v}_t. \tag{21}$$

The linearized adjoint system for the backward integration is

$$\begin{aligned}\delta\dot{\widehat{v}}_t&=\mathrm{sym}^\dagger_{\widetilde{v}_t}\delta\widehat{v}_t-\mathrm{sym}^\dagger_{\delta\widehat{v}_t}\widehat{v}_t+\delta\widehat{u}_t+(\widetilde{\mathcal{D}}\widetilde{u}_t)^T\star\delta\widehat{u}_t+(\widetilde{\mathcal{D}}\delta\widetilde{u}_t)^T\star\widehat{u}_t,\\ \delta\dot{\widehat{u}}_t&=-\widetilde{\mathcal{D}}\delta\widehat{u}_t*\widetilde{v}_t-\widetilde{\mathcal{D}}\widehat{u}_t*\delta\widetilde{v}_t-\widetilde{v}_t*\widetilde{\nabla}(\delta\widehat{u}_t)-\widehat{u}_t*\widetilde{\nabla}(\widetilde{v}_t),\end{aligned} \tag{22}$$

subject to initial conditions $\delta\widehat{u}_1=-\frac{2}{\sigma^2}\mathcal{F}[\nabla S(1)\cdot\nabla S(1)+(S(1)-T)\cdot\nabla^2S(1)]$ and $\delta\widehat{v}_1=0$, with $\mathrm{sym}^\dagger_{\widetilde{v}_t}\delta\widehat{v}_t=\mathrm{ad}_{\delta\widetilde{v}_t}\widehat{v}_t-\mathrm{ad}^\dagger_{\widehat{v}_t}\delta\widetilde{v}_t$.

We summarize the inference of our algorithm in Alg. 2. Following sections include the detailed computational derivations for gradient term and second-order Hessian approximation.

2.4 Derivations of gradient and Hessian

As we derived in Sec. 2.2, the augmented energy with the adjoint variables $\widehat{v}_t,\widehat{u}_t$ can be formed as follows after adding geodesic constraints as Lagrangian multipliers,

$$\begin{aligned}\mathbb{E}(\widetilde{v}_0)=&-\ln\,p(\widetilde{v}_0\,|\,S,T,\sigma^2)+\underbrace{\int_0^1\langle\widehat{v}_t,\dot{\widetilde{v}}_t+\mathrm{ad}^\dagger_{\widetilde{v}_t}\widetilde{v}_t\rangle}_{1}\\ &+\underbrace{\langle\widehat{u}_t,\dot{\widetilde{u}}_t+\widetilde{v}_t+\widetilde{\mathcal{D}}\widetilde{u}_t*\widetilde{v}_t\rangle\,dt}_{2}.\end{aligned} \tag{23}$$

Algorithm 2 Model inference.

Input : source image S and target image T, step size ϵ, iterations r

Output: optimal registration $\widetilde{v}_0^{opt}$, and covariance $\mathcal{H}^{-1} f(\widetilde{v}_0^{opt})$

/* Mean estimation */

for $i = 1$ to r **do**

1. Forward integrating the geodesic evolution equations Eq (5) with initial velocity $\widetilde{v}_0$, then generates a collection of time-dependent velocity fields $\widetilde{v}_t$;

2. Backward integrating adjoint equations Eq (20) to get gradient $\nabla_{\widetilde{v}_0 E}$;

3. $\widetilde{v}_0^{opt} \leftarrow \widetilde{v}_0^{opt} - \epsilon \nabla_{\widetilde{v}_0 E}$

/* Covariance estimation */

for $i = 1$ to M **do**

1. Forward shooting of $\delta\widetilde{v}_0$: forward-sweeping the linearized geodesic constraints Eq (21) to generate $\delta\widetilde{v}_t$;

2. Backward shooting of $\delta\widetilde{v}_1$: backward integrating adjoint equation Eq (22) to generate $\widetilde{\mathcal{H}} E \delta\widetilde{v}_0$;

3. Invert Hessian $\widetilde{\mathcal{H}}^{-1} E \delta\widetilde{v}_0 = \widetilde{\mathcal{H}}^{-1} f(\widetilde{v}_0^{opt})$ to generate covariance.

Next, we derive the variation of $\widetilde{E}$ w.r.t the variables $\widetilde{v}_0$ and $\widetilde{u}_0$ for $1, 2$ respectively. To simplify the notation, we drop the subscript of time index t in the following sections.

Each part of the first derivative of $\partial_{\widetilde{v}} E$ is,

$$
\begin{aligned}
1: \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \int_0^1 \langle \widehat{v}, \partial_t(\widetilde{v} + \epsilon\delta\widetilde{v}) + \mathrm{ad}^{\dagger}_{\widetilde{v}+\epsilon\delta\widetilde{v}}(\widetilde{v} + \epsilon\delta\widetilde{v})\rangle \\
2: \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \bigg(\int_0^1 \langle \widehat{u}, \widetilde{v} + \epsilon\delta\widetilde{v} + \widetilde{\mathcal{D}}\widetilde{u} * (\widetilde{v} + \epsilon\delta\widetilde{v})\rangle \bigg)
\end{aligned}
$$

We combine these two terms,

$$
\begin{aligned}
1,2: &\int_0^1 \langle \widehat{v}, \delta\dot{\widetilde{v}} + \mathrm{ad}^{\dagger}_{\widetilde{v}}\delta\widetilde{v} + \mathrm{ad}^{\dagger}_{\delta\widetilde{v}}\widetilde{v}\rangle + \int_0^1 \langle \widehat{u}, \delta\widetilde{v} + \widetilde{\mathcal{D}}u * \delta\widetilde{v}\rangle \\
&= \langle \widehat{v}, \delta\widetilde{v}\rangle\Big|_{t=0}^{t=1} - \int_0^1 \langle \dot{\widehat{v}}, \delta\widetilde{v}\rangle + \int_0^1 \langle \mathrm{ad}_{\widetilde{v}}\widehat{v} - \mathrm{ad}^{\dagger}_{\widehat{v}}\widetilde{v}, \delta\widetilde{v}\rangle \\
&\quad + \int_0^1 \langle \widehat{u} + (\widetilde{\mathcal{D}}\widetilde{u})^T \star \widehat{u}, \delta\widetilde{v}\rangle \\
&= \langle \widehat{v}, \delta\widetilde{v}\rangle\Big|_{t=0}^{t=1} + \int_0^1 \langle -\dot{\widehat{v}} + \mathrm{ad}_{\widetilde{v}}\widehat{v} - \mathrm{ad}^{\dagger}_{\widehat{v}}\widetilde{v} + \widehat{u} + (\widetilde{\mathcal{D}}\widetilde{u})^T \star \widehat{u}, \delta\widetilde{v}\rangle. \qquad (24)
\end{aligned}
$$

Each part of the first derivative $\partial_{\widetilde{u}} E$ is

$$\begin{aligned}
1,2: \partial_{\widetilde{u}} E &= \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \left(\int_0^1 \langle \widehat{u}, \partial_t(\widetilde{u} + \epsilon\delta\widetilde{u}) + \widetilde{\mathcal{D}}(\widetilde{u} + \epsilon\delta\widetilde{u}) * \widetilde{v} \rangle \right) \\
&= \int_0^1 \langle \widehat{u}, \delta\dot{u} + \widetilde{\mathcal{D}}\delta\widetilde{u} * \widetilde{u} \rangle \\
&= \langle \widehat{u}, \delta\widetilde{u} \rangle \Big|_{t=0}^{t=1} - \int_0^1 \langle \dot{\widehat{u}}, \delta\widetilde{u} \rangle - \int_0^1 \langle \delta\widetilde{u}, \widetilde{\mathcal{D}}\widetilde{u} * \widetilde{v} + \widetilde{u} * \widetilde{\nabla} \cdot \widetilde{v} \rangle \\
&= \langle \widehat{u}, \delta\widetilde{u} \rangle \Big|_{t=0}^{t=1} - \int_0^1 \langle \delta\widetilde{u}, \dot{\widehat{u}} + \widetilde{\mathcal{D}}\widetilde{u} * \widetilde{v} + \widetilde{u} * \widetilde{\nabla} \cdot \widetilde{v}) \rangle. \qquad (25)
\end{aligned}$$

By setting Eq (25) to zero, we have $\dot{\widehat{u}}_t = -\widetilde{\mathcal{D}}\delta\widehat{u}_t * \widetilde{v}_t - \delta\widehat{u}_t * (\widetilde{\nabla} \cdot \widetilde{v}_t)$. Putting Eq (24) and Eq (25) together, we obtain the adjoint equations of Eq (20),subject to initial conditions

$$\begin{aligned}
\widehat{v}_1 &= 0, \\
\widehat{u}_1 &= -\frac{1}{\sigma^2} \langle \nabla S_1, S_1 - T \rangle.
\end{aligned}$$

We develop a similar forward-backward approach to compute the second derivatives of $\widetilde{v}$ and $\widetilde{u}$. We first derive the second-order forward shooting equation of by adding auxiliary variables $\delta\widehat{v}$ and $\delta\widehat{u}$ on the geodesic shooting EpDiff Eq (5) and deformation transport equation Eq (3) respectively.
For $\delta\widetilde{v}$,

$$\begin{aligned}
\partial^2_{\delta\widetilde{v}} E &= \frac{\partial^2}{\partial \epsilon}\Big|_{\epsilon=0} \int_0^1 \langle \delta\widehat{v}, \partial_t(\delta\widetilde{v} + \epsilon\delta\widehat{v}) + \mathrm{ad}^\dagger_{\widetilde{v}+\epsilon\delta\widehat{v}}(\widetilde{v} + \epsilon\delta\widehat{v}) \rangle \\
&= \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \int_0^1 \langle \delta\widehat{v}, \delta\dot{\widetilde{v}} + \mathrm{ad}^\dagger_{\widetilde{v}+\epsilon\delta\widehat{v}}\delta\widetilde{v} + \mathrm{ad}^\dagger_{\delta\widetilde{v}}(\widetilde{v} + \epsilon\delta\widehat{v}) \rangle \\
&+ \int_0^1 \langle \delta\widehat{v}, \partial_t(\delta\widehat{v} + \epsilon\delta\widehat{v}) + \mathrm{ad}^\dagger_{\widetilde{v}+\epsilon\delta\widehat{v}}(\widetilde{v} + \epsilon\delta\widehat{v}) \rangle \\
&= \int_0^1 \langle \delta\widehat{v}, \delta\dot{\widetilde{v}} + \mathrm{ad}^\dagger_{\widetilde{v}}\delta\widetilde{v} + \mathrm{ad}^\dagger_{\delta\widetilde{v}}\widetilde{v} \rangle + \langle \delta\widetilde{v}, \dot{\widetilde{v}} + \mathrm{ad}^\dagger_{\widetilde{v}}\widetilde{v} \rangle. \qquad (26)
\end{aligned}$$

We then derive the forward shooting equation for $\delta\widetilde{u}$ as follows,

$$\begin{aligned}
\partial_{\delta\widetilde{u}} E &= \frac{\partial^2}{\partial \epsilon}\Big|_{\epsilon=0} \int_0^1 \langle \delta\widehat{u}, \partial_t(\widetilde{u} + \epsilon\delta\widehat{u}) + \widetilde{v} + \epsilon\delta\widehat{v} + \widetilde{\mathcal{D}}(\widetilde{u} + \epsilon\delta\widehat{u}) * (\widetilde{v} + \epsilon\delta\widehat{v}) \rangle \\
&= \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \int_0^1 \langle \delta\widehat{u}, \delta\dot{\widetilde{u}} + \delta\widetilde{v} + \widetilde{\mathcal{D}}\delta\widetilde{u} * \widetilde{v} + \widetilde{\mathcal{D}}(\widetilde{u} + \epsilon\delta\widehat{u}) * \delta\widetilde{v} \rangle \\
&\quad + \langle \delta\widehat{u}, \partial_t(\widetilde{u} + \epsilon\delta\widehat{u}) + \widetilde{v} + \epsilon\delta\widehat{v} + \widetilde{\mathcal{D}}(\widetilde{u} + \epsilon\delta\widehat{u}) * (\widetilde{v} + \epsilon\delta\widehat{v}) \rangle \\
&= \int_0^1 \langle \delta\widehat{u}, \delta\dot{\widetilde{u}} + \delta\widetilde{v} + \widetilde{\mathcal{D}}\delta\widetilde{u} * \widetilde{v} + \widetilde{\mathcal{D}}\widetilde{u} * \delta\widetilde{v} \rangle + \int_0^1 \langle \delta\widehat{u}, \dot{\widetilde{u}} + \widetilde{v} + \widetilde{\mathcal{D}}\widetilde{u} * \widetilde{v} \rangle. \qquad (27)
\end{aligned}$$

By setting Eq (26) and Eq (27) to zero, we obtain the second-order forward shooting for $\delta\widetilde{v}$ and $\delta\widetilde{u}$ as stated in Eq (21).

Next we construct the augmented energy by adding second-order forward shooting equations as Lagrangian multipliers.

$$\begin{aligned}\mathbb{E}(\delta\widehat{v}_0) = &-\ln\, p(\widehat{v}_0 \,|\, S, T, \sigma^2) \\ &+ \underbrace{\int_0^1 \langle \delta\widehat{v}, \delta\dot{\widetilde{v}} + \mathrm{ad}^\dagger_{\delta\widetilde{v}}\widetilde{v} + \mathrm{ad}^\dagger_{\widetilde{v}}\delta\widetilde{v}\rangle}_{1} \\ &+ \underbrace{\langle \delta\widehat{u}, \delta\dot{\widetilde{u}} + \widetilde{\mathcal{D}}\delta\widetilde{u} * \widetilde{v} - \widetilde{\mathcal{D}}\widetilde{u} * \delta\widetilde{v} - \delta\widetilde{v}\rangle}_{2}.\end{aligned}$$

Similarly, we derive the second-order backward shooting equation by adding auxiliary variables $\delta\widehat{v}$ and $\delta\widehat{u}$ on part 1 and 2.
For $\delta\widetilde{v}$,

$$\begin{aligned}\partial^2_{\delta\widetilde{v}}E &= \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\int_0^1 \langle \delta\widehat{v}, \partial_t(\delta\widetilde{v} + \epsilon\delta\widehat{v}) + \mathrm{ad}_{\delta\widetilde{v}+\epsilon\delta\widehat{v}}\widetilde{v} + \mathrm{ad}_{\widetilde{v}}(\delta\widetilde{v} + \epsilon\delta\widehat{v})\rangle \\ &\quad + \langle \delta\widehat{u}, \partial(\delta\widetilde{v} + \epsilon\delta\widehat{v}) + \widetilde{\mathcal{D}}\delta\widetilde{u} * \widetilde{v} + \widetilde{\mathcal{D}}\widetilde{u} * (\delta\widetilde{v} + \epsilon\widehat{v}) + (\delta\widetilde{v} + \epsilon\delta\widehat{v})\rangle \\ &= \langle \delta\widehat{v}, \delta\widetilde{v}\rangle\Big|_{t=0}^{t=1} + \int_0^1 \langle \delta\widehat{v}, -\delta\dot{\widetilde{v}} + \mathrm{ad}_{\widetilde{v}}\delta\widetilde{v} - \mathrm{ad}^\dagger_{\delta\widetilde{v}}\widetilde{v} + \delta\widetilde{u} + (\widetilde{\mathcal{D}}\widetilde{u})^T \star \delta\widetilde{u}\rangle \\ &\quad + \int_0^1 \langle \delta\widehat{v}, -\delta\dot{\widetilde{v}} + \mathrm{ad}_{\delta\widetilde{v}}\widetilde{v} - \mathrm{ad}^\dagger_{\widetilde{v}}\delta\widetilde{v} + (\widetilde{\mathcal{D}}\delta\widetilde{u})^T \star \widetilde{u}\rangle. \end{aligned} \tag{28}$$

For $\delta\widetilde{u}$,

$$\begin{aligned}\partial^2_{\delta\widetilde{u}}E &= \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\int_0^1 \langle \delta\widehat{u}, \partial_t(\delta\widetilde{u} + \epsilon\delta\widehat{u}) + \mathrm{ad}_{\delta\widetilde{v}}\widetilde{v} + \mathrm{ad}_{\widetilde{v}}(\delta\widetilde{v})\rangle \\ &\quad + \langle \delta\widehat{u}, \partial_t(\delta\widetilde{u} + \epsilon\delta\widehat{u}) + \widetilde{\mathcal{D}}(\delta\widetilde{u} + \epsilon\delta\widehat{u}) * \widetilde{v} + \widetilde{\mathcal{D}}\widetilde{u} * \delta\widetilde{v} - \delta\widetilde{v}\rangle \\ &= \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\int_0^1 \langle \delta\widehat{u}, \partial_t(\delta\widetilde{u} + \epsilon\delta\widehat{u}) + \widetilde{\mathcal{D}}(\widetilde{u} + \epsilon\delta\widehat{u}) * \widetilde{v}\rangle \\ &\quad + \langle \delta\widehat{u}, \partial_t(\delta\widetilde{u} + \epsilon\delta\widehat{u}) + \widetilde{\mathcal{D}}(\widetilde{u} + \epsilon\delta\widehat{u}) * \delta\widetilde{v}\rangle \\ &= \langle \delta\widehat{u}, \delta\widetilde{u}\rangle\Big|_{t=0}^{t=1} + \int_0^1 \langle -\delta\widehat{u}, \delta\dot{\widetilde{u}} + \widetilde{\mathcal{D}}\delta\widetilde{u} * \widetilde{v} + \delta\widetilde{v} * \widetilde{\nabla}\cdot\widetilde{v}\rangle \\ &\quad + \int_0^1 \langle -\delta\widehat{u}, \delta\dot{\widetilde{u}} + \widetilde{\mathcal{D}}\widetilde{u} * \delta\widetilde{v} + \widetilde{u} * \widetilde{\nabla}\cdot\delta\widetilde{v}\rangle. \end{aligned} \tag{29}$$

By setting Eq (28) and Eq (29) to 0, we arrive the second-order backward equations of $\delta\widetilde{v}$ and $\delta\widetilde{u}$ as Eq (22), subject to initial conditions $\delta\widehat{u}_1 = -\frac{2}{\sigma^2}\mathcal{F}[\nabla S(1)\cdot\nabla S(1) + (S(1) - T)\cdot\nabla^2 S(1)]$ and $\delta\widehat{v}_1 = 0$.

2.5 Experimental Evaluation

To show the effectiveness of our model, we run experiments on both 2D synthetic data and 3D real brain MRI scans. We estimate the covariance by using $\alpha = 3, c = 6$ for the operator $\widetilde{L}$. We set band-limited dimension of the initial velocity field as 16, which is similar to the settings used in the pairwise diffeomorphic image registration [206]. We run both our model and baseline algorithms till convergence, with the number of time steps for Euler integration in geodesic shooting being 10. **Data** We simulate a collection of 2D synthetic dataset starting from a binary circle with resolution 100×100. We then generate smooth initial velocity fields from the prior distribution $p(\widetilde{v}_0)$, defined in Eq (49). The deformed binary circles (used as our target images) are constructed by deforming the binary circle image with forward shooting the initial velocities.

The brain MRI scans of 60 subjects from the OASIS dataset [113], aged 60 to 90. All MRIs have the same resolution $128 \times 128 \times 128$ with the voxel size of $1.25 \times 1.25 \times 1.25mm^3$. To further evaluate the accuracy of our mean estimation (registration solution), we use 10 pairs of 3D brain MRI scans with manual segmentation labels from public released ADNI dataset [80]. All images underwent the preprocessing of skull-stripping, downsampling, intensity normalization to $[0, 1]$ interval, bias field correction, and co-registration with affine transformations

Experiments We first evaluate the proposed approach on 2D synthetic data and compare with the state-of-the-art diffeomorphic registration uncertainty quantification algorithm defined in a full-dimensional image space [195]. This baseline algorithm is defined in the original high-dimensional image space, where the full hessian is approximated via low-rank basis obtained by principal component analysis. The authors then compute the covariance matrix by inverting the approximated low-rank Hessian.

A visualization of uncertainty map estimated by the covariance matrix will be demonstrated. We set the initialization of initial velocity fields as zero and $\sigma = 0.01$ for both methods. We then validate our model on real 3D brain MRIs. Note that it is difficult to measure the accuracy of uncertainty methods since there is no registration ground truth exists. In order to demonstrate the effectiveness of our method in a statistical way, we report the difference between the Hessian without low-dimensional approximation and the Hessian estimated by our method both on 2D synthetic and 3D real data. In addition to estimate the final uncertainty maps, we show the time and memory comparison of all algorithms.

Finally, we demonstrate that another benefit of our model is the improved performance of mean estimation by standard image registration problem. Rather than finding an optimal registration result (mean) in a high-dimensional image space by using gradient descent algorithm, we instead perform an entire optimization in a low-dimensional bandlimited space. This not only makes the algorithmic inference much faster, but also reduces the risk of getting stuck in local minima. We compare our algorithm with the state-of-the-art fast diffeomorphic registration schemes: Quicksilver [194] and Voxelmorph [14]. To evaluate the estimated deformations, we perform registration-based segmentation and examine the resulting segmentation accuracy. The volume overlap, also known as Dice Similarity Coefficient (DSC), between the propagated segmentation A and the manual segmentation B for each structure is computed as $(A, B) = 2(|A| \cap |B|)/(|A| + |B|)$ where $\cap$ denotes an intersection of two regions.

Results Fig. 7 visualizes the uncertainty information estimated from both our method and the baseline algorithm performed in high-dimensional space on 2D data. We extract the local covariance matrix of each voxel and visualize it as an ellipse on the source image, with the color representing the matrix determinant. The smaller determinants are closer to the non-isotropic area (e.g., circle boundaries), which indicate more confident registration results.

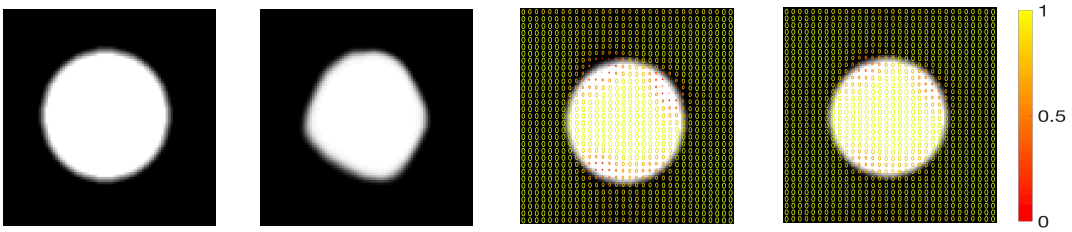


Fig. 7: Left to right: source image, target image, covariance matrix determinant estimated by the baseline algorithm and our method.

Fig. 8 visualizes an example of 3D brain registration uncertainty. Note that due to the difficulty of computing a full covariance matrix by inverse Hessian in a high-dimensional image space, we need to use an approximated low-rank Hessian with a number of dominant eigenmodes [195]. Based on our key observation that the eigenvalues of Hessian matrix decay fast (as shown on the left panel of Fig. 9), we calculated the variances with various numbers of eigenmodes and noticed that the difference between covariances goes to zero beyond 2600 eigenmodes. Following [195], we then choose the first 2600 eigenmodes to approximate the full covariance accordingly. Both two methods show that the

high uncertainty (with less confidence) appear isotropic areas (e.g., inside the ventricle), while the low uncertainty (with high confidence) appears around non-isotropic areas (e.g. ventricle boundaries). The right panel of Fig. 9 reports the comparison of time and memory consumption. Our algorithm offers significant improvements in computational efficiency. It is worth mentioning that while we adopted FLASH [206] for fast registration, the proposed Bayesian framework can be generalized to other registration models with the same level of efficiency, e.g., stationary velocity fields that remain constant over time [6], or projected subspace of velocity fields that characterizes deformations in a much lower dimensional space [179].

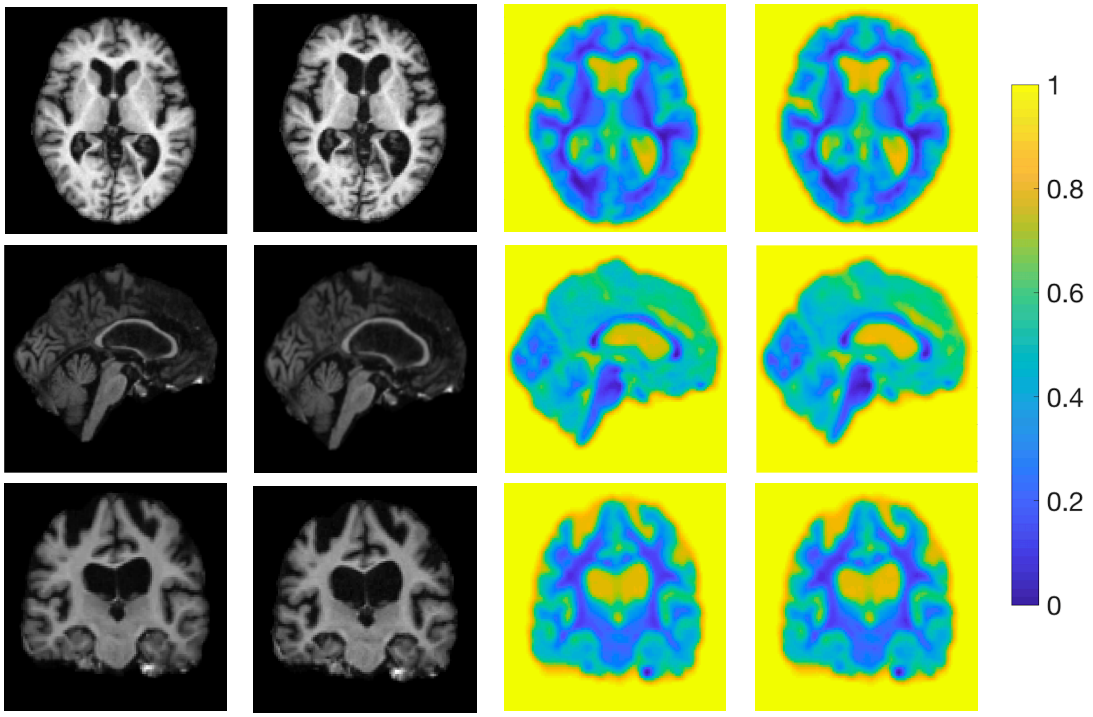


Fig. 8: Left to right: source image, target image, and uncertainty (visualized as the trace of covariance) estimated by baseline algorithm and our method.

Fig. 10 reports the absolute value of difference between the Hessian estimated by our method and the full Hessian on ten pairs of 2D synthetic data and twenty pairs of 3D MRI scans. The low errors on both 2D and 3D experiments indicate that our estimation is fairly close to the optimal uncertainty result.

Fig. 11 displays example error maps between deformed source images by the estimated mean (an optimal of registration) from all methods and target images. It shows that our

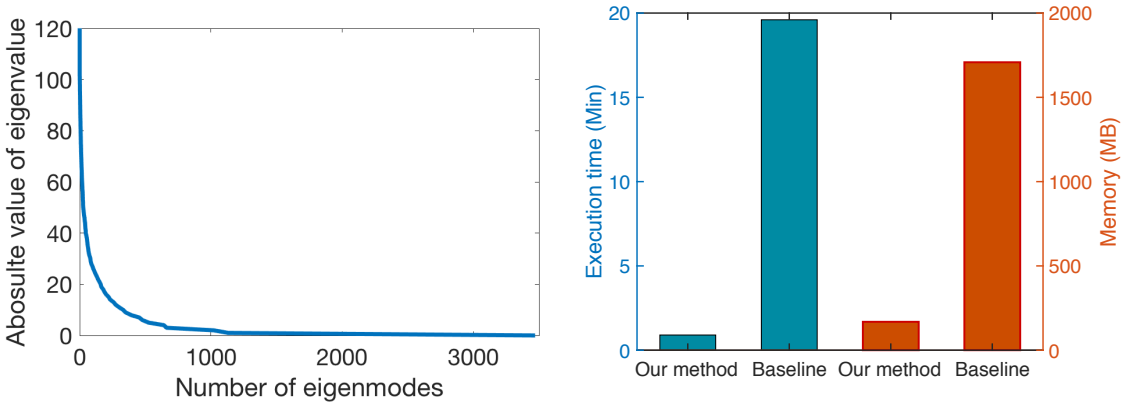


Fig. 9: Left panel: eigenvalues of the Hessian matrix estimated by the baseline algorithm; Right panel: comparison of average runtime and memory consumption for all test cases.

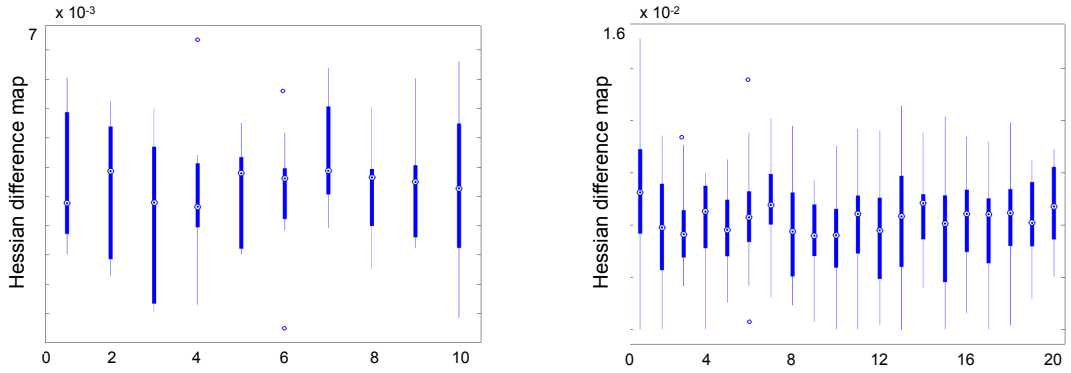


Fig. 10: Difference between the Hessian matrices estimated by our method and the full Hessian on both 2D and 3D data.

method achieves better registration results (a.k.a, smaller errors) than the two baseline algorithm [194, 14].

Fig. 12 reports the dice evaluations on brain cortex over 10 pairs of images for all methods. Our algorithm produces better dice score than the other two baseline algorithms.

2.6 Conclusion

In this section, I presented a novel Bayesian model for registration uncertainty quantification in the space of diffeomorphic transformations. In contrast to previous approaches, our method significantly reduced the computational cost of the registration posterior inference effectively via (i) a low-dimensional representation of velocity fields that are associated

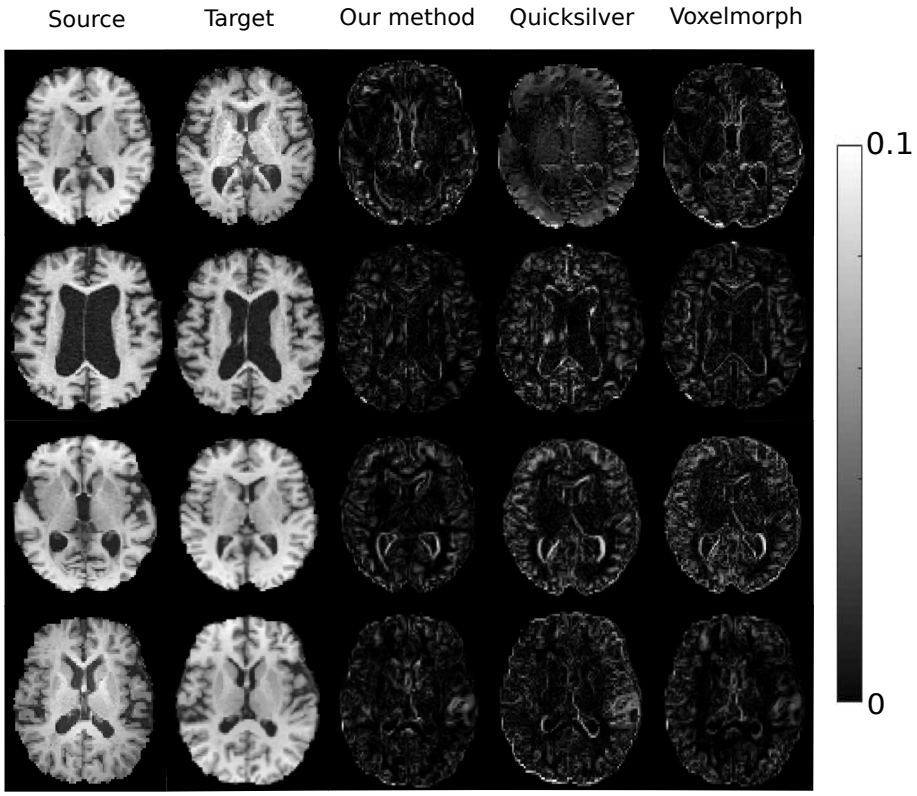


Fig. 11: Left to right: source image, target image, difference map of deformed image by estimated mean of our model, Quicksilver [194], and Voxelmorph [14].

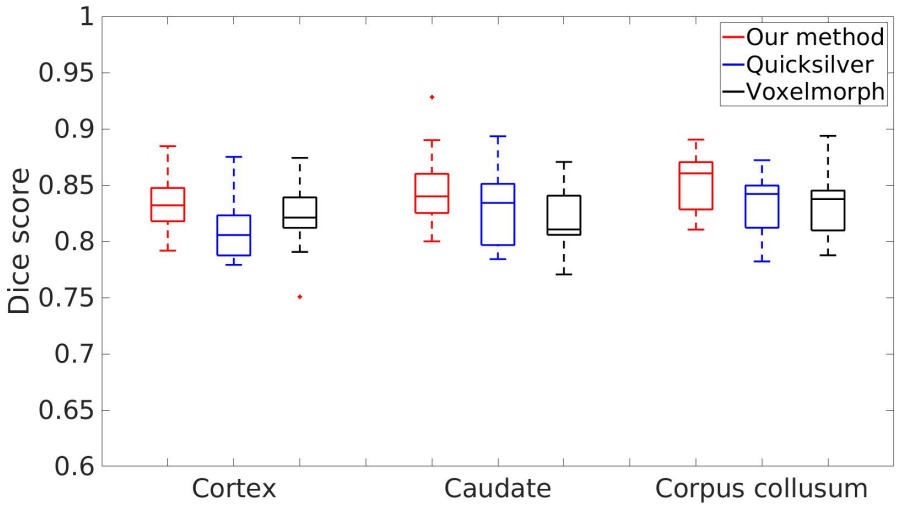


Fig. 12: Dice score evaluated on the brain structure, cortex, caudate and corpus collusum of our method, Quicksilver, and Voxelmorph.

with deformation fields; and (ii) an efficient Laplace's approximation to the posterior distribution, where the covariance was fully estimated by second-order methods. This work is the first step toward efficient probabilistic models of registration uncertainty quantification based on high-dimensional geometric transformations. Our future work will be investigating sampling methods to thoroughly assess our developed model uncertainty, and further extending its application to real clinical settings, e.g., real-time image-guided navigation system for neurosurgery. While in this work we focused on the representation of time-dependent velocity fields in the context of LDDMM, our method is general to other transformation parameterizations such as stationary velocity fields [6].

3 Bayesian Atlas Building & Regularization Parameter Estimation with Hierarchical Priors

3.1 Related Work

Deformable atlas building is to create a ‘‘mean‘‘ or averaged image and register all subjects to a common space. The resulting atlas and group transformations are powerful tools for statistical shape analysis of images [76, 107, 208], template-based segmentation [133, 146, 79], or object tracking [104, 102], just to name a few. A good quality of altas heavily relies on the registration process, which is typically formulated as a regularized optimization to solve [8, 15, 173, 203]. An issue in the current process of registration-based atlas construction is how to regularize model parameters. Having an appropriate regularization is critical to the ”sharpness” of the atlas, as well as ensuring a set of desirable properties of transformations, i.e., a smooth and invertible smooth mapping between images, also known as diffeomorphisms, to preserve the topology of original images.

Current atlas building models either exhaustively search an optimal regularization in the parameter space, or treat it as unknown variables to estimate from Bayesian models. While ad hoc parameter-tuning by users may yield satisfactory results, it requires expert domain knowledge to guide the tuning process [86, 172, 107]. Inspired by probabilistic models, several works have proposed Bayesian models of atlas building with automatically estimated regularizations [2, 3, 207]. These approaches define a posterior distribution that consists of an image matching term between a deformed atlas and each individual as a likelihood, and a regularization as a prior to support the smoothness of transformation fields. The regularization parameter is then jointly estimated with atlas after carefully

integrating out the image deformations by using Monte Carlo sampling. However, sampling in a high-dimensional transformation space (i.e., on a dense 3D image grid 128^3) is computationally expensive and often leads to a long execution time along with high memory consumption. More importantly, all aforementioned methods are limited to regularizations with single-penalty function for population studies. This greatly prohibits the model's ability to adaptively search for the best regularization parameter associated with each individual, which is critical to images with various degrees of geometric transformations.

We propose a hierarchical Bayesian model of atlas building with multi-penalty regularization in the context of Large Deformation Diffeomorphic Metric Mapping (LDDMM) algorithm [15]. In contrast to previous approaches treating the regularization of individual subjects as a single-penalty function with adhoc parameters, we develop a data-adaptive algorithm to automatically adjust the model parameters accordingly. To achieve this, we introduce a novel hierarchical prior that features (i) prior distributions with multiple regularization parameters on the group transformations in a low-dimensional bandlimited space; and (ii) a hyperprior to model the regularization parameters as latent variables generated from Gamma distribution. We then develop a Monte Carlo Expectation Maximization (MCEM) algorithm, where the expectation step integrates over the regularization parameters using Hamiltonian Monte Carlo (HMC) sampling. The joint estimation of model parameters including atlas, registration, and hyperparameters in the maximization step successfully eliminates a massive burden of multi-parameters tuning. We demonstrate the effectiveness of our algorithm on both 2D synthetic images and 3D real brain MRIs.

To the best of our knowledge, we are the first to extend the atlas building to a data-adaptive and parameter-tuning-free framework via hierarchical Bayesian learning. Experimental results show that our model provides an efficient atlas construction of population images, particularly with large variations of geometric transformations. This paves a way for an improved quality of clinical studies where atlas building is required, for example, statistical shape analysis of brain changes for neurodegenerative disease diagnosis [76], or atlas-based segmentation for in-utero placental disease monitoring [102].

3.2 Method: Bayesian Atlas Building with Hierarchical Priors

This section presents a hierarchical Bayesian model for atlas building that allows subject-specific regularization with no manual effort of parameter-tuning. We introduce a hierar-

chical prior distribution on the initial velocity fields with adaptive smoothing parameters followed by a likelihood distribution on images.

Likelihood. Assuming an independent and identically distributed (i.i.d.) Gaussian noise on image intensities, we formulate the likelihood of each observed image I_n as

$$p(I_n \,|\, I, \widetilde{v}_n, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma^2)^M} \exp\left(-\frac{1}{2\sigma^2}\|I \circ \phi_n - I_n\|_2^2\right). \tag{30}$$

Here σ^2 denotes a noise variance, M is the number of image voxels, and ϕ_n is an inverse Fourier transform of $\widetilde{\phi}_n$ at time point $t = 1$. It is worth mentioning that other noise models such as spatially varying noises [159] can also be applied.

Prior. To ensure the smoothness of transformation fields, we define a prior on each initial velocity field $\widetilde{v}_n$ as a complex multivariate Gaussian distribution

$$p(\widetilde{v}_n \,|\, \alpha_n) = \frac{1}{(2\pi)^{\frac{M}{2}} |\widetilde{\mathcal{L}}_n^{-1}(\alpha_n)|} \exp\left(-\frac{1}{2}\langle \widetilde{\mathcal{L}}_n(\alpha_n)\widetilde{v}_n, \widetilde{\mathcal{L}}(\alpha_n)\widetilde{v}_n\rangle\right), \tag{31}$$

where $|\cdot|$ is matrix determinant. The Fourier coefficients of a discrete Laplacian operator is $\widetilde{\mathcal{L}}_n(\xi_1, \ldots, \xi_d) = \left(-2\alpha_n \sum_{j=1}^{d} (\cos(2\pi\xi_j) - 1) + 1\right)^3$, with $(\xi_1, \ldots, \xi_d)$ being a d-dimensional frequency vector.

Hyperprior. We treat the subject-specific regularization parameter α_n of the prior distribution Eq (31) as a random variable generated from Gamma distribution, which is a commonly used prior to model positive real numbers [158]. Other prior such as inverse Wishart distribution [63] can also be applied. The hyperprior of our model is formulated as

$$p(\alpha_n \,|\, k, \beta) = \frac{\alpha_n^{k-1} \exp^{(-\alpha_n/\beta)}}{\Gamma(k)\beta^k}, \tag{32}$$

with k and β being positive numbers for shape and scale parameters respectively. The Gamma function $\Gamma(k) = (k-1)!$ for all positive integers of k. We finally arrive at the

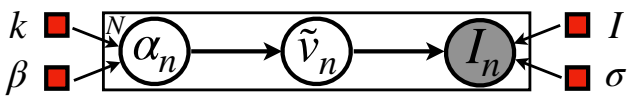


Fig. 13: Graphic model of hierarchical Bayesian atlas building.

log posterior of the diffeomorphic transformation and regularization parameters as

$$
\begin{aligned}
E(\widetilde{v}_n, \alpha_n, I, \sigma, k, \beta) &\triangleq \ln \prod_{n=1}^{N} p(I_n \,|\, I, \widetilde{v}_n, \sigma^2) \cdot p(\widetilde{v}_n \,|\, \alpha_n) \cdot p(\alpha_n \,|\, k, \beta) \\
&= \sum_{n=1}^{N} \frac{1}{2} \ln|\mathcal{L}_n| - M \ln \sigma - \frac{\|I \circ \phi_n - I_n\|_2^2}{2\sigma^2} - \frac{1}{2}(\widetilde{\mathcal{L}}\widetilde{v}_n, \widetilde{\mathcal{L}}\widetilde{v}_n) \\
&\quad (k-1) \ln \alpha_n - \frac{\alpha_n}{\beta} - k \ln \beta - \ln \Gamma(k) + \text{const.}
\end{aligned} \tag{33}
$$

3.2.1 Model Inference We develop an MCEM algorithm to infer the model parameter Θ, which includes the image atlas I, the noise variance of image intensities σ^2, the initial velocities of diffeomorphic transformations $\widetilde{v}_n$, and the hyperparameters k and β. We treat the regularization parameter α_n as latent random variables and integrate them out from the log posterior in Eq (33). Computations of two main steps (expectation and maximization) are illustrated below.

Expectation: HMC. Since the E-step does not yield a closed-form solution, we employ a powerful Hamiltonian Monte Carlo (HMC) sampling method [49] to approximate the expectation function Q with respect to the latent variables α_n. For each α_n, we draw a number of S samples from the log posterior Eq (33) by using HMC from the current estimated parameters $\widehat{\Theta}$. The Monte Carlo approximation of the expectation Q is

$$
Q(\Theta|\widehat{\Theta}) \approx \frac{1}{S} \sum_{n=1}^{N} \sum_{j=1}^{S} \ln p(\alpha_{nj} \,|\, I_n; \widehat{\Theta}). \tag{34}
$$

To produce samples of α_n, we first define the potential energy of the Hamiltonian system $H(\alpha_n, \gamma) = U(\alpha_n) + W(\gamma)$ as $U(\alpha_n) = -\ln p(\alpha_n|I_n; \Theta)$. The kinetic energy $W(\gamma)$ is a typical normal distribution on an auxiliary variable γ. This gives us Hamilton's equations to integrate

$$
\frac{\alpha_n}{dt} = \frac{\partial H}{\partial \gamma} = \gamma, \quad \frac{d\gamma}{dt} = -\frac{\partial H}{\partial \alpha_n} = -\nabla_{\alpha_n} U. \tag{35}
$$

Since α_n is a Euclidean variable, we use a standard "leap-frog" numerical integration scheme, which approximately conserves the Hamiltonian and results in high acceptance rates. The gradient of U with respect to α_n is

$$
\nabla_{\alpha_n} U = \frac{3}{2S} \sum_{j=1}^{S} [\sum_{i=1}^{d} \frac{\widetilde{\mathcal{A}}_i}{\alpha_{nj} \widetilde{A}_i + 1} - \langle 2(\alpha_{nj} \widetilde{\mathcal{A}} + 1)^5 \widetilde{\mathcal{A}} \widetilde{v}_{nj}, \widetilde{v}_{nj} \rangle], \tag{36}
$$

where $\widetilde{\mathcal{A}} = -2\sum_{i=1}^{d}(\cos(2\pi\xi_i) - 1)$. Here $\widetilde{\mathcal{A}}$ denotes a discrete Fourier Laplacian operator with a d-dimensional frequency vector.

Starting from the current point α_n and initial random auxiliary variable γ, the Hamiltonian system is integrated forward in time by Eq (35) to produce a candidate point $(\widehat{\alpha}_n, \widehat{\gamma})$. The candidate point $\widehat{\alpha}_n$ is accepted as a new point in the sample with probability $p(accept) = \min(1, -U(\widehat{\alpha}_n) - W(\widehat{\gamma}) + U(\alpha_n) + W(\gamma))$.

Maximization: Gradient Ascent. We derive the maximization step to update the parameters $\Theta = \{I, \widetilde{v}_n, \sigma^2, k, \beta\}$ by maximizing the HMC approximation of the expectation Q in Eq (34).

For updating the atlas image I, we set the derivative of the Q function with respect to I to zero. The solution for I gives a closed-form update

$$I = \frac{\sum_{j=1}^{S}\sum_{n=1}^{N}(I_n \circ \phi_{nj}^{-1}) \cdot |D\phi_{nj}^{-1}|}{\sum_{j=1}^{S}\sum_{n=1}^{N}|D\phi_{nj}^{-1}|}. \tag{37}$$

Similarly, we obtain the closed-form solution for the noise variance σ^2 after setting the gradient of Q w.r.t. σ^2 to zero

$$\sigma^2 = \frac{1}{MNS}\sum_{n=1}^{N}\sum_{j=1}^{S}\|I \circ \phi_{nj} - I_n\|_2^2. \tag{38}$$

The closed-form solutions for hyperparameters k and β are

$$k = \psi^{-1}(\frac{1}{NS}\sum_{i=1}^{N}\sum_{j=1}^{S}\ln\alpha_{nj} - \ln\beta), \quad \beta = \frac{1}{NSk}\sum_{n=1}^{N}\sum_{j=1}^{S}\alpha_{nj}. \tag{39}$$

Here ψ is a digamma function, which is the logarithmic derivative of the gamma function $\Gamma(\cdot)$. The inverse of digamma function ψ^{-1} is computed by using a fixed-point iteration algorithm [117].

As there is no closed-form update for initial velocities, we employ a gradient ascent algorithm to estimate $\widetilde{v}_{nj}$. The gradient $\nabla_{\widetilde{v}_{nj}}Q$ is computed by a forward-backward sweep approach. Details are introduced in the FLASH algorithm [204].

3.3 Experimental Evaluation

We compare the proposed model with LDDMM atlas building algorithm that employs single-penalty regularization with manually tuned parameters on 3D brain images [204].

In HMC sampling, we draw 300 samples for each subject, with initialized value of $\alpha = 10$, $k = 9.0$, $\sigma = 0.05$, and $\beta = 0.1$. An averaged image of all image intensities is used for atlas initialization.

Data. We include 100 3D brain MRI scans with segmentation maps from OASIS [56]. The dataset covers both healthy and diseased subjects, aged from 55 to 90. The MRI scans are resampled to 128^3 with the voxel size of $1.25mm^3$. All MRIs are carefully prepossessed by skull-stripping, intensity normalization, bias field correction, and co-registration with affine transformation.

Experiments. We estimate the atlas of all deformed images by using our method and compare its performance with LDDMM atlas building [204]. Final results of atlases estimated from both our model and the baseline algorithm are reported. We also compare the time and memory consumption of proposed model with the baseline that performs HMC sampling in a full spatial domain [207]. To measure the sharpness of estimated atlas I, we adopt a metric of normalized standard deviation computed from randomly selected 3000 image patches [98]. Given $N(i)$, a patch around a voxel i of an atlas I, the local measure of the sharpness at voxel i is defined as $\text{sharpness}(I(i)) = \text{sd}_{N(i)}(I)/\text{avg}_{N(i)}(I)$, where sd and avg denote the standard deviation and the mean of N_i.

To further evaluate the quality of estimated transformations, we perform atlas-based segmentation after obtaining transformations from our model. For a fair comparison, we fix the atlas for both methods and examine the registration accuracy by computing the dice similarity coefficient (DSC) [46] between the propagated segmentation and the manual segmentation on six anatomical brain structures, including cerebellum white matter, thalamus, brain stem, lateral ventricle, putamen, caudate. The significance tests on both dice and sharpness between our method and the baseline are performed.

Results. Fig. 14 visualizes a comparison of 3D atlas on real brain MRI scans. The top panel shows that our model substantially improves the quality of atlas with sharper and better details than the baseline with different values of manually set regularization parameters, e.g., $\alpha = 0.1, 3.0, 6.0, 9.0$. Despite the observation of a smaller value of $\alpha = 0.1$ produces sharper atlas, it breaks the smoothness constraints on the transformation fields hence introducing artifacts on anatomical structures (outlined in purple boxes). The mean and standard deviation of our estimated hyperprior parameters k and β in Eq (32) over 30 pairwise image registrations are $47.40/7.22$, and $0.036/0.005$. The bottom panel quantitatively reports the sharpness metric of all methods. It indicates that our algorithm outperforms the baseline by offering a higher sharpness score while preserving the topological structure of brain anatomy.

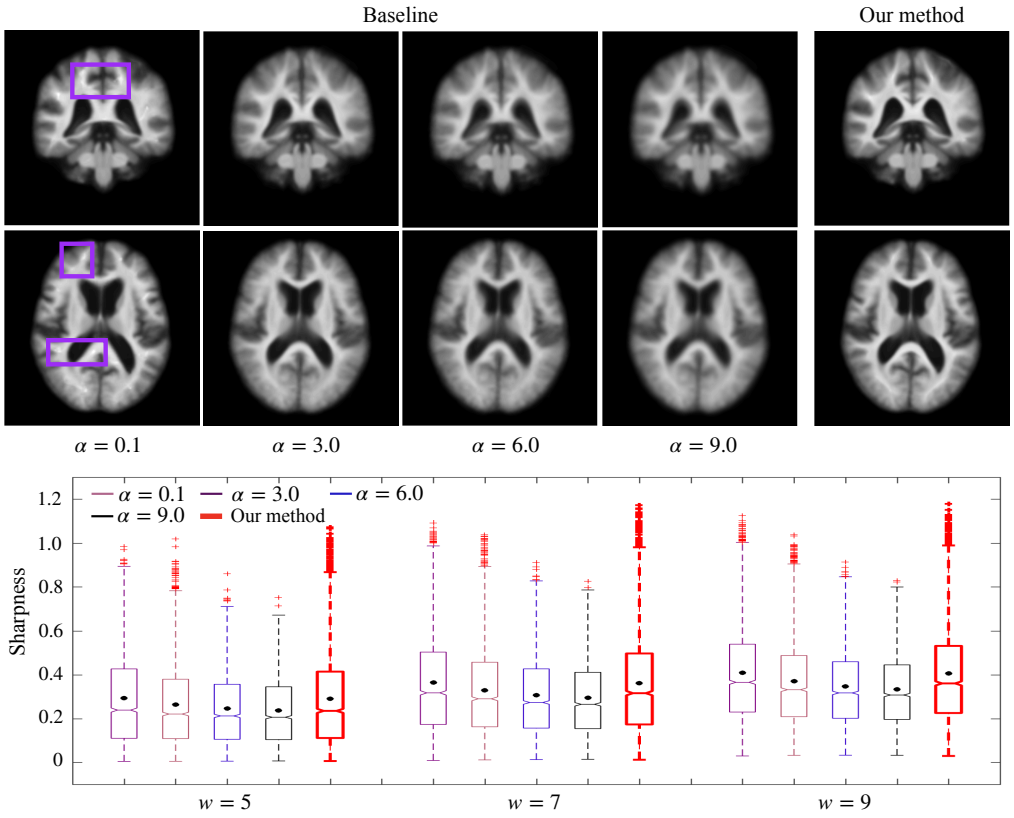


Fig. 14: Top: atlases estimated by baseline with different α and our model (artifacts introduced by small regularization are outlined in purple boxes). Bottom: sharpness measurement of atlas for all methods with different patch size w. The mean of the sharpness metric of **our method** vs. the best performance of baseline without artifacts ($\alpha = 3$) is **0.290**/0.264, **0.362**/0.323, **0.405**/0.360.

Fig. 15 reports results of fixed-atlas-based segmentation by performing the baseline with various regularization parameters and our algorithm. It shows the dice comparison on six anatomical brain structures of all image pairs. Our algorithm produces better dice coefficients without the need of parameter tuning.

The runtime of our atlas building on 100 3D brain MR images are 4.4 hours with 0.89GB memory consumption. The p-values of significance differences test on both dice ($p = 0.002$) and sharpness ($p = 0.0034$) reject the null hypothesis that there's no differences between our model estimation and baseline algorithms.

3.4 Conclusion

In this section, I present a novel hierarchical Bayesian model for unbiased diffeomorphic atlas building with subject-specific regularization. We design a new parameter choice rule

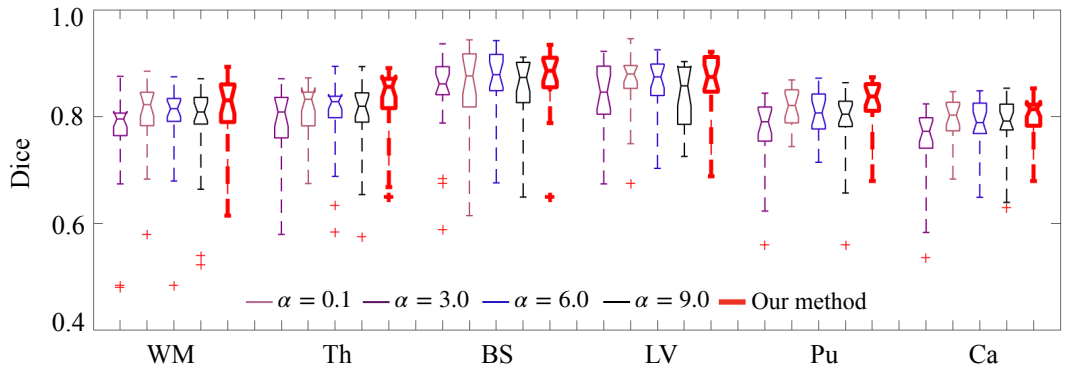


Fig. 15: A comparison of dice evaluation for fixed-atlas-based segmentation on six brain structures (cerebellum white matter (WM), thalamus (Th), brain stem (BS), lateral ventricle (LV), putamen(Pu), caudate (Ca)).

that allows adaptive regularization to control the smoothness of image transformations. We introduce a hierarchical prior that provides prior information of regularization parameters at multiple levels. The developed MCEM inference algorithm eliminates the need of manual parameter tuning, which can be tedious and infeasible in multi-parameter settings. Experimental results show that our proposed algorithm yields a better registration model as well as an improved quality of atlas. While our algorithm is presented in the setting of LDDMM, the theoretical development is generic to other deformation models, e.g., stationary velocity fields [6]. In addition, this model can be easily extended to multi-atlas building where a much higher degree of variations exist in the population studies. Our future work will focus on conducting subsequent statistical shape analysis in the resulting atlas space.

Chapter IV

Deep Neural Networks that Learn Deformable Shapes from Images

In this chapter, I introduce the developed deep neural networks that learn low-dimensional geometric deformable shape representations. Sec. 1 presents DeepFLASH, an efficient deep neural network designed for diffeomorphic image registration. It utilizes compact deformable shape representations, resulting in a significant reduction in computational complexity. This makes the model faster to perform training, testing and inferefence. Sec. 2 presents Geo-SIC, a classification network that incorporates deformable geometric shape learning into deep image classifiers. This integration leads to improved performance. Sec. 3 presents an efficient predictive approach for regularization parameter estimation. This method automatically produces regularization parameters through deep neural networks, allowing for faster training data generation and parameter estimation. The developed low-dimensional Bayesian framework maintains comparable registration results while significantly reducing computation time.

1 An Efficient Predictive Network for Learning-based Medical Image Registration

1.1 Related Work

We recall the basics of deformable image registration in Chapter II. The problem of deformable image registration is typically formulated as an optimization, seeking for a nonlinear and dense (voxelwise) spatial transformation between images. In many applications, it is desirable that such transformations be diffeomorphisms, i.e., differentiable,

bijective mappings with differentiable inverses. In this section, we focus on diffeomorphic image registration highlighting with a set of critical features: (i) it captures large deformations that often occur in brain shape variations, lung motions, or fetal movements; (ii) the topology of objects in the image remain intact; and (iii) no non-differentiable artifacts, such as creases or sharp corners, are introduced. However, achieving an optimal solution of diffeomorphic image registration, especially for large-scale or high-resolution images (e.g., a 3D brain MRI scan with the size of 256^3), is computationally intensive and time-consuming.

Attempts at speeding up diffeomorphic image registration have been made in recent works by improving numerical approximation schemes. For example, Ashburner and Friston [8] employ a Gauss-Newton method to accelerate the convergence of large deformation diffeomorphic metric mapping (LDDMM) algorithm [15]. Zhang et al. propose a low-dimensional approximation of diffeomorphic transformations, resulting in fast computation of the gradients for iterative optimization [206, 179]. While these methods have led to substantial reductions in running time, such gradient-based optimization still takes minutes to finish. Instead of minimizing a complex registration energy function [15, 204], an alternative approach has leveraged deep learning techniques to improve registration speed by building prediction models of transformation parameters. Such algorithms typically adopt convolutional neural networks (CNNs) to learn a mapping between pairwise images and associated spatial transformations in training dataset [194, 145, 14, 35, 28]. Registration of new testing images is then achieved rapidly by evaluating the learned mapping on given volumes. While the aforementioned deep learning approaches are able to fast predict the deformation parameters in testing, the training process is extremely slow and memory intensive due to the high dimensionality of deformation parameters in imaging space. In addition, enforcing the smoothness constraints of transformations when large deformation occurs is challenging in neural networks.

To address this issue, we propose a novel learning-based registration framework DeepFLASH in a low-dimensional bandlimited space, where the diffeomorphic transformations are fully characterized with much fewer dimensions. Our work is inspired by a recent registration algorithm FLASH (Fourier-approximated Lie Algebras for Shooting) [204] with the novelty of (i) developing a learning-based predictive model that further speeds up the current registration algorithms; (ii) defining a set of complex-valued operations (e.g., complex convolution, complex-valued rectifier, etc.) and complex-valued loss function of transformations entirely in a bandlimited space; and (iii) proving that our model can be easily implemented by a dual-network in the space of real-valued functions with a careful design of the network architecture. To the best of our knowledge, we are the

first to introduce the low-dimensional Fourier representations of diffeomorphic transformations to learning-based registration algorithms. In contrast to traditional methods that learn spatial transformations in a high-dimensional imaging space, our method dramatically reduces the computational complexity of the training process where iterative computation of gradient terms are required. This greatly alleviates the problem of time-consuming and expensive training for deep learning based registration networks. Another major benefit of DeepFLASH is that the smoothness of diffeomorphic transformations is naturally preserved in the bandlimited space with low frequency components. Note that while we implement DeepFLASH in the context of convolutional neural network (CNN), it can be easily adapted to a variety of other neural networks, such as fully connected network (FCN), or recurrent neural network (RNN). We demonstrate the effectiveness of our model in both 2D synthetic and 3D real brain MRI data.

1.2 Method: DeepFLASH

1.2.1 Computational Operations We introduce a learning-based registration network DeepFLASH in a low-dimensional bandlimited space $\widetilde{V}$, with newly defined operators and functions in a complex vector space $\mathbb{C}^n$. Since the spatial transformations can be uniquely determined by an initial velocity $\widetilde{v}_0$ (as introduced in Eq (5)), we naturally integrate this new parameterization in the architecture of DeepFLASH. To simplify the notation, we drop the time index of $\widetilde{v}_0$ in remaining sections.

Analogous to [194], we use optimal registration results, denoted as $\widetilde{v}^{opt}$, estimated by numerical optimization of the LDDMM algorithm as part of the training data. Our goal is then to predict an initial velocity $\widetilde{v}^{pre}$ from image patches of the moving and target images. Before introducing DeepFLASH, we first define a set of complex-valued operations and functions that provide key components of the neural architecture.

Consider a Q-dimensional complex-valued vector of input signal $\widetilde{X}$ and a real-valued kernel H, we have a complex-valued convolution as

$$H * \widetilde{X} = H * R(\widetilde{X}) + iH * I(\widetilde{X}), \tag{40}$$

where $R(\cdot)$ denotes the real part of a complex-valued vector, and $I(\cdot)$ denotes an imaginary part.

Following a recent work on complex networks [168], we define a complex-valued activation function based on Rectified Linear Unit (ReLU). We apply real-valued ReLU

separately on both of the real and imaginary part of a neuron $\widetilde{Y}$ in the output layer, i.e.,

$$\mathbb{C}\text{ReLU}(\widetilde{Y}) = \text{ReLU}(R(\widetilde{Y})) + i\text{ReLU}(I(\widetilde{Y})). \tag{41}$$

Loss function Let a labeled training dataset including pairwise images and their associated optimal initial velocity fields be $\{S_n, T_n, \widetilde{v}_n^{opt}\}_{n=1}^N$, where N is the number of pairwise images. We model a prediction function $f(S_n, T_n; W)$ by using a convolutional neural network (CNN), with W being a real-valued weight matrix for convolutional layers. We then define a loss function ℓ as

$$\ell(W) = \sum_{n=0}^{N} \|\widetilde{v}_n^{opt} - f(S_n, T_n; W)\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \tag{42}$$

where λ is a positive parameter balancing between function f and a regularity term $\text{Reg}(\cdot)$ on the weight matrix W. While we use L_2 norm as regularity, it is generic to other commonly used operators such as L_1, or L_0 norm.

The optimization problem of Eq (42) is typically solved by using gradient-based methods. The weight matrix W is updated by moving in the direction of the loss function‘‘s steepest descent, found by its gradient $\nabla_W \ell$.

1.2.2 Network Design While we are ready to design a complex-valued registration network based on CNN, the implementation of such network is not straightforward. Trabelsi et al. developed deep complex networks to specifically handle complex-valued inputs at the cost of computational efficiency [168]. In this section, we present an efficient way to decouple the proposed complex-valued registration network into a combination of regular real-valued networks. More specifically, we construct a dual-net that separates the real and imaginary part of complex-valued network in an equivalent manner.

Given the fact that the computation of real and imaginary parts in the convolution (Eq (40)) and the activation function (Eq (41)) are separable, we next show that the loss defined in Eq (42) can be equivalently constructed by the real and imaginary part individually. To simplify the notation, we define the predicted initial velocity $\widetilde{v}_n^{pre} \triangleq f(S_n, T_n; W)$ and rewrite Eq (42) as

$$\ell(W) = \sum_{n=0}^{N} \|\widetilde{v}_n^{opt} - \widetilde{v}_n^{pre}\|_{L_2}^2 + \lambda \cdot \text{Reg}(W). \tag{43}$$

Let $\widetilde{v}_n^{opt} = \beta_n + i\mu_n$ and $\widetilde{v}_n^{pre} = \delta_n + i\eta_n$ in Eq (43), we then obtain

$$\begin{aligned}
\ell(W) &= \sum_{n=0}^{N} \|(\beta_n + i\mu_n) - (\delta_n + i\eta_n)\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \\
&= \sum_{n=0}^{N} \|(\beta_n - \delta_n) + i(\mu_n - \eta_n)\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \\
&= \sum_{n=0}^{N} \|\beta_n - \delta_n\|_{L_2}^2 + \|\mu_n - \delta_n\|_{L_2}^2 + \lambda \cdot \text{Reg}(W), \\
&= \sum_{n=0}^{N} \|R(\widetilde{v}_n^{opt}) - R(\widetilde{v}_n^{pre})\|_{L_2}^2 \\
&+ \|I(\widetilde{v}_n^{opt}) - I(\widetilde{v}_n^{pre})\|_{L_2}^2 + \lambda \cdot \text{Reg}(W). \qquad (44)
\end{aligned}$$

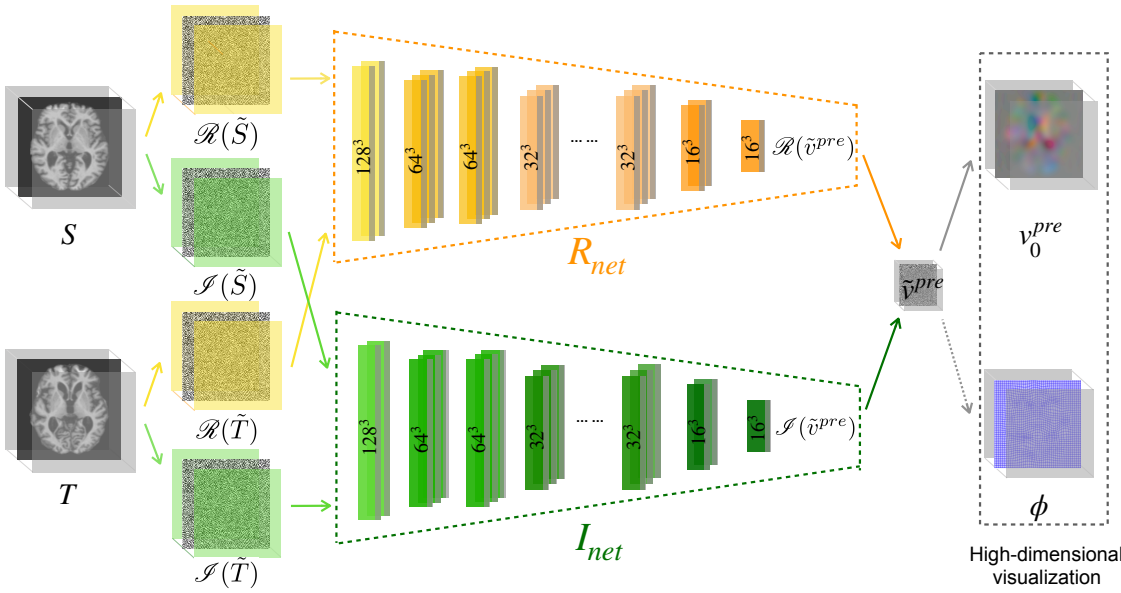


Fig. 16: Architecture of DeepFLASH with dual net: S and T denote the source and target image from high-dimensional spatial domain. $R(\widetilde{S})$ and $I(\widetilde{S})$ are the real and imaginary frequency spectrum convert from S. $R(\widetilde{T})$ and $I(\widetilde{T})$ are the real and imaginary frequency convert from T. $\widetilde{v}^{pre}$ is the low-dimensional prediction optimized from our model. v_0^{pre} and ϕ are the velocity and transformation field we recovered from our low-dimensional prediction.

Fig. 16 visualizes the architecture of our proposed model DeepFLASH. The input source and target images are defined in the Fourier space with the real part of frequencies as $R(\widetilde{S})$ and $R(\widetilde{T})$ vs. the imaginary parts as $I(\widetilde{S})$ and $I(\widetilde{T})$. We train real and imaginary parts in two individual neural networks R_{net} and I_{net}. The optimization stays in a low-dimensional bandlimited space without converting the coefficients ($R(\widetilde{v}^{pre})$, $I(\widetilde{v}^{pre})$) to

high-dimensional imaging domain back and forth. It is worthy to mention that our decoupled network is not constrained by any specific architecture. The CNN network in Fig. 16 can be easily replaced by a variety of the state-of-the-art models, e.g., U-net [147], or fully connected neural network.

1.2.3 Computational Complexity It has been previously shown that the time complexity of convolutional layers [69] is $O(\sum_{p=1}^{P} b_{p-1} \cdot h_p^2 \cdot b_p \cdot Z_p^2)$, where p is the index of a convolutional layer and P denotes the number of convolutional layers. The b_{p-1}, b_p h_p are defined as the number of input channels, output channels and the kernel size of the p-th layer respectively. Here Z_p is the output dimension of the p-th layer, i.e., $Z_p = 128^3$ when the last layer predicts transformation fields for 3D images with the dimension of 128^3.

Current learning approaches for image registration have been performed in the original high-dimensional image space [14, 194]. In contrast, our proposed model DeepFLASH significantly reduces the dimension of Z_p into a low-dimensional bandlimited space z_p (where $z_p \ll Z_p$). This makes the training of traditional registration networks much more efficient in terms of both time and memory consumption.

1.3 Experimental Evaluation

We demonstrate the effectiveness of our model by training and testing on both 2D synthetic data and 3D real brain MRI scans.

2D synthetic data. We first simulate 3000 "bull-eye" synthetic data (as shown in Fig 17) by manipulating the width a and height b of an ellipse equation, formulated as $\frac{(x-50)^2}{a^2} + \frac{(y-50)^2}{b^2} = 1$. We draw the parameter a, b randomly from a Gaussian distribution $\mathcal{N}(4, 2^2)$ for inner ellipse, and $a, b \sim \mathcal{N}(13, 4^2)$ for outer ellipse.

3D brain MRI. We include 3200 public T1-weighted 3D brain MRI scans from ADNI [80], OASIS [56], Autism Brain Imaging Data Exchange (ABIDE) [45], and LONI Probabilistic Brain Atlas Individual Subject Data (LPBA40) [156] with 1000 subjects. Due to the difficulty of preserving the diffeomorphic property across individual subjects particularly with large age variations, we carefully evaluate images from subjects aged from 60 to 90. All MRIs were all pre-processed as $128 \times 128 \times 128$, $1.25mm^3$ isotropic voxels, and underwent skull-stripped, intensity normalized, bias field corrected and pre-aligned with affine transformation.

To further validate our model accuracy through segmentation labels, we use manually delineated anatomical structures in LPBA40 dataset. We randomly select 100 pairs of

MR images with segmentation labels from LPBA40. We then carefully down-sampled all images and labels from the dimension of $181 \times 217 \times 181$ to $128 \times 128 \times 128$.

Experiments To validate the registration performance of DeepFLASH on both 2D and 3D data, we run a recent state-of-the-art image registration algorithm FLASH [203, 206] on 1000 2D synthetic data and 2000 pairs of 3D MR images to optimal solutions, which are used as ground truth. We then randomly choose 500 pairs of 2D data and 1000 pairs of 3D MRIs from the rest of the data as testing cases. We set registration parameter $\alpha = 3, c = 6$ for the operator $\widetilde{\mathcal{L}}$, the number of time steps for Euler integration in geodesic shooting as 10. We adopt the band-limited dimension of the initial velocity field $\widetilde{v}^{opt}$ as 16, which has been shown to produce comparable registration accuracy [203]. We set the batch size as 64 with learning rate $\eta = 1e-4$ for network training, then run 2000 epochs for 2D synthetic training and 5000 epochs for 3D brain training.

Next, we compare our 2D prediction with registration performed in full-spatial domain. For 3D testing, we compare our method with three baseline algorithms, including FLASH [204] (a fast optimization-based image registration method), Voxelmorph [14] (an unsupervised registration in image space) and Quicksilver [194] (a supervised method that predicts transformations in a momentum space). We also compare the registration time of DeepFLASH with traditional optimization-based methods, such as vector momenta LDDMM (VM-LDDMM) [161], and symmetric diffeomorphic image registration with cross-correlation (SyN) from ANTs [10]. For fair comparison, we train all baseline algorithms on the same dataset and report their best performance from published source codes.

To better validate the transformations generated by DeepFLASH, we perform registration based segmentation and examine the resulting segmentation accuracy over eight brain structures, including Putamen (Puta), Cerebellum (Cer), Caudate (Caud), Hippocampus (Hipp), Insular cortex (Cor), Cuneus (Cune), brain stem (Stem) and superior frontal gyrus (Gyrus). We evaluate a volume-overlapping similarity measurement, also known as Dice coefficient, between the propagated segmentation and the manual segmentation [46].

We demonstrate the efficiency of our model by comparing quantitative time and GPU memory consumption across all methods. All optimal solutions for training data are generated on an i7, 9700K CPU with 32 GB internal memory. The training and prediction procedure of all learning-based methods are performed on Nvidia GTX 1080Ti GPUs.

Results Fig. 17 visualizes the deformed images, transformation fields, and the determinant of Jacobian (DetJac) of transformations estimated by DeepFLASH and a registration method that performed in full-spatial image domain. Note that the value of DetJac indi-

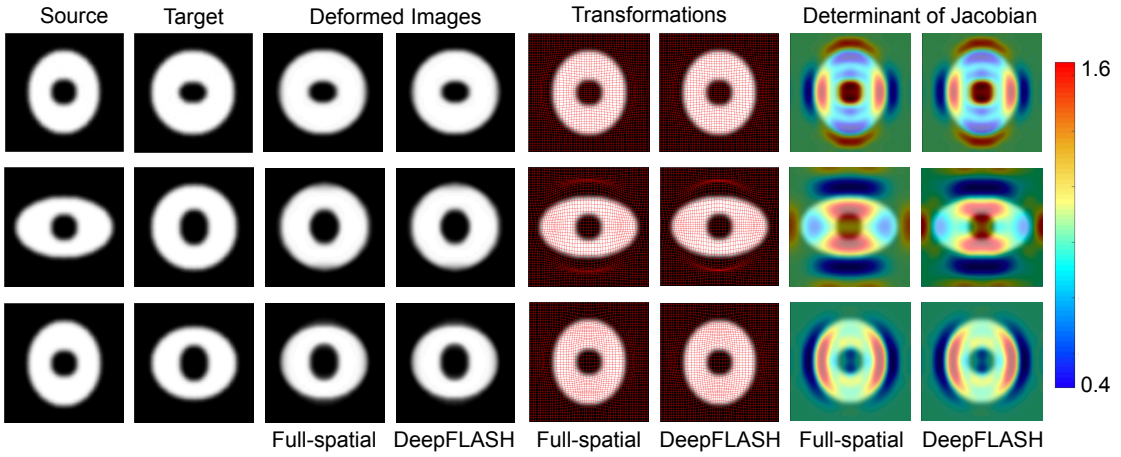


Fig. 17: Example of 2D registration results. Left to right: 2D synthetic source, target image, deformed image computed in the full-spatial domain, transformation grids overlaid with source image, and determinant of Jacobian (DetJac) of transformations.

cates how volume changes, for instance, there is no volume change when DetJac $= 1$, while volume shrinks when DetJac < 1 and expands when DetJac > 1. The value of DetJac smaller than zero indicates an artifact or singularity in the transformation field, i.e., a failure to preserve the diffeomorphic property when the effect of folding occurs. Both methods show similar patterns of volume changes over transformation fields. Our predicted results are fairly close to the estimates from registration algorithms in full spatial domain.

Fig. 18 visualizes the deformed images and the determinant of Jacobian with pairwise registration on 3D brain MRIs for all methods. It demonstrates that our method DeepFLASH is able to produce comparable registration results with little to no loss of the accuracy. In addition, our method gracefully guarantees the smoothness of transformation fields without artifacts and singularities.

The left panel of Fig. 19 displays an example of the comparison between the manually labeled segmentation and propagated segmentation deformed by transformation field estimated from our algorithm. It clearly shows that our generated segmentations align fairly well with the manual delineations. The right panel of Fig. 19 compares quantitative results of the average dice for all methods, with the observations that our method slightly outperforms the baseline algorithms.

Fig. 20 reports the statistics of dice scores (mean and variance) of all methods over eight brain structures from 100 registration pairs. Our method DeepFLASH produces comparable dice scores with significantly fast training of the neural networks.

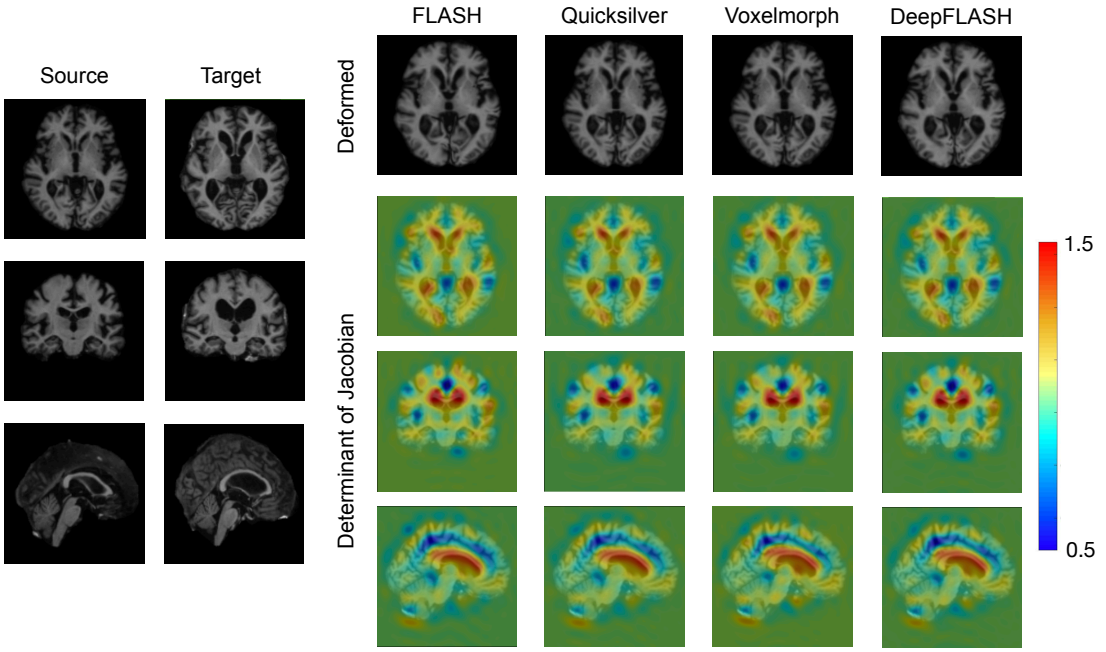
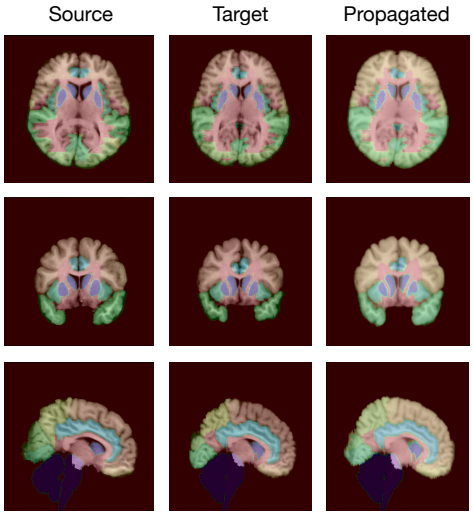


Fig. 18: Example of 3D image registration on OASIS dataset. Left panel: axial, coronal, and sagittal view of source and target images. Right panel: deformed images and determinant of Jacobian of the transformations by FLASH, Quicksilver, Voxelmorph, and DeepFLASH.



Methods	Average Dice
VM-LDDMM	0.760
ANTs(SyN)	0.770
FLASH	0.788
Quicksilver	0.762
Voxelmorph	0.774
DeepFLASH	0.780

Fig. 19: Left: source and target segmentations with manually annotated eight anatomical structures (on LPBA40 dataset), propagated segmentation label deformed by our method. Right: quantitative result of average dice for all methods.

Table.1 reports the time consumption of optimization-based registration methods, as well as the training and testing time of learning-based registration algorithms. Our method

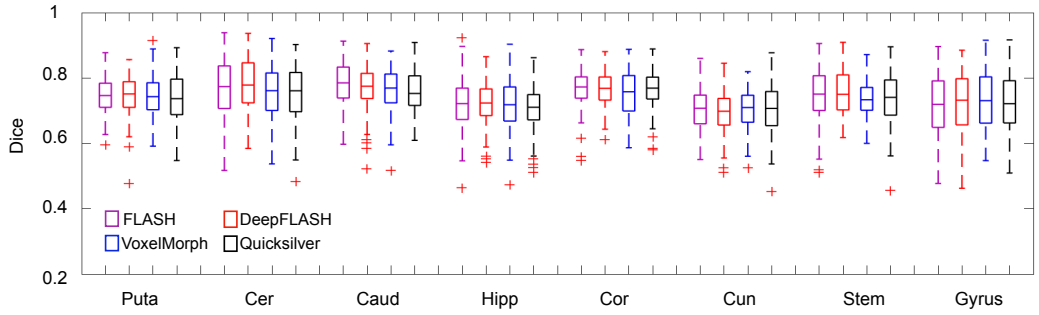


Fig. 20: Dice score evaluation by propagating the deformation field to the segmentation labels for four methods on eight brain structures (Putamen(Puta), Cerebellum (Cer), Caudate (Caud), Hippocampus (Hipp), Insular cortex (Cor), Cuneus (Cune), brain stem (Stem), and superior frontal gyrus (Gyrus)).

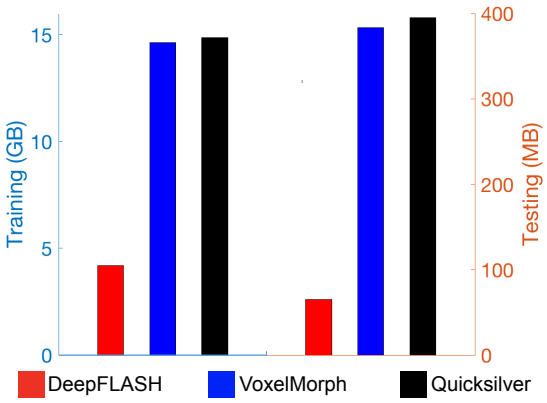


Fig. 21: Comparison of average GPU memory usage on our method and learning-based baselines for both training and testing processes.

can predict the transformation fields between images approximately 100 times faster than optimization-based registration methods. In addition, DeepFLASH outperforms learning-based registration approaches in testing, while with significantly reduced computational time in training.

Fig. 21 provides the average training and testing GPU memory usage across all methods. It has been shown that our proposed method dramatically lowers the GPU footprint compared to other learning-based methods in both training and testing.

1.4 Conclusion

In this section, I present a novel diffeomorphic image registration network with efficient training process and inference. In contrast to traditional learning-based registration meth-

Table 1: Top: quantitative results of time consumption on both CPU and GPU for optimization-based registration methods (there is no GPU-version of ANTs). Bottom: computational time of training and prediction for our method DeepFLASH and current deep learning registration approaches.

Methods	Training Time (GPU hours)	Registration time (sec)	
		CPU	GPU
VM-LDDMM	-	1210	262
ANTs(SyN)	-	6840	-
FLASH	-	286	53.4
Quicksilver	31.4	122	0.760
Voxelmorph	29.7	52	0.571
DeepFLASH	**14.1**	**41**	**0.273**

ods that are defined in the high-dimensional imaging space, our model is fully developed in a low-dimensional bandlimited space with the diffeomorphic property of transformation fields well preserved. Based on the fact that current networks are majorly designed in real-valued spaces, we are the first to develop a decoupled-network to solve the complex-valued optimization with the support of rigorous math foundations. Our model substantially lowers the computational complexity of the neural network and significantly reduces the time consumption on both training and testing, while preserving a comparable registration accuracy. The theoretical tools developed in our work is flexible/generic to a wide variety of the state-of-the-art networks, e.g., FCN, or RNN. To the best of our knowledge, we are the first to characterize the diffeomorphic deformations in Fourier space via network learning. This work also paves a way for further speeding up unsupervised learning for registration models.

2 Learning Deformable Geometric Shapes in Deep Image Classifiers

2.1 Related Work

For image classification, deformable shapes have been recognized as powerful tool [13, 61, 118] since they capture geometric features that describe changes of objects with complex structures in images, making them a reliable cue for image analysis tasks [127, 118,

193]. Bountiful literature demonstrates that the robustness of shapes to variations in image intensity and texture (e.g., noisy or corrupted data) makes it a reliable cue for image analysis tasks [127, 118, 193]. Existing methods have studied various representations of geometric shapes, including landmarks [16, 23, 40], point clouds [1], binary segmentations [25, 163], and medial axes [132]. A very recent research area in geometric deep learning [26, 139] has investigated mathematical representations of shapes in the form of analytic graphs or points and then uses them to synthesize shapes. These aforementioned techniques often ignore objects' interior structures; hence do not capture the intricacies of complex objects in images. In contrast, deformation-based shape representations (based on elastic deformations or fluid flows) focus on highly detailed shape information from images [36, 149]. With the underlying assumption that objects in many generic classes can be described as deformed versions of an ideal template, descriptors in this class arise naturally by matching the template to an input image. This procedure is also known as *atlas building* [85, 181, 207]. The resulting transformation is then considered a shape that reflects geometric changes. We will feature deformation-based shape representations that offer more flexibility in describing shape changes and variability of complex structures. However, our developed framework can be easily adapted to other types of representations, including those characterized by landmarks, binary segmentations, curves, and surfaces.

Inspired by the advantages of incorporating shape information in image analysis tasks, current deep learning-based classification networks have been mostly successful in using pre-extracted shapes from images [199, 121, 16]. However, these methods require preprocessed shape data and oftentimes achieve a suboptimal solution in identifying shape features that are most representative to differentiate different classes of images. An explicit learning of deformable shapes in deep image classifiers has been missing. This limits the power of classification models where quantifying and analyzing geometric shapes is critical.

In this work, we introduce a novel deep learning image classification model, named as Geo-SIC, that jointly learns deformable shapes in a multi-template deformation space. More specifically, Geo-SIC provides an unsupervised learning of deformation-based shape representations via a newly designed sub-network of atlas building. Different from previous deep learning based atlas building approaches [42, 73], we employ an efficient parameterization of deformations in a low-dimensional Fourier space [204] to speed up the training inference. The major contribution of Geo-SIC is three folds:

(i) In contrast to previous approaches treating shapes as preprocessed objects from images, Geo-SIC provides a more fundamental approach by merging shape features nat-

urally in the learning process of classification. To the best of our knowledge, Geo-SIC was the first to learn deformation-based shape descriptors within an image classifier. It provides an image distance function of both intensity and geometric changes that are most relevant to classify different groups.

(ii) Geo-SIC performs a simultaneous feature extraction from both image and learned shape spaces. With these integrated features, Geo-SIC achieves an improved accuracy and robustness of image classification. An additional benefit of Geo-SIC is increased model interpretability because of its access to the underlying geometric features of image data.

(iii) Geo-SIC provides an efficient geometric learning network via atlas building in a compact and low-dimensional shape space. This reduces the computational complexity of model training in atlas building, especially for high-dimensional image data (i.e., 3D brain MRIs).

We demonstrate the effectiveness of Geo-SIC on both synthetic 2D images and real 3D brain MR images. Experimental results show that our model substantially improves the classification accuracy compared to a wide variety of models without jointly learned geometric features. We then visualize the class activation maps by using gradient-weighted class activation mapping (Grad-CAM) [153]. The highlighted regions show that Geo-SIC attracts more attention to geometric shape features that positively contribute to the accuracy of classifiers.

2.2 Method: Geo-SIC

In this section, we present a novel deep image classifier (Geo-SIC) that explicitly learns geometric shape representations for an improved performance of accuracy, as well as increased model interpretability. Geo-SIC consists of two modules: an unsupervised learning of geometric shapes via an atlas building network, and a boosted classification network that integrates features from both images and learned shape spaces. Details of our network architecture are introduced as follows.

Geometric shape learning based on an atlas building network. Let (θ_g^E, θ_g^D) be the parameters of an encoder-decoder in our geometric learning network. Consider a number of J image classes, there exists a number of $N_j, j \in \{1, \ldots, J\}$ images in each class. Our atlas building network will learn the shape representations, also known as initial velocity fields $\widetilde{v}_n(\theta_g^E, \theta_g^D), n \in \{1, \cdots, N_j\}$, with an updated atlas I_j. We adopt the architecture of UNet [147] in this work, however, other network structures such as UNet++ [210] and TransUNet [33] can be easily applied.

Boosted image classification network. Let θ_c^E be the parameters of an encoder that extracts features from image spaces. We develop a feature fusion module that integrates geometric shape and image features in a latent space parameterized by $\theta_{gc}(\theta_g^E, \theta_c^E)$. This boosted classification network will predict a class label $y_{nj}(\theta_{gc})$ for each input image.

2.2.1 Network loss. The loss function of Geo-SIC includes the loss from both geometric learning and classification network. Given a set of image class labels $\widehat{y}_{nj}$, we define $\Theta = (\theta_{\mathrm{g}}^{\mathrm{E}}, \theta_{\mathrm{g}}^{\mathrm{D}}, \theta_{\mathrm{c}}^{\mathrm{E}}, \theta_{\mathrm{gc}})$ for all network parameters and formulate the total loss of Geo-SIC as

$$l(\Theta) = \sum_{n=1}^{N_j} \sum_{j=1}^{J} [\frac{1}{\sigma_j^2} \| I_j \circ \phi_{nj}^{-1} \left(\widetilde{v}_{nj}(\theta_g^E, \theta_g^D) \right) - I_{nj} \|^2 + (\widetilde{\mathcal{L}}_j \widetilde{v}_{nj}(\theta_g^E, \theta_g^D), \widetilde{v}_{nj}(\theta_g^E, \theta_g^D))$$
$$- \lambda \widehat{y}_{nj} \cdot \log y_{nj}(\theta_{gc})] + \mathrm{reg}(\Theta), \quad \text{s.t.} \quad \text{Eq (3)\& Eq (5)}, \tag{45}$$

where $\mathrm{reg}(\cdot)$ is a regularity term on the network parameters. Geo-SIC conducts a joint training for a geometric learning network, an image classifier backbone plus a feature fusion module. The entire model is optimized by the total loss Eq (45) with geodesic constraints. Testing can be efficiently performed by using the boosted classifier, which includes the well-trained boosted backbone with a feature fusion module.

An overview of our proposed Geo-SIC network is shown in Fig. 22.

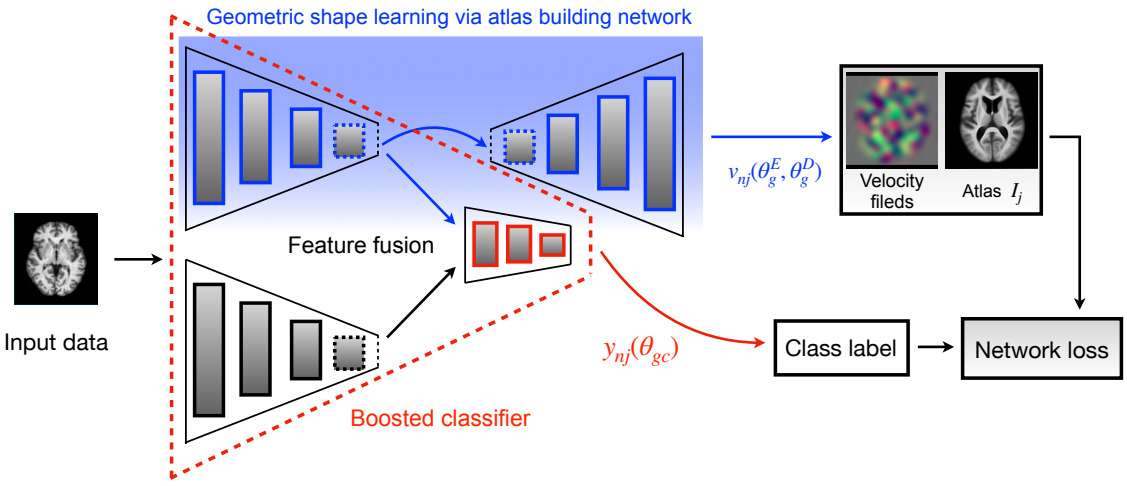


Fig. 22: An overview of our proposed Geo-SIC network.

While Geo-SIC enables the neural network to capture the geometric shape variation in the tangent space of diffeomorphisms in Fourier domain, other deformable-based shape approaches, such as image registration using piecewise functions [148], variants of LDDMM in spatial domain [161], and stationary velocity fields [6] can be easily plugged

into our proposed method. Now we are ready to introduce the network architecture of Geo-SIC. representations then present the joint learning model for image classification.

We optimize the network loss by an alternating optimization scheme, where the original loss is effectively decomposed into two subfunctions. Details of our developed network architecture of Geo-SIC and its optimization will be introduced in the following sections.

2.2.2 Alternating Optimization For Network Training We develop an alternating optimization scheme [126] to minimize the network loss defined in Eq (45). All parameters are optimized jointly by alternating between the training of geometric shape learning and image classification networks.

Training of geometric shape learning via atlas building. In contrast to current approaches that parameterize deformation-based shapes in a high-dimensional image space [73, 42], our model employs an efficient reparameterization in a low-dimensional bandlimited space [204]. This makes the computation of geodesic constraints (Eq (3) and Eq (5)) required in the loss function substantially faster in each forward/backward propagation during the training process. The loss of training in the sub-module of atlas building network is

$$l_{\text{Geo}}(\theta_g^E, \theta_g^D, I_j) = \sum_{n=1}^{N_j}\sum_{j=1}^{J}[\frac{1}{\sigma_j^2}\|I_j \circ \phi_{nj}^{-1}\left(\widetilde{v}_{nj}(\theta_g^E, \theta_g^D)\right) - I_{nj}\|^2 + (\widetilde{\mathcal{L}}_j\widetilde{v}_{nj}(\theta_g^E, \theta_g^D), \widetilde{v}_{nj}(\theta_g^E, \theta_g^D)) + \text{reg}(\theta_g^E, \theta_g^D), \quad s.t. \quad Eq\ (3) \& Eq\ (5), \tag{46}$$

Similar to [73, 42], we treat atlas I_j as a network parameter and update it accordingly. To guarantee the network optimization stays in the tangent space of diffeomorphisms, we pull back the network gradient with regard to initial velocity fields by backward integrating adjoint jacobi fields [204] each time after the forward pass. More details are included in the supplementary materials.

Training of boosted image classifier. As highlighted in the red box in Fig. 22, Geo-SIC extracts both images and geometric features and then integrates them into a feature fusion block. This boosted classifier is optimized over the loss defined as

$$l_{\text{SIC}}(\theta_{gc}) = -\lambda\sum_{n=1}^{N_j}\sum_{j=1}^{J}\widehat{y}_{nj}\cdot\log y_{nj}\left(\theta_{gc}(\theta_g^E, \theta_c^E)\right) + \text{reg}\left(\theta_{gc}(\theta_g^E, \theta_c^E)\right). \tag{47}$$

A summary of our joint learning of Geo-SIC through an alternating optimization is in Alg. 3.

Algorithm 3 Joint learning of Geo-SIC.

Input : A set of images $\{I_{nj}\}$ with class labels $\{\widehat{y}_{nj}\}$ and a number of iterations r.

Output: Predicted class label, atlas, and initial velocity fields.

for $i = 1$ to r **do**

`/* Train geometric learning network */`

Minimize the atlas building loss in Eq (46);

Output the atlases $\{I_j\}$ and predicted initial velocity fields $\{v_{nj}\}$ for all image classes;

`/* Train the boosted image classifier */`

With learned shape features, Minimizing the boosted classification loss in Eq (47);

Output the predicted class labels $\{\widehat{y}_{nj}\}$;

end

Until convergence

2.3 Experimental Evaluation

2D data set. We choose 50000 images (including five classes, circle, cloud, envelope, square, and triangle are shown in Fig 24) of Google Quickdraw dataset [83], a collection of categorized drawings contributed by online players in a drawing game. We run affine transformation within each class as prepossessing and upsample each image from 28×28 to 100×100.

3D brain MRI. For brain data, we include 373 public T1-weighted brain MRI scans from the OASIS dataset [56]. All 150 subjects are aged from 60 to 96 with Alzheimer‘‘s disease diagnosis (79 cases for dementia and 71 cases for non-demented). All MRIs were all pre-processed as $256 \times 256 \times 256$, $1.25mm^3$ isotropic voxels, and underwent skull-stripped, intensity normalized, bias field corrected, and pre-aligned with affine transformation.

Classification evaluation. We demonstrate the effectiveness of our model on both 2D synthetic data and 3D brain MRI scans. We select four classification backbones (AlexNet [92], a five-block 3D CNN, ResNet18 [71], and VGG19 [157]) as baseline methods. For CNN, we use a 3D convolutional layer with a $5 \times 5 \times 5$ convolutional kernel size, a batch normalization (BN) layer with activation functions (PReLU or ReLU), and a $2 \times 2 \times 2$ 3D max-pooling in each CNN block. For a fair comparison, we show the results of Geo-SIC by replacing the backbone in our model with all baselines (named as Geo-SIC:Alex, Geo-SIC:CNN, Geo-SIC:Res, and Geo-SIC:VGG).

To further investigate the advantages of our joint learning, we compare with two-step approaches (Two-step Alex, Two-step CNN, Two-step Res, and Two-step VGG), where the geometric learning network is treated as a preprocessing step for geometric feature ex-

traction. We report the average accuracy, F1-score, AUC, sensitivity, and specificity for all methods. We also show the receiver operating characteristic (ROC) curves by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

Robustness and interpretability. We demonstrate the robustness of Geo-SIC to variations in image intensity by adding different scales of universal adversarial noises [120, 119] in both 2D and 3D images. We adopt an iterative algorithm [119] that computes the universal perturbations to send perturbed images outside of the decision boundary of the classifier while fooling most images without visibly changing image data. We then compare the image classification accuracy for the baseline (selected from the best performance in backbones) and our model.

To better understand the model interpretability in terms of network attention of all models, we visualize the Grad-CAM [153] of the last neural network layer.

Atlas evaluation. We also evaluate the performance of our newly designed atlas building network, which serves as a core part of Geo-SIC. We compare the estimated atlas of Geo-SIC that achieves the highest AUC with three atlas-building methods: a diffeomorphic autoencoder [73] (LagoMorph), a deep learning based conditional template estimation method [42] (Con-Temp) and a Bayesian atlas building framework with hyper priors [181] (Hier-Baye).

We report the total computational time and memory consumption on 3D real brain images. To measure the sharpness of estimated atlas I, we adopt a metric of normalized standard deviation computed from randomly selected 4000 image patches [98]. Given $M(i)$, a patch around a voxel i of an atlas I, the local measure of the sharpness at voxel i is defined as $\text{sharpness}(I(i)) = \text{sd}_{M(i)}(I)/\text{avg}_{M(i)}(I)$, where sd and avg denote the standard deviation and the mean of M_i.

Parameter Setting. We set an optimal dimension of the low-dimensional shape representation as 16^2 for 2D dataset and 32^3 for 3D dataset. We set parameter $\alpha = 3$ for the operator $\widetilde{\mathcal{L}}$, the number of time steps for Euler integration in EPDiff (Eq (5)) as 10. We set the noise variance $\sigma = 0.02$. We set the batch size as 16 and use the cosine annealing learning rate schedule that starts from a learning rate $\eta = 1e-3$ for network training. We run 1000 epochs with the Adam optimizer and save networks with the best validation performance for all models. All networks are trained with an i7, 9700K CPU with 32 GB internal memory. The training and prediction procedure of all learning-based methods are performed on four Nvidia GTX 2080Ti GPUs. For both 2D and 3D datasets, we split the

images by using 70% as training images, 15% as validation images, and 15% as testing images.

Classification on 2D synthetic images Table. 2 reports the classification performance across all methods over model accuracy and five micro-averaged evaluation metrics. Our method achieves the highest model accuracy with comparable micro-averaged AUC, F-1 score, precision, sensitivity, and specificity. Compare to all baselines including two-step approaches, Geo-SIC achieves the best classification performance.

Table 2: Classification performance comparison on 2D synthetic data over six metrics (micro-average).

Models	Accuracy	AUC	F-1 score	Precision	Sensitivity	Specificity
AlexNet	0.880	0.931	0.876	0.914	0.880	0.920
Two-step AlexNet	0.891	0.955	0.891	0.906	0.891	0.905
Geo-SIC: Alex	**0.928**	**0.977**	**0.927**	**0.950**	**0.928**	**0.952**
CNN	0.861	0.928	0.857	0.874	0.861	0.881
Two-step CNN	0.869	0.931	0.866	**0.914**	0.869	**0.918**
Geo-SIC: CNN	**0.897**	**0.954**	**0.898**	0.911	**0.897**	0.910
ResNet18	0.875	0.941	0.877	0.885	0.875	0.883
Two-step Res	0.911	0.972	0.912	0.935	0.911	0.883
Geo-SIC: Res	**0.935**	**0.983**	**0.928**	**0.948**	**0.935**	**0.956**
VGG19	0.883	0.946	0.882	0.904	0.883	0.905
Two-step VGG	0.895	0.951	0.888	0.909	0.895	0.919
Geo-SIC: VGG	**0.927**	**0.980**	**0.924**	**0.957**	**0.927**	**0.960**

Fig. 23 visualizes the micro-averages ROC curve for baselines and our proposed Geo-SIC. For four sets of comparisons, all curves produced by our classifiers are with a larger AUC and are closer to the top-left corner, which indicates a better classification performance.

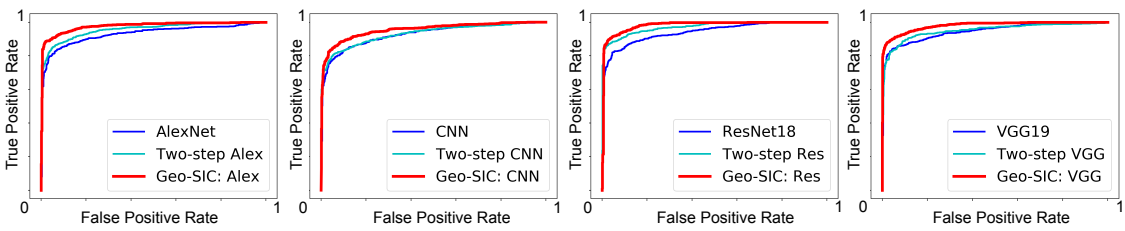


Fig. 23: ROC curves of multi-class classification comparison between baselines and the proposed method.

Fig. 24 (left) displays the heat maps overlaid with 2D QuickDraw data for all methods. Geo-SIC produces more explainable heat maps that are geometrically aligned with the original shapes. It indicates that the latent shape space attracts more attention to geometric features that positively contribute to the performance of classifiers. Fig. 24 (right) visualizes a comparison of atlases generated by baselines and Geo-SIC. It shows that Geo-SIC produces image atlases with the best visual quality, e.g., clearer circle/cloud edges, and sharper triangle/square corners.

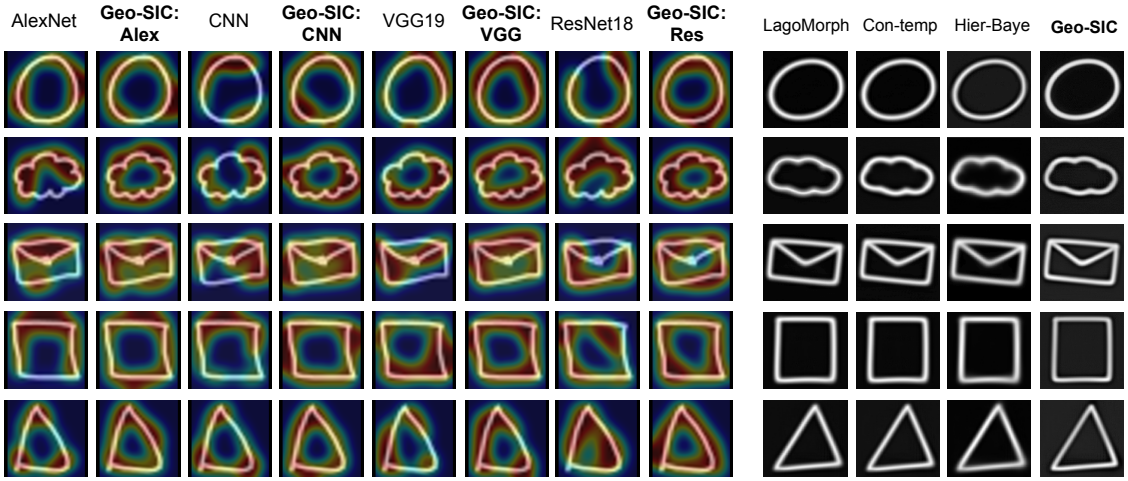


Fig. 24: Left: visualization of Grad-CAMs on mutil-atlas building task of five different geometric shapes.; Right: atlas comparison between state-of-the-arts and Geo-SIC.

Classification on 3D real brain MRIs Table. 3 reports the model performance for brain images. Our method Geo-SIC outperforms all baselines with comparable accuracy, AUC, F1-score, precision, sensitivity, and specificity. It shows that Geo-SIC substantially improves the classification performance with a reduced misclassification rate.

Fig. 25 visualizes the ROC curves for all algorithms. For four sets of comparisons, our boosted classifiers offer curves closer to the top-left corner with higher AUC values. It shows that Geo-SIC has better performance in distinguishing between healthy control and Alzheimer's disease groups.

Robustness and interpretability Fig. 26 shows that Geo-SIC consistently achieves better classification accuracy ($\sim 10\%$ higher in average) than baseline algorithms (i.e., VGG19) across different levels of adversarial attacks (i.e., $\epsilon = 5e{-}3, 5e{-}2, 5e{-}1$) on image intensity. This indicates that our model is able to improve the classification model robustness by providing jointly learned geometric features.

Table 3: Classification performance comparison on 3D brain images over six metrics.

Models	Accuracy	AUC	F-1 score	Precision	Sensitivity	Specificity
AlexNet	0.791	0.861	0.796	0.787	0.806	0.775
Two-step AlexNet	0.800	0.887	0.809	0.822	0.796	0.804
Geo-SIC: Alex	**0.835**	**0.917**	**0.827**	**0.823**	**0.831**	**0.839**
CNN	0.779	0.860	0.773	0.804	0.744	0.814
Two-step CNN	0.788	0.883	0.791	0.794	0.787	0.789
Geo-SIC: CNN	**0.824**	**0.902**	**0.828**	**0.833**	**0.822**	**0.825**
ResNet18	0.805	0.887	0.811	0.813	0.809	0.800
Two-step Res	0.833	0.917	0.836	0.804	0.870	0.797
Geo-SIC: Res	**0.873**	**0.933**	**0.874**	**0.874**	**0.874**	**0.872**
VGG19	0.805	0.871	0.813	0.807	0.818	0.790
Two-step VGG	**0.865**	0.880	0.838	**0.822**	0.854	**0.826**
Geo-SIC: VGG	0.852	**0.930**	**0.850**	0.820	**0.881**	0.825

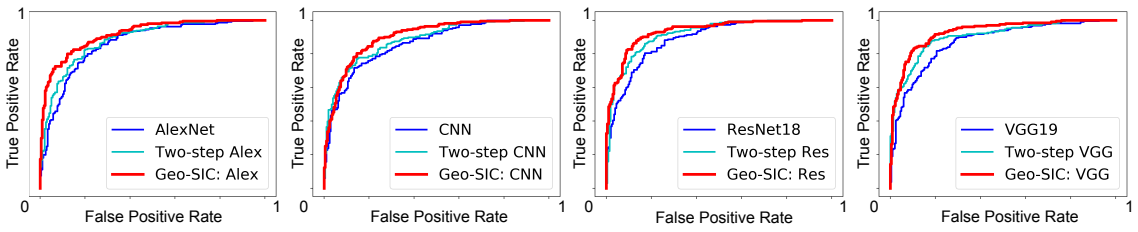


Fig. 25: ROC curves of binary classification comparison for baselines, two-step approaches, and Geo-SIC.

Fig. 27 displays the heat maps (produced by Grad-CAM) overlaid with brain MRIs. Our generated heat maps are fairly aligned with brain structures (e.g., ventricle), which are critical to Alzheimer's disease diagnosis. Our model highlights how latent geometric features guide neural networks to identify the most anatomically meaningful brain regions for distinguishing between healthy and disease groups. Specifically, the heat maps generated by Geo-SIC for both dementia and non-dementia cases are more explainable than those produced by other models. This increased interpretability demonstrates the potential for our model to improve the analysis of Alzheimer's disease in a clinical setting.

Atlas of 3D images Fig. 28 (top) visualizes a comparison of the atlas on real brain MRI scans. With the benefits of reparameterizing the deformation fields in a low-dimensional bandlimited space [204], our model obtains better quality of the atlas with sharper details. More specifically, Geo-SIC offers a better brain atlas with clearer anatomical structures, e.g., ventricle, grey and white matter.

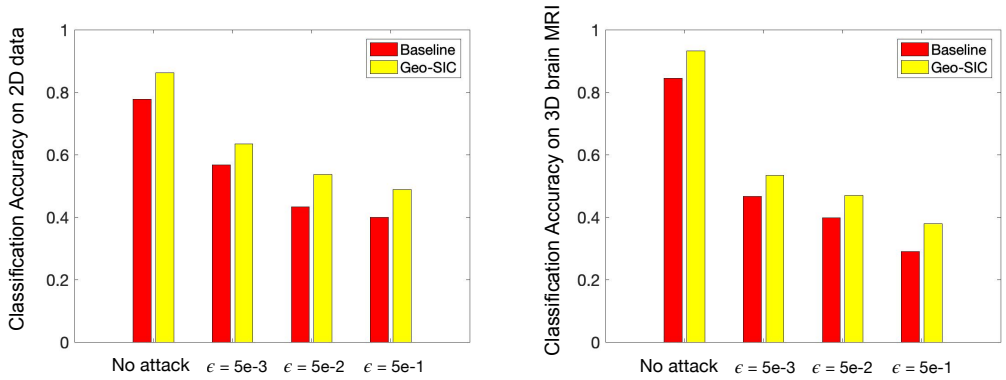


Fig. 26: Classification accuracy comparison between baseline and Geo-SIC under different scales of adversarial noise attack for 2D and 3D image classification.

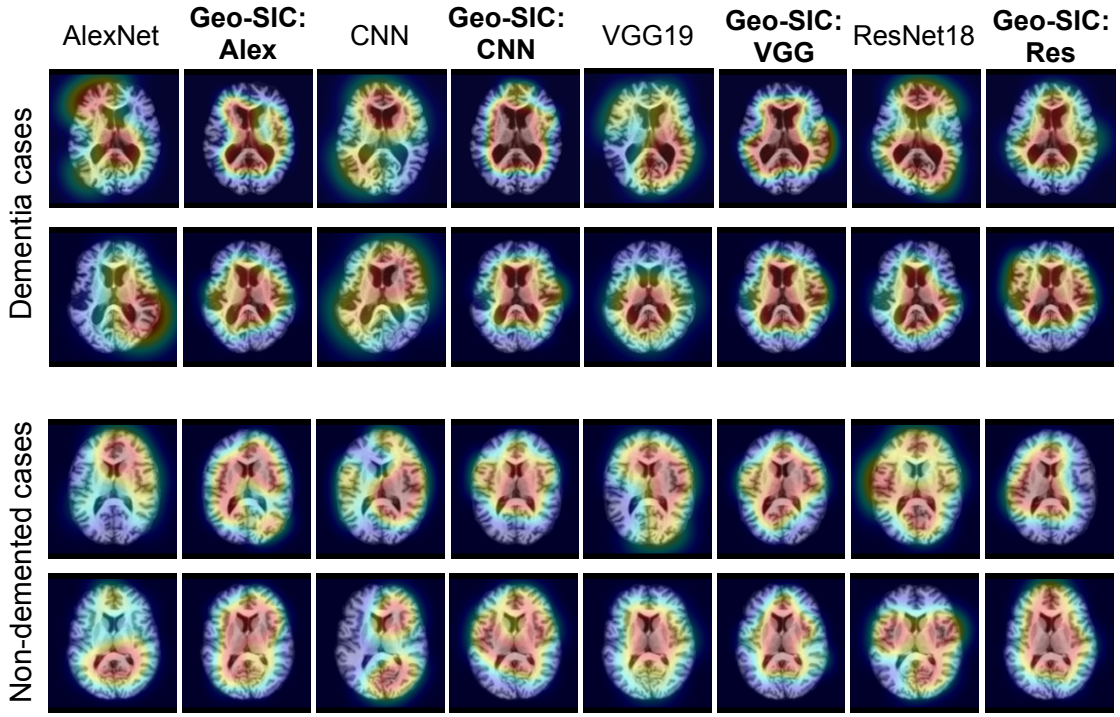


Fig. 27: Visualization of Grad-CAMs on single-atlas building of dementia and non-dementia. Left to right: Grad-CAM heatmaps generated by AlexNet, Geo-SIC:Alex, CNN, Geo-SIC:CNN, ResNet18, Geo-SIC:Res, VGG19 and Geo-SIC:VGG.

Fig. 28 (bottom left) quantitatively reports the sharpness metric of all methods. Fig. 28 (bottom right) shows the comparison of computational time and memory consumption across all methods. Although Con-Temp is slightly faster than Geo-SCI due to a very different parameterization of velocity fields (stationary velocity rather than the time-

dependent velocity in other methods), it achieves a less sharp atlas and still requires larger memory consumption than our model. Compared with the other methods (Lagomorph and Hier-Baye) that employ time-dependent velocity fields, Geo-SIC substantially redu

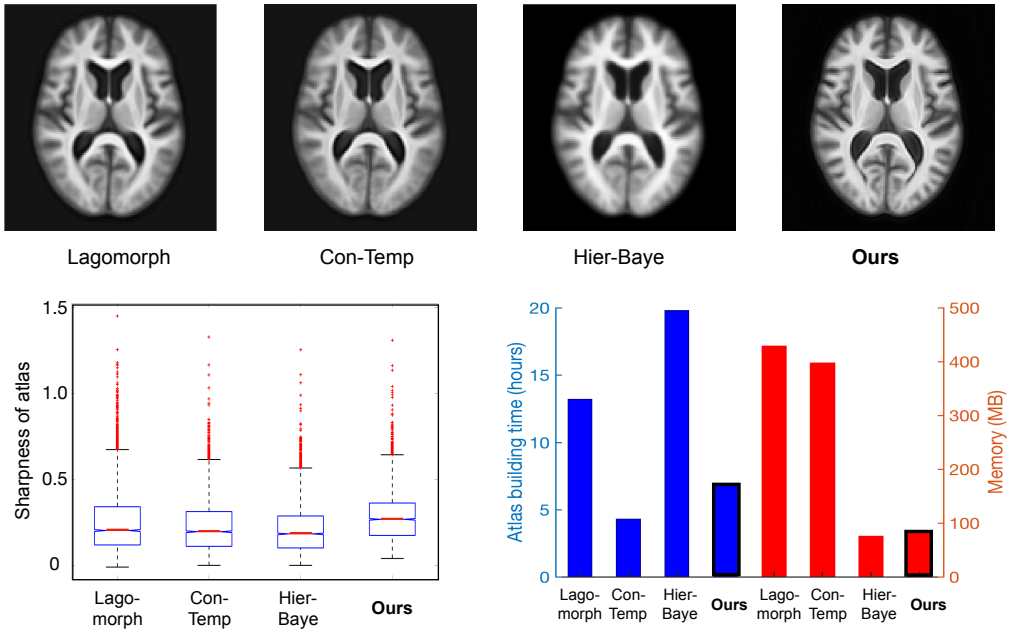


Fig. 28: Top: atlas comparison between state-of-the-arts and Geo-SIC. Bottom left: sharpness metric evaluation of atlas (the higher the better).The mean values of the sharpness metric of three baselines and **Geo-SIC** are, 0.259, 0.235, 0.218, and **0.2867**; bottom left: average time and memory consumption comparison for atlas building.

2.4 Conclusion

In this section, I present a novel deep learning model, named as Geo-SIC, that for the first time incorporates deformable geometric shape learning into deep image classifiers. We jointly learn a boosted classifier with an unsupervised shape learning network via atlas building. To achieve this goal, we define a new joint loss function with an alternating optimization scheme. An additional benefit is that Geo-SIC provides efficient shape representations in a low-dimensional bandlimited space. Experimental results on both 2D synthetic data and 3D brain MRI scans show that our model gains an improved classification performance while producing a sharper atlas with better visual quality. In addition, compared with the state-of-the-arts, our model is more explainable in terms of interpreting the network attention on geometric features. The theoretical tools developed in this

work are generic to a wide variety of combinations of shape representations and classification backbones. Geo-SIC not only has a great potential to impact clinically diagnostic routines, such as Alzheimer's disease detection, or post-treatment for patient care, but also bridges the gap between the developed deformable shape learning theories and classification-based applications. Future work to extend our Geo-SIC can be (i) modeling multiple templates within each class to capture multimodal image distributions, and (ii) incorporating images with missing data values that are caused by occlusions, or appearance changes such as tumor growth.

3 Deep Learning for Regularization Parameter Estimation

3.1 Related Work

The nice properties of diffeomorphisms keep topological structures of objects intact in images. Artifacts (i.e., tearing, folding, or crossing) that generate biologically meaningless images can be effectively avoided, especially when large deformation occurs. As we addressed in Chapter II, the problem of diffeomorphic image registration is typically formulated as an optimization over transformation fields, such as a free-form deformation using B-splines [148], a LogDemons algorithm based on stationary velocity fields (SVF) [6], and a large diffeomorphic deformation metric mapping (LDDMM) method utilizing time-varying velocity fields [15].

To ensure the smoothness of transformation fields, a regularization term defined on the tangent space of diffeomorphisms (called velocity fields) is often introduced in registration models. Having such a regularity with proper model parameters is critical to registration performance because they greatly affect the estimated transformations. Either too large or small-valued regularity can not achieve satisfying registration results (as shown in Fig .29). Models of handling the regularity parameter mainly include (i) direct optimizing a Bayesian model or treating it as a latent variable to integrate out via Expectation Maximization (EM) algorithm [2, 3, 207, 181], (ii) exhaustive search in the parameter space [81, 170, 150], and (iii) utilizing parameter continuation methods [64, 65, 111, 112]. Direct optimization approaches define a posterior of transformation fields that includes an image matching term as a likelihood and a regularization as a prior to support the smoothness of transformations [211, 3, 166]. Estimating regularization parameters of these models using direct optimization is not straightforward due to the complex structure of the

posterior distribution. Simpson et al. infer the level of regularization in small deformation registration model by mean-field VB inference [84], which allows tractable approximation of full Bayesian inference in a hierarchical probabilistic model [159, 158]. However, these aforementioned algorithms are heavily dependent on initializations, and are prone to getting stuck in the local minima of high-dimensional and non-linear functions in the transformation space. A stochastic approximative expectation maximization (SAEM) algorithm [3] was developed to marginalize over the posterior distribution of unknown parameters using a Markov Chain Monte Carlo (MCMC) sampling method. Later, Zhang et al. estimate the model parameters of regularization via a Monte Carlo Expectation Maximization (MCEM) algorithm for unbiased atlas building problem [207]. A recent model of Hierarchical Bayesian registration [181] further characterizes the regularization parameters as latent variables generated from Gamma distribution, and integrates them out by an MCEM method.

Despite the achievement of the aforementioned methods, estimating the regularization parameter in a high-dimensional and nonlinear space of 3D MRIs (i.e., dimension is typically 128^3 or higher) inevitably leads to expensive computational cost through iterative optimizations. To address this issue, we present a deep learning approach to fast predict registration parameters. While there exist learning-based registration models for transformations [91, 14, 17], we are particularly interested in learning the relationship between pairwise images and optimal regularizations of transformations via regression. In order to produce "ground truth" regularization parameters, we first introduce a low-dimensional Bayesian model of image registration to estimate the best regularity from the data itself. Following a recent work of [177], we construct a posterior distribution of diffeomorphic transformations entirely in a bandlimited space with much lower dimensions. This greatly reduces the computational cost of data generation in training. The theoretical tools devel-

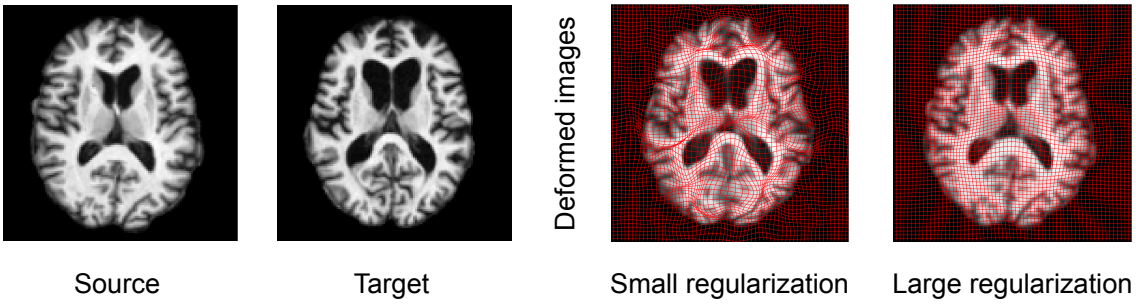


Fig. 29: Left to right: examples of transformation fields overlaid with deformed images with under-regularized and over-regularized registration models. A small regularization introduces crossing artifacts on the transformations vs. a large regularization discourages sufficient transformations between images.

oped in this work are generic to various deformable registration models, e.g, stationary velocity fields that remain constant over time [6]. The model recommends optimal registration parameters for registration in real-time and has great potential in clinical applications (i.e., image-guided navigation system for brain shift compensation during surgery [106]). To summarize, our main contributions are three folds:

- To the best of our knowledge, we are the first to present a predictive regularization estimation method for diffeomorphic image registration through deep learning.
- We develop a low-dimensional Bayesian framework in a bandlimited Fourier space to speed up the training data generation.
- Our model significantly speeds up the parameter estimation, while maintaining comparable registration results.

The section is organized as follows. In sec. 3.2.1, we first develop a low-dimensional posterior distribution that is parametrized by bandlimited velocity fields. We then estimate the regularization parameter by maximizing the posterior. In sec. 3.2.2, we design a deep convolutional neural network that takes an image pair as input and adaptively predicts the optimal smoothness level for registration. In sec. 3.3, we validate our model on both 2D synthetic data and 3D brain MRI scans.

3.2 Method: Efficient Predictive Approach for Regularization Parameter Estimation

In this section, we present a supervised learning model based on CNN to predict the regularity of image registration for a given image pair. Analogous to [194], we run optimization-based image registration to obtain training data. We introduce an efficient Bayesian model of image registration to produce appropriate regularization parameters for training.

3.2.1 Low-dimensional Bayesian Model of Registration In contrast to previous approaches, our proposed model is parameterized in a bandlimited velocity space $\widetilde{V}$, with parameter α enforcing the smoothness of transformations.

Assuming an independent and identically distributed (i.i.d.) Gaussian noise on image intensities, we obtain the likelihood

$$p(T \mid S \circ \phi_1^{-1}, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma^2)^M} \exp\left(-\frac{1}{2\sigma^2}||S \circ \phi_1^{-1} - T||_2^2\right), \tag{48}$$

where σ^2 is the noise variance and M is the number of image voxels. The deformation ϕ_1^{-1} corresponds to $\widetilde{\phi}_1^{-1}$ in Fourier space via the Fourier transform $\mathcal{F}(\phi_1^{-1}) = \widetilde{\phi}_1^{-1}$, or its inverse $\phi_1^{-1} = \mathcal{F}^{-1}(\widetilde{\phi}_1^{-1})$. The likelihood is defined by the residual error between a target image and a deformed source at time point $t = 1$. We assume the definition of this distribution is after the fact that the transformation field is observed through geodesic shooting; hence is not dependent on the regularization parameters.

Analogous to [176], we define a prior on the initial velocity field $\widetilde{v}_0$ as a complex multivariate Gaussian distribution, i.e.,

$$p(\widetilde{v}_0|\alpha) = \frac{1}{(2\pi)^{\frac{M}{2}}|\widetilde{\mathcal{L}}^{-1}(\alpha)|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\widetilde{\mathcal{L}}(\alpha)\widetilde{v}_0, \widetilde{v}_0)\right), \tag{49}$$

where $|\cdot|$ is matrix determinant. The Fourier coefficients of $\widetilde{\mathcal{L}}$ is, i.e., $\widetilde{\mathcal{L}} = \left(\alpha\widetilde{\mathcal{A}} + 1\right)^3$, $\widetilde{\mathcal{A}}(\xi_1, \ldots, \xi_d) = -2\sum_{j=1}^{d}(\cos(2\pi\xi_j) - 1)$. Here $\widetilde{\mathcal{A}}$ denotes a negative discrete Fourier Laplacian operator with a d-dimensional frequency vector $(\xi_1, \ldots, \xi_d)$, where d is the dimension of the bandlimited Fourier space.

Combining the likelihood in Eq (48) and prior in Eq (49) together, we obtain the negative log posterior distribution on the deformation parameter parameterized by $\widetilde{v}_0$ as

$$-\ln\, p(\widetilde{v}_0 \,|\, S, T, \sigma^2, \alpha) = \frac{1}{2}(\widetilde{\mathcal{L}}\widetilde{v}_0, \widetilde{v}_0) + \frac{\|S \circ \phi_1^{-1} - T\|_2^2}{2\sigma^2} - \frac{1}{2}\ln|\widetilde{\mathcal{L}}| + 2M\ln\sigma + M\ln(2\pi). \tag{50}$$

Next, we optimize Eq (50) over the regularization parameter α and the registration parameter $\widetilde{v}_0$ by maximum a posterior (MAP) estimation using gradient descent algorithm. Other optimization schemes, such as BFGS [134], or the Gauss-Newton method [8] can also be applied.

Gradient of $\boldsymbol{\alpha}$. To simplify the notation, first we define $f(\widetilde{v}_0) \triangleq -\ln\, p(\widetilde{v}_0 \,|\, S, T, \sigma^2, \alpha)$. Since the discrete Laplacian operator $\widetilde{\mathcal{L}}$ is a diagonal matrix in Fourier space, its determinant can be computed as $\prod_{j=1}^{d}(\alpha\widetilde{\mathcal{A}}_j + 1)^3$. Therefore, the log determinant of $\widetilde{\mathcal{L}}$ operator is

$$\ln|\widetilde{\mathcal{L}}| = 3\sum_{j=1}^{d}(\alpha\widetilde{\mathcal{A}}_j + 1).$$

We then derive the gradient term $\nabla_\alpha f(\widetilde{v}_0)$ as

$$\nabla_\alpha f(\widetilde{v}_0) = -\frac{3}{2}[\sum_{j=1}^{d}\frac{\widetilde{\mathcal{A}}_j}{\alpha\widetilde{\mathcal{A}}_j + 1} - \langle(\alpha\widetilde{\mathcal{A}} + 1)^5\widetilde{\mathcal{A}}\widetilde{v}_0, \widetilde{v}_0\rangle]. \tag{51}$$

Gradient of $\widetilde{v}_0$. We compute the gradient with respect to $\widetilde{v}_0$ by using a forward-backward sweep approach developed in [206]. Steps for obtaining the gradient $\nabla_{\widetilde{v}_0} f(\widetilde{v}_0)$ are as follows:

(i) Forward integrating the geodesic shooting equation Eq (5) to compute $\widetilde{v}_1$,

(ii) Compute the gradient of the energy function Eq (1) with respect to $\widetilde{v}_1$ at $t = 1$,

$$\nabla_{\widetilde{v}_1} f(\widetilde{v}_0) = \widetilde{\mathcal{L}}^{-1}(\alpha) \left(\frac{1}{\sigma^2} (S \circ \phi_1^{-1} - T) \cdot \nabla (S \circ \phi_1^{-1}) \right). \tag{52}$$

(iii) Bring the gradient $\nabla_{\widetilde{v}_1} f(\widetilde{v}_0)$ back to $t = 0$ by integrating adjoint Jacobi fields backward in time [206],

$$\frac{d\widehat{v}}{dt} = -\mathrm{ad}^{\dagger}_{\widetilde{v}} \widehat{h}, \quad \frac{d\widehat{h}}{dt} = -\widehat{v} - \mathrm{ad}_{\widetilde{v}} \widehat{h} + \mathrm{ad}^{\dagger}_{\widehat{h}} \widetilde{v}, \tag{53}$$

where $\widehat{v} \in V$ are introduced adjoint variables with an initial condition $\widehat{h} = 0, \widehat{v} = \nabla_{\widetilde{v}_1} f(\widetilde{v}_0)$ at $t = 1$.

A summary of the optimization is in Alg. 4. It is worthy to mention that our Bayesian framework developed in the low-dimensional bandlimited space dramatically speeds up the computational time of generating training regularity parameters by approximately ten times comparing with high-dimensional frameworks.

3.2.2 Network Architecture Now we are ready to introduce our network architecture by using the estimated α and given image pairs as input training data. Fig. 30 shows an overview flowchart of our proposed learning model. With the optimal registration regularization parameter α^{opt} obtained from an image pair (as described in Sec. 3.2.1), a two-stream CNN-based regression network takes source images, target images as training data to produce a predictive regularization parameter. We optimize the network $\mathcal{H}(S, T; W)$ with the followed objective function,

$$E_{\text{loss}} = \text{Err}[\mathcal{H}(S, T; W), \alpha^{opt}] + \text{Reg}(W),$$

where $\text{Reg}(W)$ denotes the regularization on convolution kernel weights W. $\text{Err}[\cdot, \cdot]$ denotes the data fitting term between the ground truth and the network output. In our model, we use L_2 norm for both terms. Other network architectures, e.g. 3D Residual Networks (3D-ResNet) [67] and Very Deep Convolutional Networks (3D-VGGNet) [157, 191] can be easily applied as well.

In our network, we input the 3D source and target images into separate channels that include four convolutional blocks. Each 3D convolutional block is composed of a 3D

Algorithm 4 MAP of low-dimensional Bayesian registration model for training data generation.

Input : source image S and target image T, step size ϵ, step size τ, iterations r, algorithm stop criterion rate u, counter q, minimal convergence iteration q_{min}

Output: optimal smoothness level α^{opt}, and registration solution $\widetilde{v}_0$

`/* Low-dimensional MAP Estimation for` α `*/`

for $i = 1$ to r **do**

1. Compute gradient $\nabla_\alpha f(\widetilde{v}_0)$ by Eq (51);

`/* Update` α `when its gradient is greater than zero;*/`

2. **if** $\|\nabla_\alpha f(\widetilde{v}_0)\|^2 > 1e-6$ **then**

$\alpha^{opt} \leftarrow \alpha^{opt} - \tau \nabla_\alpha f(\widetilde{v}_0)$;

else

break;

3. Forward integrating the geodesic evolution equation in Eq (5) with initial velocity $\widetilde{v}_0$;

4. Compute gradient $\nabla_{\widetilde{v}_1} f(\widetilde{v}_0)$ by Eq (52) then backward integrating adjoint equations Eq (53) to generate $\nabla_{\widetilde{v}_0} f(\widetilde{v}_0)$ at $t = 0$;

`/* Update` $\widetilde{v}_0$ `when its gradient is greater than zero;*/`

5. **if** $\|\nabla_{\widetilde{v}_0} f(\widetilde{v}_0)\|^2 > 1e-6$ **then**

$\widetilde{v}_0 \leftarrow \widetilde{v}_0 - \epsilon \nabla_{\widetilde{v}_0} f(\widetilde{v}_0)$;

else

break;

`/* Compute algorithm stop rate */`

6. Compute total energy Eq (50) for $i-$th iteration as $\mathrm{Obj_i}$;

7. **if** $\frac{\mathrm{Obj_i} - \mathrm{Obj_{i-1}}}{\mathrm{Obj_i}} < u$ **then**

$q = q + 1$;

else

continue;

`/* Convergence check */`

8. **if** $q \geq q_{min}$ **then**

break ;

else

continue;

End

convolutional layer, a batch normalization (BN) layer with activation functions (PReLU or ReLU), and a 3D max-pooling layer. Specifically, we apply $5 \times 5 \times 5$ convolutional kernel and $2 \times 2 \times 2$ max-pooling layer to encode a batch (size as B) of source and target images (128^3) to feature maps (16^3). After extracting the deep features from source and target channels, we combine them into a fusion channel, which includes three convolutional

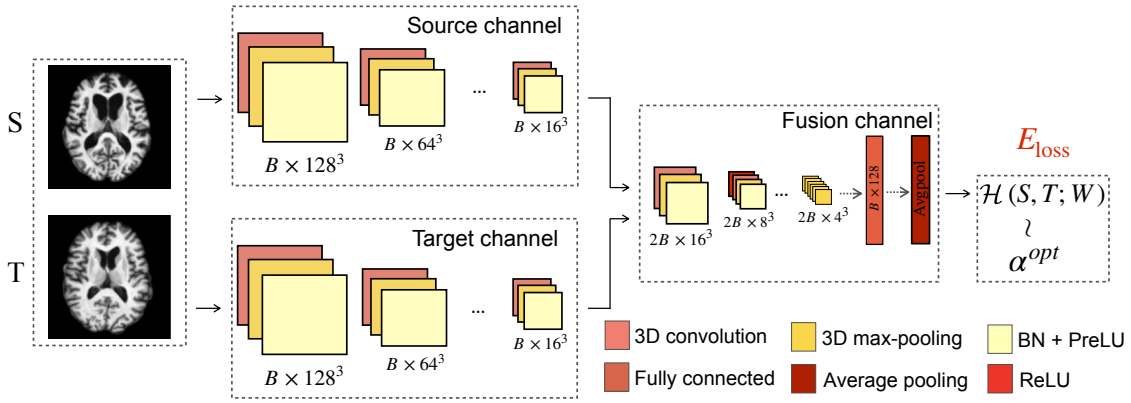


Fig. 30: Illustration of our proposed network. From left to right: training data includes pairwise images, a CNN-based neural network, and a loss computed between the network output and the optimal parameter α^{opt}, which is estimated from Alg. 4.

blocks, a fully connected layer, and an average pooling layer to produce the network output.

3.3 Experimental Evaluation

To demonstrate the effectiveness of the proposed low-dimensional Bayesian registration model, we validate it through three sets of experiments. For 2D synthetic data registration, we deform a binary source image with velocity fields sampled from the prior distribution in Eq (49) with known regularization parameters to simulate target images. We show three convergence graphs of our MAP estimation and compare them with the ground truth parameters.

Similarly, we synthesize 900 image pairs using regularization parameters at different scales respectively, i.e., $\alpha = \{0.1, 1.0, 10.0\}$, to test the performance of our predictive model. We then use the predicted parameter α to run registration model and show the error maps between target and deformed images.

For 3D brain MRI registration, we show results on both MAP and our network prediction. We first show the numerical difference between the MAP estimation and our prediction (i.e. predicted regularization parameter), and then report the mean error of deformed images between both methods across all datasets. We visualize the transformation grids and report the value of regularization parameters for both methods. To further investigate the accuracy of parameters generated by our model, we perform registration-based segmentation and examine the resulting segmentation accuracy over nine brain structures, in-

cluding cortex, putamen, cerebellum, caudate, gyrus, brain stem, precuneus, cuneus, and hippocampus. We evaluate a volume-overlapping similarity measurement, also known as Sørensen−Dice coefficient [46], between the propagated segmentation and the manual segmentation. The statistics of dice evaluation over 150 registration pairs are reported.

We last compare the computational efficiency on both time and memory consumption of the proposed method with a baseline model that performs Bayesian estimation of regularization parameter in the full spatial domain [207].

To generate the training data of initial velocities, we run the proposed low-dimensional Bayesian registration algorithms until convergence. We use the Euler integrator in geodesic shooting and set the of integration steps as 10. We set algorithm stop rate u as $1e-6$ and minimal convergence iteration q_{min} as 30. We use optimal truncated dimension for $\widetilde{v}_0$ as 16 and $\sigma = 0.03$ according to [206]. For the network setting, We initialize the convolution kernel weights using the He normal initializer [70] and use the Adam optimizer with a learning rate of $5e-4$ until convergence. We set 16 and $1.0e-4$ as batch size and weight decay. The maximum epoch for 2D and 3D network training is 1000.

2D synthetic data. We generate synthetic bull-eye images with the size of 100×100 (as shown in Fig. 31). We manipulate the width a and height b of two ellipses by using equation $\frac{(x-50)^2}{a^2} + \frac{(y-50)^2}{b^2} = 1$.

3D brain MRIs. We include 1500 public T1-weighted brain MRI scans from ADNI dataset [80], OASIS [56], and LPBA40 [156], among which 260 subjects have manual delineated segmentation labels. All 3D data were carefully pre-processed as $128\times128\times128$, $1.25mm^3$ isotropic voxels, and underwent skull-stripped, intensity normalized, bias field corrected, and pre-aligned with affine transformation.

For both 2D and 3D datasets, we split the images by using 70% as training images, 15% as validation images, and 15% as testing images such that no subjects are shared across the training, validation, and testing stage. We evaluate the hyperparameters of models and generate preliminary experiments on the validation dataset. The testing set is only used for computing the final results.

Results Fig. 31 displays our MAP estimation of registration results including appropriate regularity parameters on 2D synthetic images. The middle panel of Fig. 31 reports the convergence of α estimation vs. ground truth. It indicates that our low-dimensional Bayesian model provides trustworthy regularization parameters that are fairly close to ground truth for network training. The bottom panel of Fig. 31 shows the convergence graph of the total energy for our MAP approach.

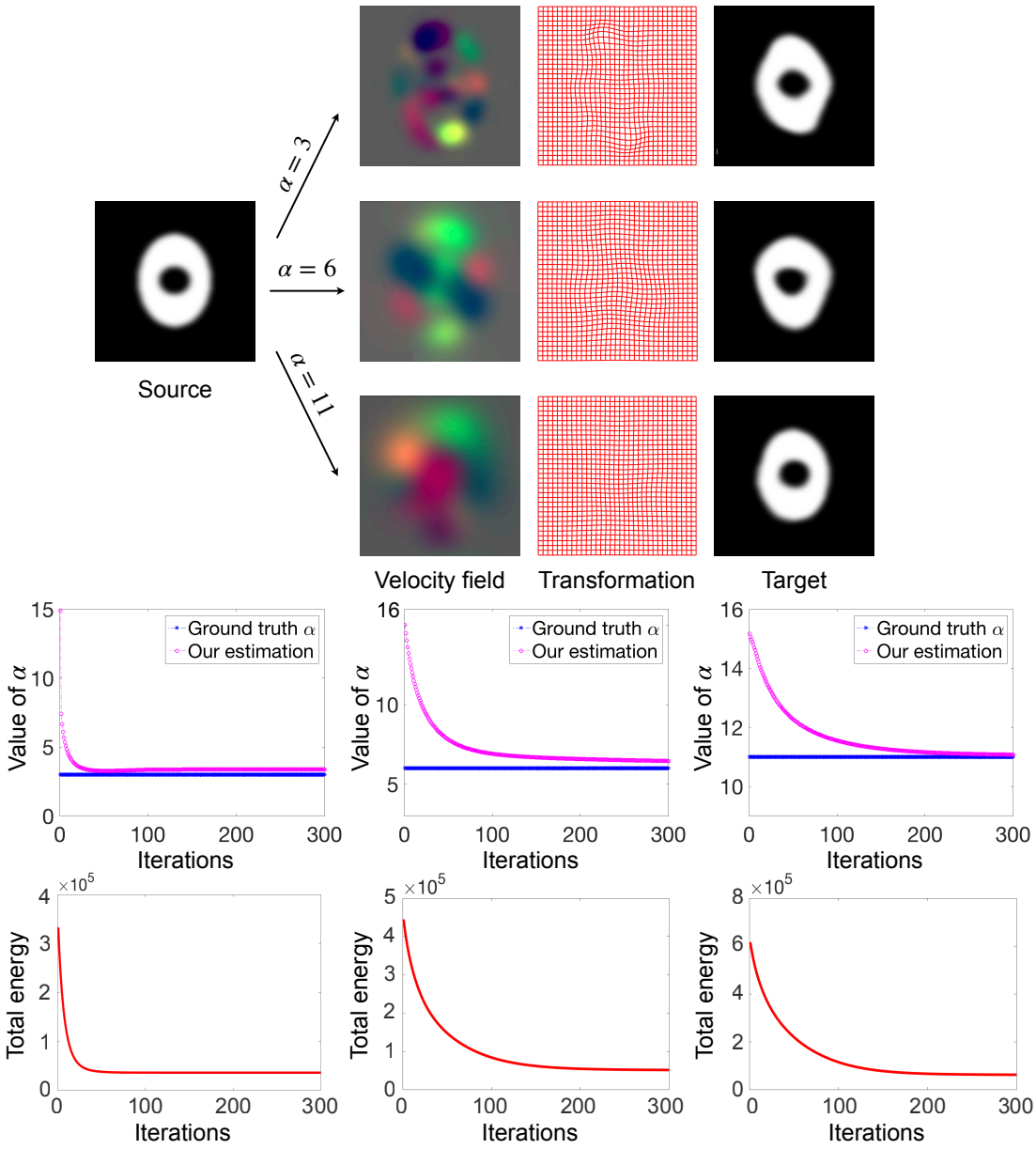


Fig. 31: Top panel: source image, velocity fields generated from prior distribution and transformation fields (with known regularization parameter $\alpha = 3, 6, 11$), and target images produced by deformed source images; Middle panel: convergence graphs of estimated α by MAP for training data; Bottom panel: convergence graphs of total energy Eq (50).

Fig. 32 further investigates the consistency of our network prediction. The left panel shows estimates of regularization parameter at multiple scales, i.e., $\alpha = 0.1, 1.0, 10.0$, over 900 2D synthetic image pairs respectively. The right panel shows the mean error of image differences between deformed source images by transformations with predicted α

and target images. While there are small variations on estimated regularization parameters, the registration results are very close (with averaged error at the level of 10^{-5}).

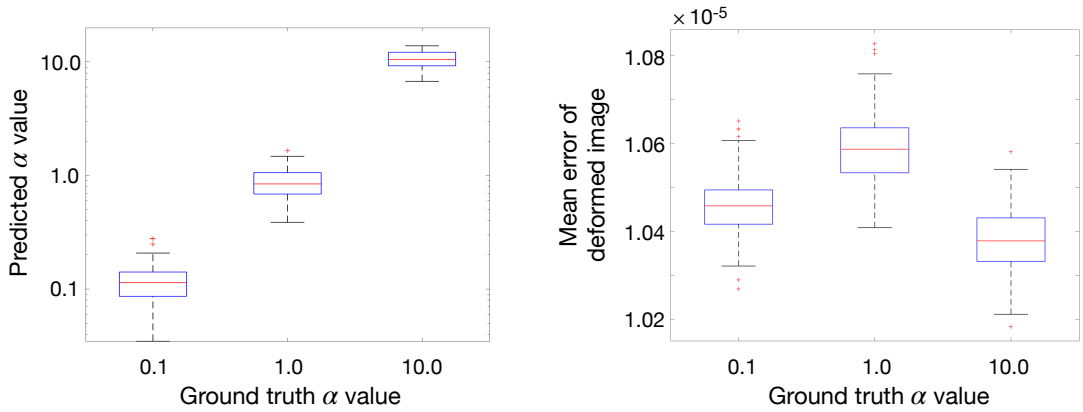


Fig. 32: Left: our prediction of regularization parameters over 900 image pairs synthesized with different ground truth parameters $\alpha = 0.1, 1.0, 10.0$; Right: error maps of image differences between deformed and target images.

Fig. 33 shows examples of 2D pairwise image registration with regularization estimated by MAP and our predictive deep learning model. We obtain the regularization parameter $\alpha = 11.34$ (MAP) vs. $\alpha = 13.20$ (network prediction), and $\alpha = 5.44$ (MAP) vs. $\alpha = 6.70$ (network prediction). The error map of deformed images indicates that both estimations obtain fairly close registration results.

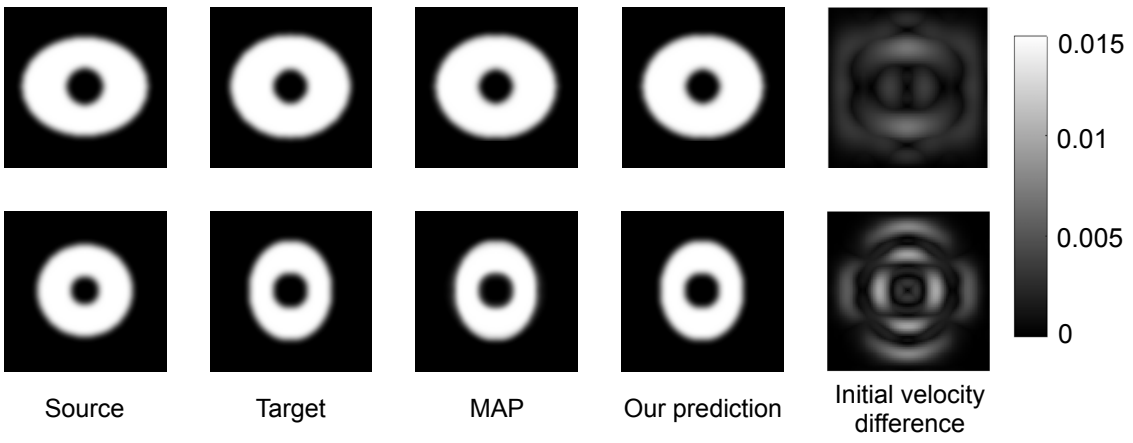


Fig. 33: Left to right: two examples of 2D synthetic source, target, deformed images by MAP and predictive deep learning method, and error maps of deformed image intensity differences.

Fig. 34 displays deformed 3D brain images overlaid with transformation grids for both methods. The parameter estimated from our model produces a comparable registration result. From our observation, a critical pattern between the optimal α^{opt} and its associate image pairs is that the value of α is relatively smaller when large deformation occurs. This is because the image matching term (encoded in the likelihood) requires a higher weight to deform the source image.

Fig. 35 investigates the consistency of our network prediction over three different datasets of 3D brain MRI. The left panel shows the absolute value of numerical differences between predicted regularization parameters and MAP estimations. The right panel shows the voxel-wise mean error of image differences between deformed images by transformations with predicted α and deformed images by MAP. While slight numerical difference on estimated regularization parameters exists, the 3D deformed images are fairly close (with averaged voxel-wise error at the level of 10^{-6}).

Fig. 36 visualizes three views of the deformed brain images (overlay with propagated segmentation label) that are registered by MAP and our prediction. Our method produces a registration solution, which is highly close to the one estimated by MAP. The propagated segmentation label fairly aligns with the target delineation for each anatomical structure. While we show different views of 2D slices of brains, all computations are carried out fully in 3D.

Fig. 37 reports the volume overlapping of nine anatomical structures for both methods, including Cor(cortex), Puta (putamen), Cere (cerebellum), Caud (caudate), gyrus, Stem (brain stem), Precun (precuneus), Cun (cuneus), and Hippo (hippocampus). Our method produces comparable dice scores comparing with MAP estimations. This indicates that the segmentation-based registration by using our estimation achieves comparable registration performance with little to no loss of accuracy.

Table. 4 quantitatively reports the averaged time and memory consumption of MAP estimation in full spatial image domain and our method. The proposed predictive model provides appropriate regularization parameters approximately 1000 times faster than the conventional optimization-based registration method with a much lower memory footprint.

3.4 Conclusion

In this section, I proposed a deep learning-based approach to model the relationship between the regularization of image registration and the input image data. We first developed

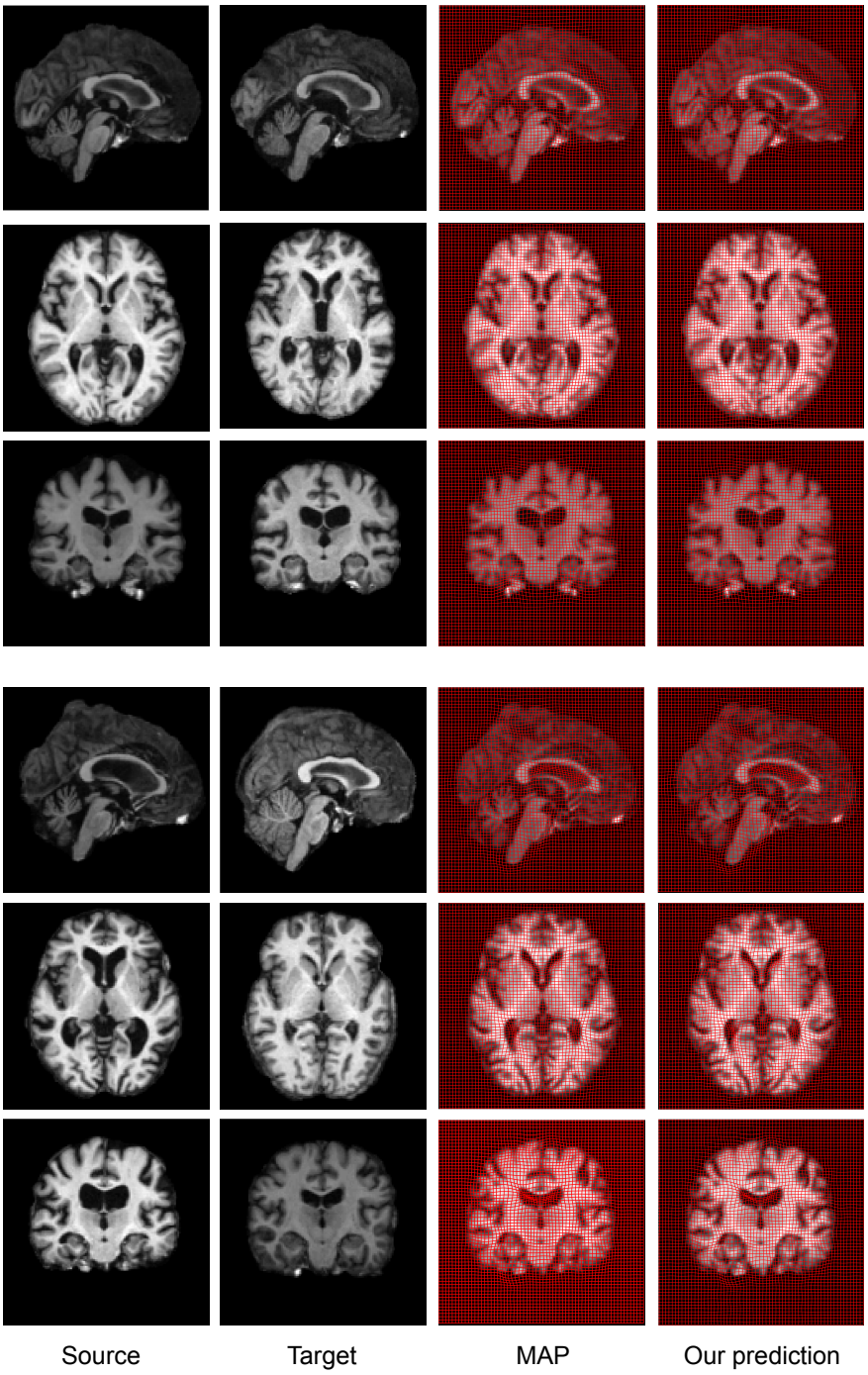


Fig. 34: Left to right: source, target, deformed images overlaid with the transformation grids generated by both methods. The optimal and predicted value of α for two showcases are $(8.90, 9.20)$ and $(3.44, 2.60)$.

a low-dimensional Bayesian model that defines image registration entirely in a bandlimited space. We then learned the mapping between regularization parameters and spatial images through a CNN-based neural network. To the best of our knowledge, we are the first to predict the optimal regularization parameter of diffeomorphic image registration

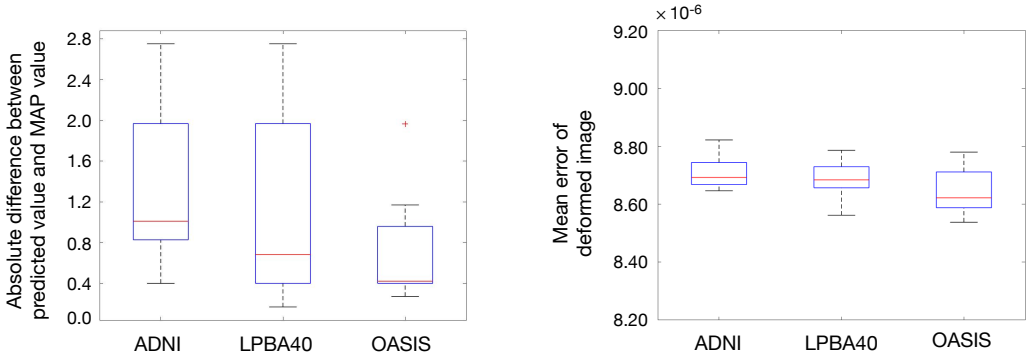


Fig. 35: Left: statistics of the difference between our prediction and MAP estimation over 300 image pairs from three data sets, ADNI, LPBA40, and OASIS; Right: error maps of image differences between deformed image by our prediction and MAP estimation.

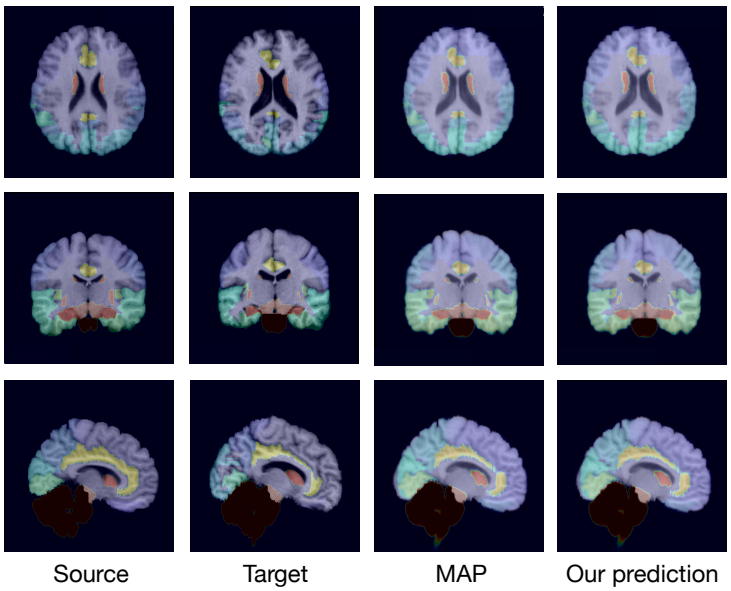


Fig. 36: Axial, coronal and sagittal view of 3D segmentation labels with nine anatomical structures overlaid with source, target, deformed images by our low-dimensional MAP ($\alpha = 8.91$) and our network prediction ($\alpha = 6.80$).

Table 4: Time and memory consumption of MAP estimation of regularization in full-dimensional image space vs. our proposed low-dimensional Bayesian model, as well as network prediction.

Methods	Full-spatial MAP	Low-dimensional MAP	Network Prediction
Runtime (Sec)	1901	257	**2.16**
Memory (MB)	450	119	**34.4**

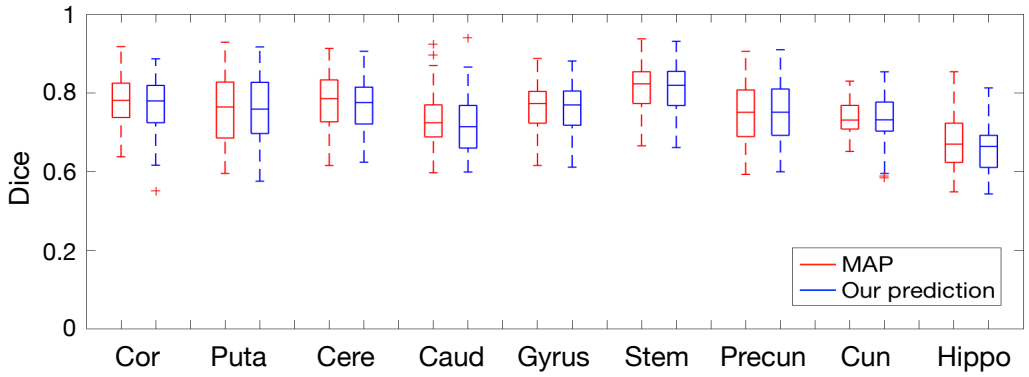


Fig. 37: Dice scores of propagated manual segmentations for both methods for 150 registration pairs. Evaluations are performed over nine anatomical brain structures, Cor (cortex), Puta (putamen), Cere (cerebellum), Caud (caudate), gyrus, Stem (brain stem), Precun (precuneus), Cun (cuneus) and Hippo (hippocampus).

by deep learning approaches. In contrast to existing methods, our developed model substantially improves the efficiency and the robustness of image registration. Our work has great potential in a variety of clinical applications, e.g. image-guided navigation system for neurosurgery in real-time. Potential future work may include: i) extending the current model to further consider adversarial examples, i.e., image outliers with significant differences; and ii) developing an unsupervised learning of registration parameter estimation to eliminate the need of training data generation (ground truth labels).

Chapter V

Learning Metamorphic Image Transformation

In this chapter, a reliable model for predictive metamorphic image registration called MetaMorph is introduced. This model is designed for the metamorphic registration of images that have undergone appearance changes due to factors such as brain tumors. The MetaMorph model incorporates a novel piecewise regularization technique on the tangent space of diffeomorphic transformations, which is developed using segmentation maps of abnormal regions obtained through learning. The model is capable of jointly learning both the geometric transformation and appearance changes, resulting in mutually beneficial outcomes. MetaMorph has the potential to be applied in various image-guided clinical interventions, such as real-time image-guided navigation systems for tumor removal surgery.

1 Learning Metamorphic Image Transformation With Appearance Change

1.1 Related Work

Deformable image registration is an important tool in a variety of medical image analysis tasks, such as multi-modality image alignment [109, 38, 136], statistical analysis for population image studies [205, 138, 181], atlas-guided image segmentation [140, 175], and object tracking with anomaly detection [34, 135]. In many clinical applications, it is desirable that the estimated transformations are diffeomorphisms (i.e., bijective, smooth, and inverse smooth mappings) because they produce anatomically plausible images [15]. Despite recent achievements in treating the problem of diffeomorphic image registration

as a fast learning task, current approaches oftentimes have an assumption that the topology of objects presented in images is intact [14, 180, 89, 32]. Existing algorithms fail badly in cases where appearance changes occur (e.g., missing data caused by pathology, such as tumors, myocardial scars, multiple sclerosis, and etc.) because they have little to no control over these unknown variables.

To address this issue, a few algorithms of image metamorphosis have been developed to incorporate the modeling of appearance changes in registration functions [130, 74, 57, 124, 22, 59]. Existing metamorphic image registration methods mainly fall into two categories: (i) exclude appearance changes via manually delineated segmentations of abnormal regions [130, 124], and (ii) treat the appearance changes as unknown variables estimated out from images [57, 22]. These approaches either heavily depend on manually segmented labels of 3D volumetric data that are time and labor-consuming, or struggle with balancing between the effects of appearance vs. geometric changes. A recent work [22] has developed a metamorphic autoencoder that estimates the deformation and appearance variations by decoupling the geometric and appearance representations in latent spaces. However, such a model is highly sensitive to parameter-tuning due to its difficulty in differentiating changes caused by geometric transformations vs. appearances.

In this section, we develop a novel learning-based model of metamorphic image registration, named as MetaMorph, that provides more robust and accurate registration results in images with appearance changes. In contrast to previous approaches [130, 57, 124, 22], we incorporate a new *appearance-aware regularization* in the network loss function that enforces a piecewise constraint on geometric transformation fields. Such a constraint will be learned simultaneously from a jointly optimized segmentation task. In addition, we effectively augment the segmentation labels by utilizing the learned transformations in the training process. This not only substantially improves the segmentation performance, but also reduces the requirement of massive groundtruth segmentation labels. The main contributions of our proposed MetaMorph are summarized as three folds:

- To the best of our knowledge, MetaMorph is the first predictive registration algorithm that utilizes jointly learned segmentation maps to model appearance changes.
- MetaMorph learns a new appearance-aware regularization that piecewisely constrains the variations of image intensities caused by geometric transformations separately from appearance changes.
- The joint learning scheme of MetaMorph maximizes the mutual benefits of metamorphic image registration and segmentation.

To demonstrate the effectiveness of our model, we validate MetaMorph on real 3D human brain tumor MRIs. Experimental results show that MetaMorph outperforms the state-of-the-art learning-based registration models [14, 22] with substantially increased accuracy. The developed MetaMorph has great potential in various image-guided clinical interventions, e.g., real-time image-guided navigation systems for tumor removal surgery.

1.2 Method: MetaMorph

1.2.1 Metamorphic Image Registration with Appearance-aware regularization The objective function of diffeomorphic image registration in Eq (**??**) works well under the condition that images are ideally of good quality with preserved topology. This assumption breaks when corruptions such as appearance changes or occlusions occur. In this section, we first define an objective function of the metamorphic image registration that considers the modeling of appearance changes. An appearance-aware regularization is developed to effectively suppress the negative influences of appearance changes in typical diffeomorphic image registration algorithms. We then develop a joint learning framework that includes i) a segmentation network for appearance change detection, and ii) a metamorphic registration network incorporating the newly formulated objective function as part of the network loss.

Appearance-aware regularization. The purpose of metamorphic image registration is to find an optimal transformation $\psi(v_0, \delta)$ that is composed of two variables: the optimal initial velocity field v_0, and the appearance change δ. A recent work proposed to learn these variables via disentangled latent representations in an encoder-decoder neural network [22]. However, it is extremely challenging for this algorithm to differentiate the variations of image intensities caused by geometric transformations from appearance changes since they unavoidably compensate each other. The ambiguity introduced by optimizing two compensating variables without any guidance fails to search for accurate registration solutions. In addition, this makes the algorithm highly sensitive to network parameters and with much-increased risk of poor convergence. To alleviate this issue, we introduce an appearance-aware regularization in the registration framework, guided by learned segmentations of the appearance changing areas.

Assume U is a union of the learned segmentations of appearance changing areas from the source image S and the target image T. Analogous to Eq (3), we define the appearance-aware regularization $\mathbf{Reg}^*(\cdot)$ in the space of initial velocity fields. To suppress the effects of appearance variations, we piecewisely constrain the initial velocity fields through a

segmentation indicator, i.e.,

$$\mathbf{Reg}^*(v_0) = (L(v_0 \odot (1-U)), v_0 \odot (1-U)), \text{ s.t. Eq (4)}, \tag{54}$$

where $\odot$ represents an element-wise multiplication between a vector field and a scalar field. For the purpose of notation simplicity, we define $\widehat{v}_0 \triangleq v_0 \odot (1-U)$ in the following sections.

With the newly defined regularization in Eq (54), we arrive at the objective function of metamorphic image registration as

$$\mathbf{E}^*[\widehat{\psi}(\widehat{v}_0)] = \mathbf{Dist}^*[\widehat{S} \circ \widehat{\psi}_1(\widehat{v}_0), \widehat{T}] + \mathbf{Reg}^*(\widehat{v}_0), \tag{55}$$

where $\widehat{S}$ and $\widehat{T}$ denotes the source and target images with appearance changes masked out, i.e., $\widehat{S} = S \odot (1-U)$, and $\widehat{T} = T \odot (1-U)$. Here, the $\mathbf{Dist}^*[\cdot,\cdot]$ is the image dissimilarity term that measures the dissimilarity between the consistent area between the deformed image and target.

1.2.2 Predictive Metamorphic Image Registration We develop a deep learning framework to jointly learn the segmentation for appearance change and the masked-out velocity field $\widehat{v}_0$. An overview of our proposed MetaMorph architecture is shown in Fig. 38.

Appearance change can be masked by a fixed foreground segmentation via pre-running image segmentation algorithms [124, 130]. However, performing manual annotations of segmentation labels is time and labor-consuming. In this work, instead of using a fixed mask, we treat the appearance change as a variable from the segmentation network and jointly optimize it with the optimal registration solution. We utilize an encoder-decoder based neural network to learn the segmentation masks and then apply them to the associate image pairs for masking out the appearance change. Although we adopt UNet-based architecture for segmentation in this work [147], other networks such as recurrent residual neural networks [4], transformer-based networks [33, 68] can also be easily plugged into the proposed method.

With the developed segmentation network, now we are ready to formulate the loss function of MetaMorph,

$$\ell = \mathbf{Dist}^*[\widehat{S} \circ \widehat{\psi}_1(\widehat{v}_0), \widehat{T}] + \mathbf{Reg}^*(\widehat{v}_0) + \gamma \cdot \ell_{seg}, \text{ s.t. Eq (54)}. \tag{56}$$

Here γ is a weighting parameter that balances segmentation and registration loss. ℓ_{seg} is the segmentation loss that maximizes the Sørensen−Dice coefficient [46] between ground truth y and the predicted $\widehat{y}$,

$$\ell_{seg} = 1 - \text{Dice}(y, \widehat{y}), \tag{57}$$

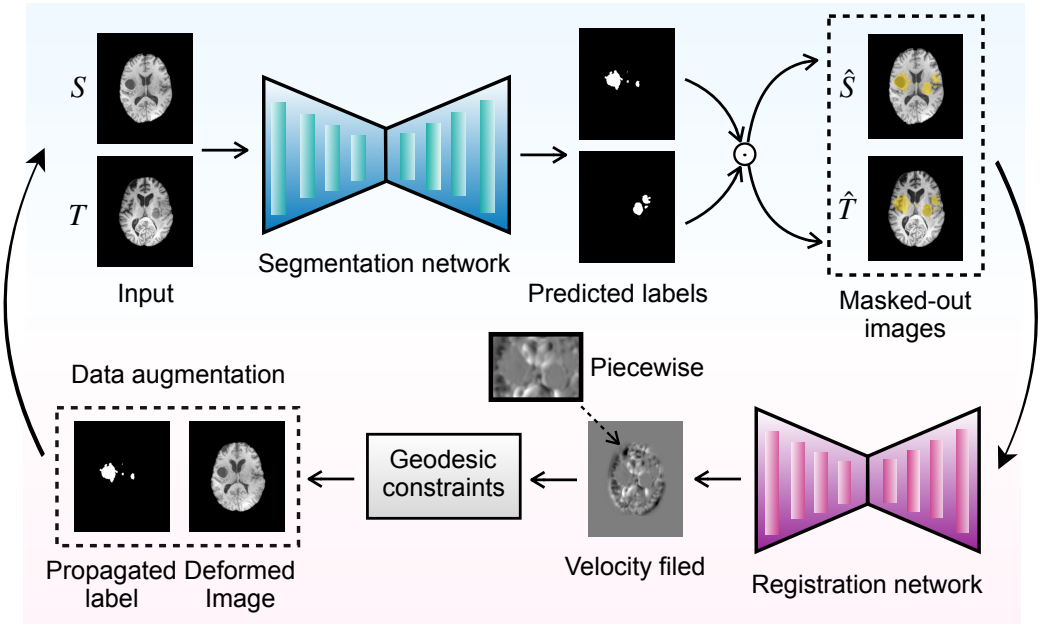


Fig. 38: An illustration of the network architecture for MetaMorph. Top left to right: input a pair of images into a segmentation network, and apply predicted labels onto images to mask out the appearance change. Bottom right to left: input a pair of images (with masked-out appearance change) to the registration network and predict a piecewise velocity field, integrate geodesic constraints, and produce a deformed image and transformation-propagated segmentation. The deformed images and labels are circulated into the segmentation network as augmented data.

where $\text{Dice}(y, \widehat{y}) = 2(|y| \cap |\widehat{y}|)/(|y| + |\widehat{y}|)$.

We adopt an approximated region-based mutual information (RMI) term [209], which is a broadly-used metric that can be applied to measure the dissimilarity term for images from different domains. For simplicity, we let $\widehat{S}_\psi$ denote the deformed image. Let $f(\widehat{S}_\psi)$ and $f(\widehat{T})$ denote the probability density functions for the deformed image and target respectively, and their joint probability density function is $f(\widehat{S}_\psi, \widehat{T})$. The image dissimilarity with an RMI formula is,

$$\begin{aligned} \textbf{Dist}^*[\widehat{S}_\psi, \widehat{T}] = \text{RMI}(\widehat{S}_\psi, \widehat{T}) &= \int_{\widehat{S}_\psi} \int_{\widehat{T}} f(\widehat{S}_\psi, \widehat{T}) \log \frac{f(\widehat{S}_\psi, \widehat{T})}{f(\widehat{S}_\psi) f(\widehat{T})} \\ &\approx l_{ce}(\widehat{S}_\psi, \widehat{T}) - \frac{1}{B} \sum_{b=1}^{B} I_b(\widehat{T}; \widehat{S}_\psi), \end{aligned} \tag{58}$$

where $L_{ce}(\cdot, \cdot)$ is a cross entropy loss between two images. $I_b(\cdot; \cdot)$ is a batch-wise lower-bound that $I_b(\widehat{T}; \widehat{S}_\psi) = \frac{1}{2} \log[\det(\Sigma_{\widehat{T}|\widehat{S}_\psi})]$, where $\Sigma_{\widehat{T}|\widehat{S}_\psi}$ is the posterior covariance matrix

of $\widehat{T}$, given $\widehat{S}_\psi$. It is a symmetric positive semi-definite matrix. B denotes the number of images in a mini-batch b. Please refer to [209] for more derivation details.

We develop an alternating optimization scheme [126] to minimize the network loss defined in Eq (56). All network parameters are optimized jointly by alternating between the training of segmentation and image registration. A summary of our joint learning of MetaMorph is in Alg. 5.

Algorithm 5 Joint learning of MetaMorph.

Input : Source and target images, the number of iterations q.

Output: Segmentation labels, the deformed image, and the transformation.

for $i = 1$ to q **do**

`/* Train image segmentation network */`

Minimize the segmentation loss in Eq (57);

Output the predicted segmentations and adopt both labels to mask appearance change in images;

`/* Train appearance-aware registration network */`

Minimizing the metamorphic loss in Eq (55) with appearance-aware geodesic constraints;

Output the predicted velocity field and the deformed image;

end

Until convergence

1.3 Experimental Evaluation

To demonstrate the effectiveness of the proposed model, we compare both segmentation and registration tasks with state-of-the-arts.

Data. For 3D brain tumor MRI scans with tumor segmemtation labels, we include 100 public T1-weighted brain scans of different subjects from Brain Tumor Segmentation (BraTS) [12, 115] challenge 2021. We also include 28 bio-markers (16 for brain ventricle and 12 for corpus callosum) that are annotated by clinicians to better evaluate the image registration performance. All MRIs are $155 \times 240 \times 240$, $1.25mm^3$ isotropic voxels. As a preprocessing step, we run affine registration, intensity normalization, and bias field correction on all images.

Experiments. We compare our metamorphic image registration method with two registration baselines, an unsupervised predictive diffeomorphic registration method (VoxelMorph as VM) [14], and a metamorphic autoencoder (MAE) [22] via learning disentangled appearance and shape representations. To better visualize the deformations, we show

predicted transformation grids and deformed images with transformation-propagated bio-markers for all methods. Quantitatively, we compute the L_2 distance as registration error between propagated bio-markers and bio-markers on target images for 60 pairs.

We evaluate the brain tumor segmentation via computing Dice similarity [46] for our model with three segmentation backbones, U-Net architecture [147], U-Net based on recurrent residual convolutional neural network (R2-Unet) [4], and transformer-based Unet (UnetR) [68]. For a fair comparison, we show the results of MetaMorph by replacing the segmentation module in our model with all backbones (named MetaMorph:Unet, MetaMorph:R2-Unet, and MetaMorph:UnetR). We visualize the predicted segmentation labels with testing images across all methods.

Parameter Settings. We set parameter $\alpha = 3$ for the operator L, the number of time steps for Euler integration in EPDiff (Eq (4)) as 10. We set the weight parameter $\gamma = 0.5$ and the batch size as 4. We use an adaptive cosine annealing learning rate scheduler that starts from an initial value at $\eta = 5e - 4$ for network training. We run all models for 100 epochs with Adam optimizer and save the networks with the best validation performance. The training and prediction procedure of all learning-based methods are performed on two Nvidia GTX 2070Ti GPUs. We run five-fold cross validation and split the images by using 70% as training images, 20% as validation images, and 10% as testing images.

Results. Fig. 39 visualizes the image registration prediction of two 3D brain MRIs of study across all methods. It shows our method MetaMorph significantly outperforms both VM and MAE. General diffeomorphic registration model (e.g., VM) without appearance-control mechanism may fail and produces less satisfied deformed images without sufficient deformations. MAE offers accurate deformations to a certain level while it produces artifacts. By excluding the appearance change, MetaMorph is able to more accurately deform the regions without tumors (e.g., ventricles and corpus callosum). It shows that our propagated bio-markers align best with the target frame.

Fig. 40 shows two examples of image segmentation performance comparison for all methods. It indicates that MetaMorph-based models predict better segmentation labels (closer to ground truth) than original backbones. Note that the predicted label by MetaMorph has a slightly better segmentation of the brain tumor boundary. This is because we use deformed images and labels that are produced by a joint registration framework as augmented data for each subject; thus learning a broader spectrum for appearance variation in data and offering more detailed prediction when testing data arrives.

Fig. 41 (left panel) statistical reports the Dice coefficient comparison. It indicates that MetaMorph consistently achieves a higher segmentation accuracy than backbones. Trans-

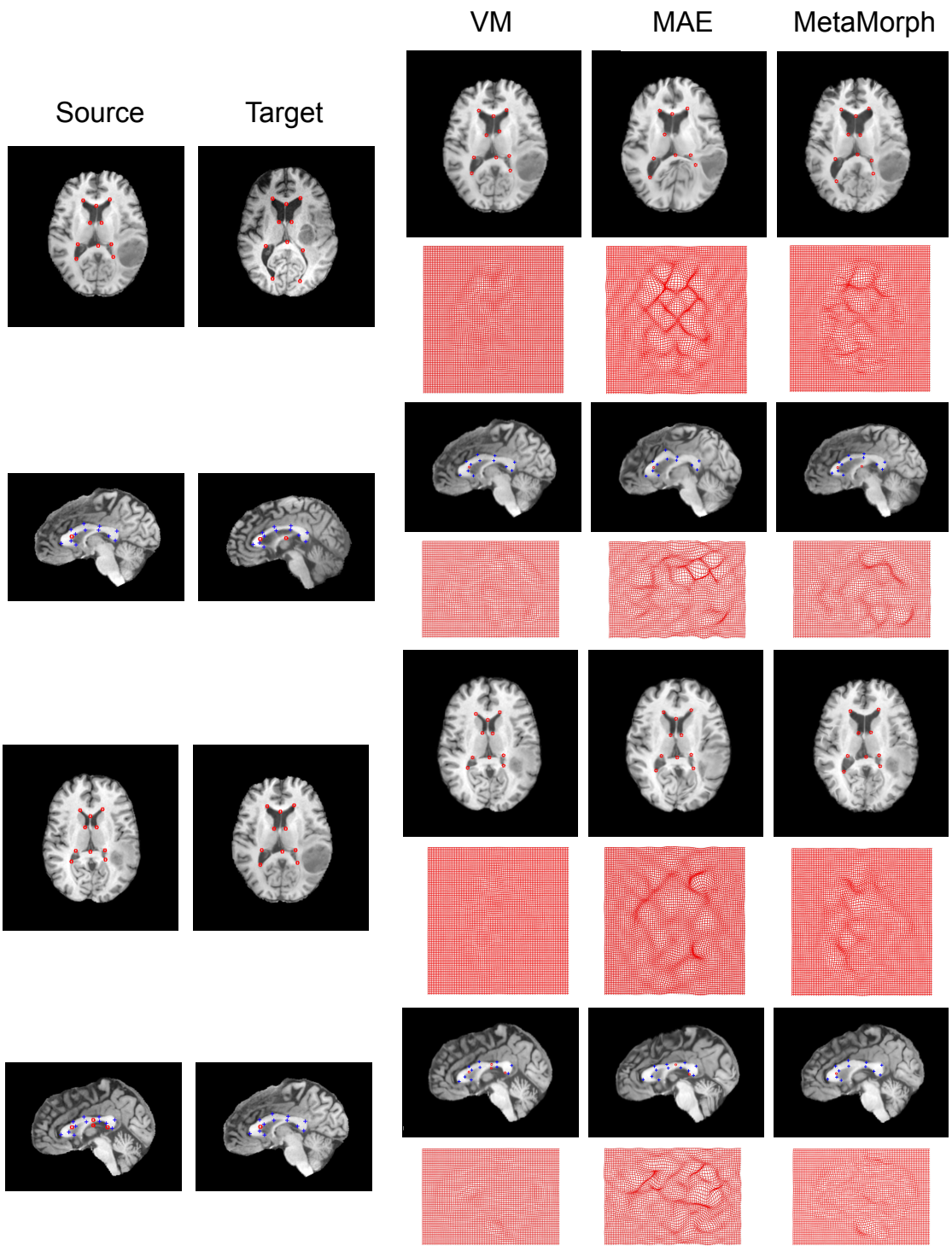


Fig. 39: Image registration performance comparison for all methods. From left to right, source, target, deformed images by VoxelMorph (VM), metamorphic autoencoder (MAE), and our method. All images are overlaid with annotated landmarks (red circle for ventricle and blue cross for corpus callosum).

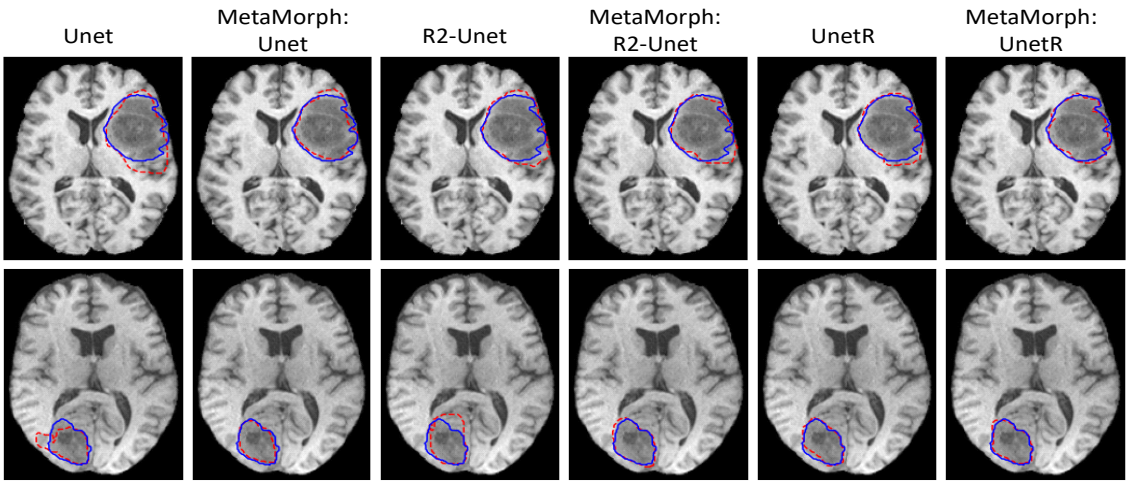


Fig. 40: Image segmentation comparison for all methods. Left to right: segmentation map difference (red) between the predicted label and the ground truth (green) for Unet, MetaMorph: Unet, R2-Unet, MetaMorph: R2-Unet, UnetR and MetaMorph: UnetR.

former based methods (UnetR-based) produce the highest Dice for all methods. Fig. 41 (right panel) reports the quantitative registration error between landmarks in the target image and landmarks in the deformed image. MetaMorph outperforms other methods with the lowest error, indicating our proposed method finishes the metamorphic image registration task with higher accuracy.

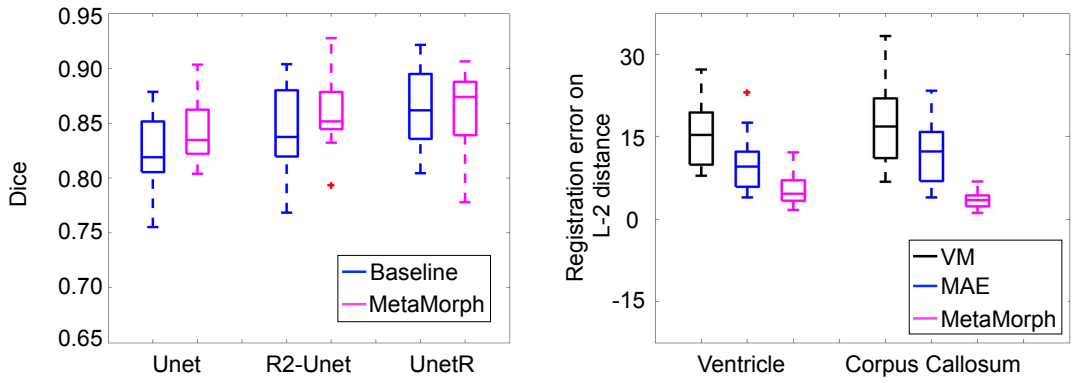


Fig. 41: Left: Dice comparison on brain tumor segmentation across all methods over images. The means of baseline vs. **our method** are 0.815/**0.834**, 0.835/**0.856**, 0.861/**0.874**; Right: registration error (computed on L_2 distance) of two anatomical landmarks for 60 brain pairs. The means of errors for VM vs. MAE vs. **our method** are 15.02/10.53/**4.64**, 16.48/13.59/**4.10**.

1.4 Conclusion

We present a predictive metamorphic image registration model, MetaMorph, via deep neural networks. Different from existing models that have limited control over appearance change, we develop a joint learning framework that adopts a segmentation module to accurately guide the registration network to learn diffeomorphic transformation fields. The developed segmentation module maximally excludes the disadvantageous effect caused by appearance change for learned deformations; thus enabling more precise correspondence alignment between deformed and target frames. Experimental results on 3D brain MRIs with real tumors show that our proposed framework yields a better registration as well as a segmentation model. While our algorithm is presented in the setting of LDDMM with geodesic shooting, the theoretical development is generic to other deformation models, e.g., stationary velocity fields [6] and a deformation method using piecewise polynomials (B-splines) [148]. Our model has great clinical potential on solving one of the most challenging registration problems, e.g., real-time brain shift estimation between preoperative and intraoperative MRI scans with missing data values. Interesting future works of MetaMorph will be i) building a probabilistic model to quantify the registration uncertainty along the boundary of tumor areas and ii) extending the proposed method to more challenging clinical scenarios that appearance changes are difficult to track, e.g., real-time automated image registration on ultrasound images.

Chapter VI

Conclusion and Future Work

In this dissertation, my research bridges the gap between the developed algorithmic foundations of deformable shape learning models and deep learning-based applications by presenting predictive diffeomorphic and metamorphic image registration, registration regularization estimation, atlas building, image segmentation, and image classification. In Chapter III, I show that the developed low-dimensional shape representation learning can be effectively applied in multiple medical image analysis tasks, such as image registration [180, 179], uncertainty quantification [176, 177], parameter estimation [182] and atlas building [181]. In Chapter IV, the learned deformation-based geometric descriptors allow us efficiently perform deep neural network training, testing, inference, and population studies. In chapter V, I show the developed predictive metamorphic registration is able to precisely predict transformations for images with appearance changes (i.e., caused by brain tumors). Such a method has great potential in various image-guided clinical interventions, e.g., real-time image-guided navigation systems for tumor removal surgery [106, 178, 57]. Except for the image analysis tasks mentioned in this research, the theoretical tools developed from all methods are generic to various other deep-learning based image analysis models, for instance, image reconstruction [82] and object tracking [58]. The final outputs of this research include a number of novel algorithms with publicly released codes online. The following section summarizes the contributions of this dissertation and discusses possible future works that are extended from this research.

1 Summary of contributions

This section reviews the dissertation presented in Chapter III to Chapter V. Each contribution is revisited with a summary of how it was fulfilled.

(a) *Bayesian models that enable efficient uncertainty quantification, hierarchical atlas building, and regularization parameter estimation.*

According to the Statement 1 in Chapter I, Chapter III presents an efficient Bayesian approach to quantifying image registration uncertainty based on a low-dimensional representation of geometric deformations. In contrast to previous methods, I develop a Bayesian diffeomorphic registration framework in a bandlimited space, rather than a high-dimensional image space. It shows that a dense posterior distribution on deformation fields can be fully characterized by much fewer parameters, which dramatically reduces the computational complexity of model inferences. I also develop a novel hierarchical Bayesian model for unbiased atlas building with adaptive regularizations of image registration. It automatically enables model to select parameters to control the smoothness of diffeomorphic transformations. A hierarchical prior distribution on the regularization parameters is developed, and it further allows multiple penalties on images with various degrees of geometric transformations. I treat the regularization parameters as latent variables and integrate them out from the model by using the Monte Carlo Expectation Maximization (MCEM) algorithm. Another advantage of this model is that it eliminates the need for manual parameter tuning, which can be tedious and infeasible. Experimental results on both 2D synthetic data and real 3D brain MRI scans demonstrate that the proposed Bayesian models are significantly efficient than the state-of-the-art, and provides improve quality atlas.

(b) *A predictive approach for fast image registration with low-dimensional network training and inference.*

We recall the Statement 1 in Chapter I, Chapter IV first presents DeepFLASH, a novel network with efficient training and inference for learning-based medical image registration. Different from existing approaches that learn spatial transformations from training data in the high dimensional imaging space, a new registration network is entirely developed in a low dimensional bandlimited space. This dramatically reduces the computational cost and memory footprint of an expensive training and inference. To achieve this goal, I first introduce complex-valued operations and representations of neural architectures that provide key components for learning-based registration models. DeepFLASH employed an explicit loss function of transformation fields fully characterized in a bandlimited space with much fewer parameterizations. Experimental results on both 2D and 3D data show that our method is significantly faster than the state-of-the-art deep learning based image registration methods, while producing equally accurate alignment.

(c) *A classification model that explicitly learns geometric deformable shape information.*
We then revisit the Statement 2 in Chapter I, Chapter IV also presents Geo-SIC, the

first deep learning model to learn deformable shapes in a deformation space for an improved performance of image classification. I introduce a newly designed framework that (i) simultaneously derives features from both image and latent shape spaces with large intra-class variations; and (ii) gains increased model interpretability by allowing direct access to the underlying geometric features of image data. In particular, a boosted classification network is presented with an unsupervised learning of geometric shape representations characterized by diffeomorphic transformations within each class. Geo-SIC provides a more fundamental approach by naturally learning the most relevant shape features jointly with an image classifier. Experiments on both simulated 2D images and real MRIs show that this model substantially improves the image classification accuracy with an additional benefit of increased model interpretability.

(d) *A predictive approach for metamrophic image registration with appearance changes.* We revisit the Statement 3 in Chapter I, Chapter V presents a novel predictive model, MetaMorph, for metamorphic registration of images with appearance changes (i.e., caused by brain tumors). In contrast to previous learning-based registration methods that have little or no control over appearance-changes, MetaMorph introduces a new regularization that can effectively suppress the negative effects of appearance changing areas. In particular, I develop a piecewise regularization on the tangent space of diffeomorphic transformations (also known as initial velocity fields) via learned segmentation maps of abnormal regions. The geometric transformation and appearance changes are treated as joint tasks that are mutually beneficial. MetaMorph is more robust and accurate when searching for an optimal registration solution under the guidance of segmentation, which in turn improves the segmentation performance by providing appropriately augmented training labels. Experimental results on 3D human brain tumor MRI scans show that this model outperforms the state-of-the-art learning-based registration models. MetaMorph has great potential in various image-guided clinical interventions, e.g., real-time image-guided navigation systems for tumor removal surgery.

2 Future work

In this section, I will discuss potential future works that can be extended from this dissertation. The main research directions include i) the robustness of deformable shape learning models and ii) the deformation estimation in clinical scenarios where have severe motion artifacts occur, e.g., real-time fetal brain monitoring system.

(a) *Robustness of Deformable Shape Learning Models.*

Adversarial attacks can be a potential concern in medical imaging because incorrect diagnoses or treatment decisions have severe consequences for patients' health [54, 24, 108]. Attackers may try to subvert medical image analysis systems by introducing subtle changes to the images, such as adding noise, altering the contrast or brightness, or making small modifications to the image's pixels [24, 129]. Such changes can lead to incorrect diagnoses or treatment recommendations, potentially causing harm to patients. Adversarial attacks can be especially concerning in situations where the model is used to make critical decisions, such as in the diagnosis of cancer or other life-threatening conditions.

To address this issue, a variety of methods haven been developed to make medical image analysis models more robust against adversarial attacks [100, 43] . These techniques include adversarial training, which involves trains the model on both clean and adversarial examples and utilizes data augmentation and model regularization to improve robustness [31]. While learning-based deformable shape modeling have shown promising results in improving the robustness of networks, there is still a lack of formal research on their robustness. In order to assess the robustness of the proposed research, a simple approach is to produce a series of adversarial examples, and subsequently gauge the effectiveness of models by their capacity to correctly classify these instances. If the model's accuracy significantly drops on the adversarial examples, it is considered to be less robust. For adversarial robustness testing, the ultimate goal is to evaluate the ability of models to resist attacks from adversarial examples. In Chapter IV, experimental results show that Geo-SIC consistently achieves better classification accuracy ($\sim 10\%$ higher in average) than baseline algorithms across different levels of adversarial attacks on image intensity. This is due to the superior capability of Geo-SIC in recognizing shape features that are relatively immune to a certain degree of noise, as the geometric shape features are ensured to remain smooth within the tangent space of diffeomorphisms. However, the classification accuracy of Geo-SIC still drops dramatically when adding a high level of universal adversarial perturbations.

A possible solution for addressing adversarial attacks in deformable shape learning networks is to introduce a discriminator module based on the Generative Adversarial Network (GAN) framework [62]. The discriminator module takes in an input deformed image (either a real target or a deformed image generated from a sample under adversarial attack) and produces an output indicating whether the input is real or fake. The goal of the discriminator is to correctly classify the input data as either real or fake with high accuracy. On the other hand, the deformable shape learning network can be treated as a geometric shape generator that generates indistinguishable data

from real data. This algorithm aims to fool the discriminator by generating data that is classified as real. The geometric shape generator will be improved by minimizing the discriminator's ability to differentiate between the real and fake data, while the discriminator will be improved by maximizing its ability to differentiate between the two. By iterating between the geometric shape generator and the discriminator, the GAN learns to generate increasingly realistic deformed image and atlas that is difficult for the discriminator to distinguish from real data [44].

This approach can be effective in detecting adversarial attacks in deformable shape learning networks, as the discriminator is able to identify if the deformed image (or atlas) is under adversarial attack. By training the GAN on large datasets of both real and adversarial examples, the model will learn to accurately classify and distinguish between the two, thereby improving the robustness and reliability of the deformable shape learning network.

(b) *Joint Learning for Motion Correction via Deformable Shape Learning and Image Reconstruction in Real-time Clinical Settings Where Have Severe Motion Artifacts.*

The motion artifacts in fetal brain MRI can be problematic because of the need for high spatial resolution to visualize the small, developing structures in the fetal brain. The fetal brain is constantly growing and developing, which makes it particularly susceptible to motion artifacts during MRI scans. The motion artifacts are majorly caused by i) fetal movement, ii) maternal breathing, iii) long scan times, and iv) limitations in spatial resolution. Fetal movement during MRI scans causes the fetal brain to shift within the MRI volume, resulting in blurred or distorted images. Similarly, maternal breathing introduces motion artifacts due to the subtle shifts in fetal position caused by diaphragm and chest wall movement. Longer scan times also increase the chance of motion artifacts as small fetal movements can result in image blurring or distortion. The impact of motion artifacts is exacerbated by limitations in spatial resolution, which is often necessary to visualize small and developing structures in the fetal brain. Higher spatial resolution images are more susceptible to motion artifacts as even subtle movements have a larger impact on image quality. To properly handle the motion artifacts in fetal brain MRI, motion correction and image reconstruction methods based on rigid/affine transformation are presented to minimize artifacts in fetal brain MRI [165, 189, 78, 151]. These techniques aim to correct the motions during scans and improve the quality of images produced.

Sui et al. [165] developed a real-time slice-to-volume image registration algorithm, named SLIMM, with a parallelized patch-based approach to enable a fetal brain monitoring system that offers improved online model accuracy. Xu et al. [188] further proposed a slice-to-volume registration transformer (SVoRT) to map multiple stacks of fe-

tal MR slices into a canonical 3D space and achieved comparable model performance for slice-to-volume registration and 3D reconstruction. However, due to fetal motion is highly non-rigid, aforementioned methods have limitations in handling challenging scenarios where the brain contains large deformable shape variations, occlusions, and local image distortions. The anatomical structures of the brain exhibit varying degrees of local deformations, which means that the image voxels move in different directions. Affine transformations assume that the image is undergoing a rigid transformation, in which all structures of the image move globally. As a result, learning affine transformations to correct for motion artifacts in fetal MRI scans may lead to misalignment with inaccurate registration. In addition, fetal MRI scans are oftentimes acquired in multiple planes, which result in image distortion due to differences in slice thickness and orientation, making it more challenging to apply affine transformations consistently across all planes.

To address these limitations, a comprehensive framework that includes a deformable motion correction method and image reconstruction is needed. A deformable motion correction method will nicely handle the non-rigid movement of the fetal brain and improve the accuracy of registration. Meanwhile, image reconstruction techniques help to reduce image distortion and improve the quality of images produced. The use of such a comprehensive framework can be more appropriate for accurately aligning fetal MRI scans, especially in challenging scenarios where large deformations and severe image distortions are present.

In this future work, my goal is to develop a joint learning framework that optimizes both deformable geometric motion correction and image reconstruction tasks, in order to improve the accuracy and speed of the overall process. The major components of this model are:

(i) A joint objective function that combines the metrics used for motion correction and image reconstruction. The objective function includes a similarity between the motion-corrected and reconstructed images, a dissimilarity between motion-corrected and ground truth images, and a regularity term that measures the smoothness of the deformation field. To ensure the preservation of anatomical information, I will employ the same regularization (geodesic constraints Eq (3)& Eq (5) in the tangent space of diffeomorphisms), which is developed in this research.

(ii) An optimization scheme that efficiently optimizes over the entire model. To accomplish this, I will employ an alternative optimization strategy that iteratively updates the deformation field and the reconstructed image until convergence, similar to the optimization schemes in Geo-SIC and MetaMorph described in Chapter. IV and Chapter. V, respectively.

(iii) An efficient deformable registration algorithm to perform deformable motion correction and guarantees the high efficiency for real-time clinical applications. I will utilize a recent work based on the Fourier Neuron Operator [101, 187], which parameterized computation of EPDiff (Eq (5)) via deep neural networks. This will be efficient in testing, as it completely eliminates the need to integrate PDEs with high computational cost.

(iv) An iterative reconstruction algorithm will be required to reconstruct the image from the motion-corrected data. The choice of algorithm will depend on the imaging modality and acquisition parameters. Analogous to the iterative image reconstruction technique employed in SVoRT [188], I will commence this work by implementing the same scheme.

By jointly optimizing image registration and image reconstruction, this algorithm takes advantage of the additional information provided by the registration process to guide the reconstruction process. For example, by aligning the images before reconstruction, the amount of artifacts and noise in the resulting images can be reduced. In return, image reconstruction improves the accuracy and robustness of the registration process by incorporating more accurate image information from the reconstructed data. Overall, joint optimization for image registration and image reconstruction will lead to significant improvements in both tasks, particularly in challenging scenarios where data has severe motion artifacts; hence making it a promising future direction based on this research.

Bibliography

[1] Achlioptas, P., Diamanti, O., Mitliagkas, I., and Guibas, L. (2018). Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49.

[2] Allassonnière, S., Amit, Y., and Trouvé, A. (2007). Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):3–29.

[3] Allassonnière, S. and Kuhn, E. (2008). Stochastic algorithm for parameter estimation for dense deformable template mixture model. *arXiv preprint arXiv:0802.1521*.

[4] Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*.

[5] Arnold, V. (1966). Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. In *Annales de l'institut Fourier*, volume 16, pages 319–361.

[6] Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A log-euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 924–931. Springer.

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

[8] Ashburner, J. and Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and gauss–newton optimisation. *NeuroImage*, 55(3):954–967.

[9] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41.

[10] Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044.

[11] Bai, X., Liu, W., and Tu, Z. (2009). Integrating contour and skeleton for shape classification. In *2009 IEEE 12th international conference on computer vision workshops,*

ICCV workshops, pages 360–367. IEEE.

[12] Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F. C., Pati, S., et al. (2021). The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*.

[13] Baker, N., Lu, H., Erlikhman, G., and Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS computational biology*, 14(12):e1006613.

[14] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2019). Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800.

[15] Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157.

[16] Bhalodia, R., Elhabian, S. Y., Kavan, L., and Whitaker, R. T. (2018). Deepssm: a deep learning framework for statistical shape modeling from raw images. In *International Workshop on Shape in Medical Imaging*, pages 244–257. Springer.

[17] Biffi, C., Cerrolaza, J. J., Tarroni, G., Bai, W., De Marvao, A., Oktay, O., Ledig, C., Le Folgoc, L., Kamnitsas, K., Doumou, G., et al. (2020). Explainable anatomical shape analysis through deep hierarchical generative models. *IEEE transactions on medical imaging*, 39(6):2088–2099.

[18] Blinkouskaya, Y. and Weickenmeier, J. (2021). Brain shape changes associated with cerebral atrophy in healthy aging and alzheimer‘‘s disease. *Frontiers in Mechanical Engineering*, page 64.

[19] Blum, H. (1967). A transformation for extracting new descriptors of shape. In Wathen-Dunn, W., editor, *Models for Perception of Speech and Visual Form*. MIT Press, Cambridge, MA.

[20] Boissonnat, J.-D. (1984). Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics (TOG)*, 3(4):266–286.

[21] Bône, A., Louis, M., Colliot, O., Durrleman, S., Initiative, A. D. N., et al. (2019). Learning low-dimensional representations of shape data sets with diffeomorphic autoencoders. In *International conference on information processing in medical imaging*, pages 195–207. Springer.

[22] Bône, A., Vernhet, P., Colliot, O., and Durrleman, S. (2020). Learning joint shape and appearance representations with metamorphic auto-encoders. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 202–211. Springer.

[23] Bookstein, F. L. (1997). *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press.

[24] Bortsova, G., González-Gonzalo, C., Wetstein, S. C., Dubost, F., Katramados, I., Hogeweg, L., Liefers, B., van Ginneken, B., Pluim, J. P., Veta, M., et al. (2021). Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. *Medical Image Analysis*, 73:102141.

[25] Brechbühler, C., Gerig, G., and Kübler, O. (1995). Parametrization of closed surfaces for 3-d shape description. *Computer vision and image understanding*, 61(2):154–170.

[26] Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017). Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42.

[27] Bullo, F. (1995). Invariant affine connections and controllability on lie groups. Technical report, technical Report for Geometric Mechanics, California Institute of Technology.

[28] Cao, T., Singh, N., Jojic, V., and Niethammer, M. (2015). Semi-coupled dictionary learning for deformation prediction. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 691–694. IEEE.

[29] Chatfield, K., Philbin, J., and Zisserman, A. (2009). Efficient retrieval of deformable shape classes using local self-similarities. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 264–271. IEEE.

[30] Chen, C., Biffi, C., Tarroni, G., Petersen, S., Bai, W., and Rueckert, D. (2019). Learning shape priors for robust cardiac mr segmentation from multi-view images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 523–531. Springer.

[31] Chen, C., Qin, C., Qiu, H., Ouyang, C., Wang, S., Chen, L., Tarroni, G., Bai, W., and Rueckert, D. (2020). Realistic adversarial data augmentation for mr image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, pages 667–677. Springer.

[32] Chen, J., Frey, E. C., He, Y., Segars, W. P., Li, Y., and Du, Y. (2022). Transmorph: Transformer for unsupervised medical image registration. *Medical Image Analysis*, 82:102615.

[33] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

[34] Chen, M., Kanade, T., Pomerleau, D., and Rowley, H. A. (1999). Anomaly detection through registration. *Pattern Recognition*, 32(1):113–128.

[35] Chou, C.-R., Frederick, B., Mageras, G., Chang, S., and Pizer, S. (2013). 2d/3d image registration using regression learning. *Computer Vision and Image Understanding*, 117(9):1095–1106.

[36] Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1993). A deformable neuroanatomy textbook based on viscous fluid mechanics. In *27th Ann. Conf. on Inf. Sciences and Systems*, pages 211–216.

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

[38] Chung, A., Wells, W. M., Norbash, A., and Grimson, W. E. L. (2002). Multi-modal image registration by minimising kullback-leibler distance. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 525–532. Springer.

[39] Cockburn, B. and Shu, C.-W. (1989). Tvb runge-kutta local projection discontinuous galerkin finite element method for conservation laws. ii. general framework. *Mathematics of computation*, 52(186):411–435.

[40] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59.

[41] Dai, Z., Liu, H., Le, Q., and Tan, M. (2021). Coatnet: Marrying convolution and attention for all data sizes. *Advances in Neural Information Processing Systems*, 34.

[42] Dalca, A., Rakic, M., Guttag, J., and Sabuncu, M. (2019). Learning conditional deformable templates with convolutional networks. *Advances in neural information processing systems*, 32.

[43] Daza, L., Pérez, J. C., and Arbeláez, P. (2021). Towards robust general medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 3–13. Springer.

[44] Dey, N., Ren, M., Dalca, A. V., and Gerig, G. (2021). Generative adversarial registration for improved conditional deformable templates. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3929–3941.

[45] Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659.

[46] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

[47] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[48] Douillard, B., Underwood, J., Kuntz, N., Vlaskine, V., Quadros, A., Morton, P., and Frenkel, A. (2011). On the segmentation of 3d lidar point clouds. In *2011 IEEE International Conference on Robotics and Automation*, pages 2798–2805. IEEE.

[49] Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, 195(2):216–222.

[50] Durrleman, S., Prastawa, M., Gerig, G., and Joshi, S. (2011). Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 123–134. Springer.

[51] El-ghazal, A., Basir, O., and Belkasim, S. (2012). Invariant curvature-based fourier shape descriptors. *Journal of Visual Communication and Image Representation*, 23(4):622–633.

[52] Erler, P., Guerrero, P., Ohrhallinger, S., Mitra, N. J., and Wimmer, M. (2020). Points2surf learning implicit surfaces from point clouds. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V*, pages 108–124. Springer.

[53] Esser, P., Sutter, E., and Ommer, B. (2018). A variational u-net for conditional appearance and shape generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8857–8866.

[54] Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., and Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289.

[55] Foret, P., Kleiner, A., Mobahi, H., and Neyshabur, B. (2020). Sharpness-aware minimization for efficiently improving generalization. *arXiv preprint arXiv:2010.01412*.

[56] Fotenos, A. F., Snyder, A., Girton, L., Morris, J., and Buckner, R. (2005). Normative estimates of cross-sectional and longitudinal brain volume decline in aging and ad. *Neurology*, 64(6):1032–1039.

[57] François, A., Gori, P., and Glaunès, J. (2021). Metamorphic image registration using a semi-lagrangian scheme. In *International Conference on Geometric Science of Information*, pages 781–788. Springer.

[58] Galvez, R. L., Bandala, A. A., Dadios, E. P., Vicerra, R. R. P., and Maningo, J. M. Z. (2018). Object detection using convolutional neural networks. In *TENCON 2018-2018 IEEE Region 10 Conference*, pages 2023–2027. IEEE.

[59] Gao, Y., Zhang, M., Grewen, K., Fletcher, P. T., and Gerig, G. (2016). Image registration and segmentation in longitudinal mri using temporal appearance modeling. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 629–632. IEEE.

[60] Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.-S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., et al. (2009). Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage*, 47(4):1476–1486.

[61] Golland, P., Grimson, W. E. L., Shenton, M. E., and Kikinis, R. (2001). Deformation analysis for shape based classification. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 517–530. Springer.

[62] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

[63] Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2017). A bayesian framework for joint morphometry of surface and curve meshes in multi-object complexes. *Medical image analysis*, 35:458–474.

[64] Haber, E., Ascher, U. M., and Oldenburg, D. (2000). On optimization techniques for solving nonlinear inverse problems. *Inverse problems*, 16(5):1263.

[65] Haber, E. and Modersitzki, J. (2006). A multilevel method for image registration. *SIAM journal on scientific computing*, 27(5):1594–1607.

[66] Hajek, B. and Wong, E. (1989). Stochastic processes in information and dynamical systems.

[67] Hara, K., Kataoka, H., and Satoh, Y. (2017). Learning spatio-temporal features with 3d residual networks for action recognition. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 3154–3160.

[68] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., and Xu, D. (2022). Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 574–584.

[69] He, K. and Sun, J. (2015). Convolutional neural networks at constrained time cost. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5353–5360.

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

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

[72] Heimann, T. and Meinzer, H.-P. (2009). Statistical shape models for 3d medical image segmentation: a review. *Medical image analysis*, 13(4):543–563.

[73] Hinkle, J., Womble, D., and Yoon, H.-J. (2018). Diffeomorphic autoencoders for lddmm atlas building.

[74] Holm, D., Trouvé, A., and Younes, L. (2009). The euler-poincaré theory of metamorphosis. *Quarterly of Applied Mathematics*, 67(4):661–685.

[75] Holmes, P., Lumley, J. L., Berkooz, G., and Rowley, C. W. (2012). *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press.

[76] Hong, Y., Golland, P., and Zhang, M. (2017). Fast geodesic regression for population-based image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 317–325. Springer.

[77] Hong, Y., Joshi, S., Sanchez, M., Styner, M., and Niethammer, M. (2012). Metamorphic geodesic regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 197–205. Springer.

[78] Hou, B., Khanal, B., Alansary, A., McDonagh, S., Davidson, A., Rutherford, M., Hajnal, J. V., Rueckert, D., Glocker, B., and Kainz, B. (2018). 3-d reconstruction in canonical co-ordinate space from arbitrarily oriented 2-d images. *IEEE transactions on medical imaging*, 37(8):1737–1750.

[79] Iglesias, J. E., Sabuncu, M. R., and Van Leemput, K. (2012). Incorporating parameter uncertainty in bayesian segmentation models: Application to hippocampal subfield volumetry. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 50–57. Springer.

[80] Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L. Whitwell, J., Ward, C., et al. (2008). The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691.

[81] Jaillet, L., Yershova, A., La Valle, S. M., and Siméon, T. (2005). Adaptive tuning of the sampling domain for dynamic-domain rrts. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2851–2856. IEEE.

[82] Jiang, H., Cai, J., and Zheng, J. (2019). Skeleton-aware 3d human shape reconstruction from point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5431–5441.

[83] Jongejan, J., Rowley, H., Kawashima, T., Kim, J., and Fox-Gieg, N. (2016). The quick, draw!-ai experiment. *Mount View, CA, accessed Feb*, 17(2018):4.

[84] Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233.

[85] Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004a). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23, Supplement1:151–160.

[86] Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004b). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160.

[87] Just, M., Rösler, H., Higer, H., Kutzner, J., and Thelen, M. (1991). Mri-assisted radiation therapy planning of brain tumors-clinical experiences in 17 patients. *Magnetic resonance imaging*, 9(2):173–177.

[88] Katz, R. A. and Pizer, S. M. (2003). Untangling the blum medial axis transform. *International Journal of Computer Vision*, 55.

[89] Kim, B., Han, I., and Ye, J. C. (2021). Diffusemorph: Unsupervised deformable image registration along continuous trajectory using diffusion models. *arXiv preprint arXiv:2112.05149*.

[90] Kirillov, A., Wu, Y., He, K., and Girshick, R. (2020). Pointrend: Image segmentation as rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9799–9808.

[91] Krebs, J., Delingette, H., Mailhé, B., Ayache, N., and Mansi, T. (2019). Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging*, 38(9):2165–2176.

[92] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

[93] Kunttu, I., Lepisto, L., Rauhamaa, J., and Visa, A. (2004). Multiscale fourier descriptor for shape-based image retrieval. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 765–768. IEEE.

[94] Kybic, J. (2010). Bootstrap resampling for image registration uncertainty estimation without ground truth. *IEEE Transactions on Image Processing*, 19(1):64–73.

[95] Kybic, J. and Unser, M. (2003). Fast parametric elastic image registration. *IEEE transactions on image processing*, 12(11):1427–1442.

[96] Le Folgoc, L., Delingette, H., Criminisi, A., and Ayache, N. (2017). Quantifying registration uncertainty with sparse bayesian modelling. *IEEE transactions on medical imaging*, 36(2):607–617.

[97] Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S., and Wiechert, A. (2010). Point clouds. *Photogrammetric Engineering & Remote Sensing*, 76(10):1123–1134.

[98] Legouhy, A., Commowick, O., Rousseau, F., and Barillot, C. (2019). Online atlasing using an iterative centroid. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 366–374. Springer.

[99] Li, S., Shi, F., Pu, F., Li, X., Jiang, T., Xie, S., and Wang, Y. (2007). Hippocampal shape analysis of alzheimer disease based on machine learning methods. *American Journal of Neuroradiology*, 28(7):1339–1345.

[100] Li, X., Pan, D., and Zhu, D. (2021). Defending against adversarial attacks on medical imaging ai system, classification or detection? In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1677–1681. IEEE.

[101] Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.

[102] Liao, R., Turk, E. A., Zhang, M., Luo, J., Adalsteinsson, E., Grant, P. E., and Golland, P. (2019). Temporal registration in application to in-utero mri time series. *arXiv preprint arXiv:1903.02959*.

[103] Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current biology*, 5(5):552–563.

[104] Lorenzo-Valdés, M., Sanchez-Ortiz, G. I., Mohiaddin, R., and Rueckert, D. (2002). Atlas-based segmentation and tracking of 3d cardiac mr images using non-rigid registration. In *International conference on medical image computing and computer-assisted intervention*, pages 642–650. Springer.

[105] Luo, J., Sedghi, A., Popuri, K., Cobzas, D., Zhang, M., Preiswerk, F., Toews, M., Golby, A., Sugiyama, M., Wells, W. M., et al. (2019). On the applicability of registration uncertainty. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 410–419. Springer.

[106] Luo, J., Toews, M., Machado, I., Frisken, S., Zhang, M., Preiswerk, F., Sedghi, A., Ding, H., Pieper, S., Golland, P., et al. (2018). A feature-driven active framework for ultrasound-based brain shift compensation. *arXiv preprint arXiv:1803.07682*.

[107] Ma, J., Miller, M. I., Trouvé, A., and Younes, L. (2008). Bayesian template estimation in computational anatomy. *NeuroImage*, 42(1):252–261.

[108] Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., and Lu, F. (2021). Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition*, 110:107332.

[109] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16(2):187–198.

[110] Maillard, M., François, A., Glaunès, J., Bloch, I., and Gori, P. (2022). A deep residual learning implementation of metamorphosis. *arXiv preprint arXiv:2202.00676*.

[111] Mang, A. and Biros, G. (2015). An inexact newton–krylov algorithm for constrained diffeomorphic image registration. *SIAM journal on imaging sciences*, 8(2):1030–1069.

[112] Mang, A., Gholami, A., Davatzikos, C., and Biros, G. (2019). Claire: A distributed-memory solver for constrained large deformation diffeomorphic image registration. *SIAM Journal on Scientific Computing*, 41(5):C548–C584.

[113] Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Cross-sectional mri data in young, middle aged, nondemented and demented older adults. *Cognitive Neuroscie*, pages 1489–1507.

[114] McInerney, T. and Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108.

[115] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024.

[116] Miller, M. I., Trouvé, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, 24(2):209–228.

[117] Minka, T. (2000). Estimating a dirichlet distribution.

[118] Mohanty, N., Rath, T. M., Lee, A., and Manmatha, R. (2005). Learning shapes for image classification and retrieval. In *International Conference on Image and Video Retrieval*, pages 589–598. Springer.

[119] Moosavi-Dezfooli, S.-M., Fawzi, A., Fawzi, O., and Frossard, P. (2017). Universal adversarial perturbations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1765–1773.

[120] Moosavi-Dezfooli, S.-M., Fawzi, A., and Frossard, P. (2016). Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2574–2582.

[121] Muneer, A. and Fati, S. M. (2020). Efficient and automated herbs classification approach based on shape and texture features using deep learning. *IEEE Access*, 8:196747–196764.

[122] Nabavi, A., McL. Black, P., Gering, D. T., Westin, C.-F., Mehta, V., Pergolizzi Jr, R. S., Ferrant, M., Warfield, S. K., Hata, N., Schwartz, R. B., et al. (2001). Serial intraoperative magnetic resonance imaging of brain shift. *Neurosurgery*, 48(4):787–798.

[123] Newman, A. J. (1996). Model reduction via the karhunen-loeve expansion part i: An exposition. Technical report.

[124] Niethammer, M., Hart, G. L., Pace, D. F., Vespa, P. M., Irimia, A., Horn, J. D. V., and Aylward, S. R. (2011a). Geometric metamorphosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 639–646. Springer.

[125] Niethammer, M., Huang, Y., and Vialard, F. X. (2011b). Geodesic regression for image time-series. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 655–662. Springer.

[126] Nocedal, J. and Wright, S. J. (1999). *Numerical optimization*. Springer.

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

[128] Nunes, J. F., Moreira, P. M., and Tavares, J. M. R. (2010). Shape based image retrieval and classification. In *5th Iberian Conference on Information Systems and Technologies*, pages 1–6. IEEE.

[129] Ozbulak, U., Van Messem, A., and De Neve, W. (2019). Impact of adversarial examples on deep learning models for biomedical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 300–308. Springer.

[130] Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011). A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3):907–922.

[131] Persoon, E. and Fu, K.-S. (1977). Shape discrimination using fourier descriptors. *IEEE Transactions on systems, man, and cybernetics*, 7(3):170–179.

[132] Pizer, S. M., Fritsch, D. S., Yushkevich, P. A., Johnson, V. E., and Chaney, E. L. (1999). Segmentation, registration, and measurement of shape variation via image object shape. *IEEE transactions on medical imaging*, 18(10):851–865.

[133] Pohl, K. M., Fisher, J., Grimson, W. E. L., Kikinis, R., and Wells, W. M. (2006). A bayesian model for joint segmentation and registration. *NeuroImage*, 31(1):228–239.

[134] Polzin, T., Niethammer, M., Heinrich, M. P., Handels, H., and Modersitzki, J. (2016). Memory efficient lddmm for lung ct. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 28–36. Springer.

[135] Prastawa, M., Bullitt, E., Ho, S., and Gerig, G. (2004). A brain tumor segmentation framework based on outlier detection. *Medical image analysis*, 8(3):275–283.

[136] Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., and Kamen, A. (2019). Unsupervised deformable registration for multi-modal images via disentangled representations. In *International Conference on Information Processing in Medical Imaging*, pages 249–261. Springer.

[137] Qiu, A., Younes, L., and Miller, M. I. (2012). Principal component based diffeomorphic surface mapping. *Medical Imaging, IEEE Transactions on*, 31(2):302–311.

[138] Rao, A., Aljabar, P., and Rueckert, D. (2008). Hierarchical statistical shape analysis and prediction of sub-cortical brain structures. *Medical image analysis*, 12(1):55–68.

[139] Rematas, K., Martin-Brualla, R., and Ferrari, V. (2021). Sharf: Shape-conditioned radiance fields from a single view. *arXiv preprint arXiv:2102.08860*.

[140] Riklin-Raviv, T., Van Leemput, K., Menze, B. H., Wells III, W. M., and Golland, P. (2010). Segmentation of image ensembles via latent atlases. *Medical image analysis*, 14(5):654–665.

[141] Risholm, P., Balter, J., and Wells, W. M. (2011). Estimation of delivered dose in radiotherapy: the influence of registration uncertainty. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 548–555. Springer.

[142] Risholm, P., Pieper, S., Samset, E., and Wells, W. M. (2010a). Summarizing and visualizing uncertainty in non-rigid registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 554–561. Springer.

[143] Risholm, P., Samset, E., and Wells, W. (2010b). Bayesian estimation of deformation and elastic parameters in non-rigid registration. In *International Workshop on Biomedical Image Registration*, pages 104–115. Springer.

[144] Roberts, D. W., Hartov, A., Kennedy, F. E., Miga, M. I., and Paulsen, K. D. (1998). Intraoperative brain shift and deformation: a quantitative analysis of cortical displacement in 28 cases. *Neurosurgery*, 43(4):749–758.

[145] Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., and Pennec, X. (2017). Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 266–274. Springer.

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

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

[148] Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., and Hammers, A. (2006). Diffeomorphic registration using b-splines. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 702–709. Springer.

[149] Rueckert, D., Frangi, A. F., and Schnabel, J. A. (2003). Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE transactions on medical imaging*, 22(8):1014–1025.

[150] Ruppert, G. C., Chiachia, G., Bergo, F. P., Favretto, F. O., Yasuda, C. L., Rocha, A., and Falcão, A. X. (2017). Medical image registration based on watershed transform from greyscale marker and multi-scale parameter search. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 5(2):138–156.

[151] Sedghi, A., Salomons, G., Jutras, J.-D., Gooding, J., Schreiner, J., Wells III, W. M., and Mousavi, P. (2020). Image registration with deep probabilistic classifiers: application in radiation therapy. In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11315, page 1131509. International Society for Optics and Photonics.

[152] Sedlacek, D. and Zara, J. (2009). Graph cut based point-cloud segmentation for polygonal reconstruction. In *Advances in Visual Computing: 5th International Symposium, ISVC 2009, Las Vegas, NV, USA, November 30-December 2, 2009. Proceedings, Part II 5*, pages 218–227. Springer.

[153] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.

[154] Shah, A., Xing, W., and Triantafyllidis, V. (2017). Reduced-order modelling of parameter-dependent, linear and nonlinear dynamic partial differential equation models. *Proceedings. Mathematical, physical, and engineering sciences*, 473(2200).

[155] Shan, J. and Sampath, A. (2017). Building extraction from lidar point clouds based on clustering techniques. In *Topographic Laser Ranging and Scanning*, pages 421–444. CRC Press.

[156] Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M., and Toga, A. W. (2008). Construction of a 3d probabilistic atlas of human cortical structures. *Neuroimage*, 39(3):1064–1080.

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

[158] Simpson, I. J., Cardoso, M. J., Modat, M., Cash, D. M., Woolrich, M. W., Andersson, J. L., Schnabel, J. A., Ourselin, S., Initiative, A. D. N., et al. (2015). Probabilistic non-linear registration with spatially adaptive regularisation. *Medical image analysis*, 26(1):203–216.

[159] Simpson, I. J., Schnabel, J. A., Groves, A. R., Andersson, J. L., and Woolrich, M. W. (2012). Probabilistic inference of regularisation in non-rigid registration. *NeuroImage*, 59(3):2438–2451.

[160] Simpson, I. J., Woolrich, M. W., Cardoso, M. J., Cash, D. M., Modat, M., Schnabel, J. A., and Ourselin, S. (2013). A bayesian approach for spatially adaptive regularisation

in non-rigid registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 10–18. Springer.

[161] Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. (2013a). A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1219–1222. IEEE.

[162] Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. (2013b). A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In *International Symposium on Biomedial Imaging (ISBI)*.

[163] Staib, L. H. and Duncan, J. S. (1992). Boundary finding with parametrically deformable models. *IEEE transactions on pattern analysis and machine intelligence*, 14(11):1061–1075.

[164] Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E. (2015). Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953.

[165] Sui, Y., Afacan, O., Gholipour, A., and Warfield, S. K. (2020). Slimm: Slice localization integrated mri monitoring. *NeuroImage*, 223:117280.

[166] Toews, M. and Wells, W. M. (2009). Bayesian registration via local image regions: information, selection and marginalization. In *International Conference on Information Processing in Medical Imaging*, pages 435–446. Springer.

[167] Toshev, A., Makadia, A., and Daniilidis, K. (2009). Shape-based object recognition in videos using 3d synthetic object models. In *2009 IEEE conference on computer vision and pattern recognition*, pages 288–295. IEEE.

[168] Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., and Pal, C. J. (2017). Deep complex networks. *arXiv preprint arXiv:1705.09792*.

[169] Vaillant, M., Miller, M. I., Younes, L., and Trouvé, A. (2004). Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23:S161–S169.

[170] Valsecchi, A., Dubois-Lacoste, J., Stützle, T., Damas, S., Santamaria, J., and Marrakchi-Kacem, L. (2013). Evolutionary medical image registration using automatic parameter tuning. In *2013 IEEE Congress on Evolutionary Computation*, pages 1326–1333. IEEE.

[171] van der Niet, T., Zollikofer, C. P., de León, M. S. P., Johnson, S. D., and Linder, H. P. (2010). Three-dimensional geometric morphometrics for studying floral shape variation. *Trends in plant science*, 15(8):423–426.

[172] Vialard, F.-X., Risser, L., Holm, D. D., and Rueckert, D. (2011). Diffeomorphic atlas estimation using karcher mean and geodesic shooting on volumetric images. In *MIUA*, pages 55–60.

[173] Vialard, F. X., Risser, L., Rueckert, D., and Cotter, C. J. (2012). Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241.

[174] Vranic, D. and Saupe, D. (2001). 3d shape descriptor based on 3d fourier transform. In *EURASIP*, pages 271–274.

[175] Wachinger, C. and Golland, P. (2014). Atlas-based under-segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 315–322. Springer.

[176] Wang, J., Wells, W. M., Golland, P., and Zhang, M. (2018). Efficient laplace approximation for bayesian registration uncertainty quantification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 880–888. Springer.

[177] Wang, J., Wells III, W. M., Golland, P., and Zhang, M. (2019a). Registration uncertainty quantification via low-dimensional characterization of geometric deformations. *Magnetic resonance imaging*, 64:122–131.

[178] Wang, J., Xing, J., Druzgal, J., Wells III, W. M., and Zhang, M. (2023). Metamorph: Learning metamorphic image transformation with appearance changes. *arXiv preprint arXiv:2303.04849*.

[179] Wang, J., Xing, W., Kirby, R. M., and Zhang, M. (2019b). Data-driven model order reduction for diffeomorphic image registration. In *International conference on information processing in medical imaging*, pages 694–705. Springer.

[180] Wang, J. and Zhang, M. (2020). Deepflash: An efficient network for learning-based medical image registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4444–4452.

[181] Wang, J. and Zhang, M. (2021). Bayesian atlas building with hierarchical priors for subject-specific regularization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 76–86. Springer.

[182] Wang, J., Zhang, M., et al. (2022). Deep learning for regularization prediction in diffeomorphic image registration. *Machine Learning for Biomedical Imaging*, 1(February 2022 issue):1–10.

[183] Wassermann, D., Toews, M., Niethammer, M., and Wells, W. (2014). Probabilistic diffeomorphic registration: Representing uncertainty. In *International Workshop on Biomedical Image Registration*, pages 72–82. Springer.

[184] Wei, H., Cao, H., Cao, Y., Zhou, Y., Xue, W., Ni, D., and Li, S. (2020). Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 623–632. Springer.

[185] Wells III, W. M., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical image analysis*, 1(1):35–51.

[186] Wong, W. H. and Li, B. (1992). Laplace expansion for posterior densities of non-linear functions of parameters. *Biometrika*, 79(2):393–398.

[187] Wu, N. and Zhang, M. (2023). Neurepdiff: Neural operators to predict geodesics in deformation spaces. *arXiv preprint arXiv:2303.07115*.

[188] Xu, J., Moyer, D., Grant, P. E., Golland, P., Iglesias, J. E., and Adalsteinsson, E. (2022). Svort: Iterative transformer for slice-to-volume registration in fetal brain mri. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 3–13. Springer.

[189] Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L., et al. (2020). Pathological findings of covid-19 associated with acute respiratory distress syndrome. *The Lancet respiratory medicine*, 8(4):420–422.

[190] Yampolsky, M., Salafia, C. M., Shlakhter, O., Haas, D., Eucker, B., and Thorp, J. (2008). Modeling the variability of shapes of a human placenta. *Placenta*, 29(9):790–797.

[191] Yang, C., Rangarajan, A., and Ranka, S. (2018). Visual explanations from deep 3d convolutional neural networks for alzheimer‘‘s disease classification. In *AMIA Annual Symposium Proceedings*, volume 2018, page 1571. American Medical Informatics Association.

[192] Yang, C., Tiebe, O., Pietsch, P., Feinen, C., Kelter, U., and Grzegorzek, M. (2014). Shape-based object retrieval by contour segment matching. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2202–2206. IEEE.

[193] Yang, J. and Duncan, J. S. (2003). 3d image segmentation of deformable objects with shape-appearance joint prior models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–580. Springer.

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

[195] Yang, X. and Niethammer, M. (2015). Uncertainty quantification for lddmm using a low-rank hessian approximation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 289–296. Springer.

[196] Yang, Z., Wang, S., Manivasagam, S., Huang, Z., Ma, W.-C., Yan, X., Yumer, E., and Urtasun, R. (2021). S3: Neural shape, skeleton, and skinning fields for 3d human modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13284–13293.

[197] Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13–es.

[198] Zachiu, C., Ries, M., Ramaekers, P., Guey, J.-L., Moonen, C. T., and de Senneville, B. D. (2017). Real-time non-rigid target tracking for ultrasound-guided clinical interventions. *Physics in Medicine & Biology*, 62(20):8154.

[199] Zhang, C., Zheng, Y., Guo, B., Li, C., and Liao, N. (2021). Scn: a novel shape classification algorithm based on convolutional neural network. *Symmetry*, 13(3):499.

[200] Zhang, D. and Lu, G. (2004). Review of shape representation and description techniques. *Pattern recognition*, 37(1):1–19.

[201] Zhang, M. and Fletcher, P. T. (2014). Bayesian principal geodesic analysis in diffeomorphic image registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 121–128. Springer.

[202] Zhang, M. and Fletcher, P. T. (2015a). Bayesian principal geodesic analysis for estimating intrinsic diffeomorphic image variability. *Medical image analysis*, 25(1):37–44.

[203] Zhang, M. and Fletcher, P. T. (2015b). Finite-dimensional lie algebras for fast diffeomorphic image registration. In *International Conference on Information Processing in Medical Imaging*, pages 249–260. Springer.

[204] Zhang, M. and Fletcher, P. T. (2019). Fast diffeomorphic image registration via fourier-approximated lie algebras. *International Journal of Computer Vision*, 127(1):61–73.

[205] Zhang, M. and Golland, P. (2016). Statistical shape analysis: From landmarks to diffeomorphisms. *Medical Image Analysis*, 33:155–158.

[206] Zhang, M., Liao, R., Dalca, A. V., Turk, E. A., Luo, J., Grant, P. E., and Golland, P. (2017). Frequency diffeomorphisms for efficient image registration. In *International Conference on Information Processing in Medical Imaging*, pages 559–570. Springer.

[207] Zhang, M., Singh, N., and Fletcher, P. T. (2013). Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In *International conference on information processing in medical imaging*, pages 37–48. Springer.

[208] Zhang, M., Wells, W. M., and Golland, P. (2016). Low-dimensional statistics of anatomical variability via compact representation of image deformations. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III 19*, pages 166–173. Springer.

[209] Zhao, S., Wang, Y., Yang, Z., and Cai, D. (2019). Region mutual information loss for semantic segmentation. *Advances in Neural Information Processing Systems*, 32.

[210] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in med-*

ical image analysis and multimodal learning for clinical decision support, pages 3–11. Springer.

[211] Zöllei, L., Jenkinson, M., Timoner, S., and Wells, W. (2007). A marginalized map approach and em optimization for pair-wise registration. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 662–674. Springer.