Fast Learning-based Registration of Sparse 3D Clinical Images

by

Kathleen M. Lewis

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Author .....
Department of Electrical Engineering and Computer Science
May 2 2019

Certified by .....
John V. Guttag
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Certified by .....
Adrian V. Dalca
Instructor of Radiology, Harvard Medical School
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Fast Learning-based Registration of Sparse 3D Clinical Images

by

Kathleen M. Lewis

Submitted to the Department of Electrical Engineering and Computer Science
on May 2 2019, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

We introduce SparseVM, a method to register clinical 3D MR scans faster and more accurately than previously possible. Deformable alignment, or registration, of clinical scans is a fundamental task for many medical image applications such as longitudinal population studies. Most registration algorithms are designed for high-resolution research-quality scans, and under-perform when applied to clinical data. Clinical scans present unique challenges because, in contrast to research-quality scans, clinical scans are often sparse, missing up to 85% of the slices available in research scans. We build on a state-of-the-art learning-based registration method to improve the accuracy of sparse clinical image registration. We evaluate our method on both a clinically-acquired MRI dataset of stroke patients, and on a simulated sparse clinical scan dataset. SparseVM registers MR scans in under a second on a GPU, which is over 1000× faster than the most accurate clinical registration methods, without compromising accuracy. SparseVM has statistically significant improvements in runtimes over all baselines, and a statistically significant improvement in accuracy over the fastest baseline. Because of these contributions, SparseVM enables clinical analyses that were not previously possible. Our code is publicly available at `voxelmorph.mit.edu`.

Thesis Supervisor: John V. Guttag
Title: Professor of Electrical Engineering and Computer Science

Thesis Supervisor: Adrian V. Dalca
Title: Instructor of Radiology, Harvard Medical School

Acknowledgments

I am very thankful for all of the kindness and support I have received since starting MIT.

I would especially like to thank John for teaching me and helping me to grow. He has pushed me to improve in all areas of research while consistently providing moral support. I enjoy talking about healthcare-related applications that motivate us both.

I would also like to thank Adrian for mentoring me and for bringing me into the world of medical imaging. Under his mentorship, I have learned many technical skills and found an area of research I am excited to explore further.

My academic advisor, Vivienne Sze, has also been extremely helpful in guiding me through MIT. I always enjoy our conversations and meetings.

My family is everything to me. I couldn't have made it through without their constant silliness, support, and love.

My friends have provided me with both comic relief and advice. I appreciate that they are always there for me in both times of celebration and in times of hardship.

I am thankful for my labmates. I value their technical feedback during meetings. I also enjoy all of the coffees and late night chats we have had. I love that our lab is both technically rigorous and collaborative.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Deformable medical image registration, or alignment, is a fundamental step in medical image analyses such as longitudinal population studies. Deformable registration establishes a dense, non-linear correspondence between a pair of medical scans. Given two medical scans, registration finds a deformation field that warps one scan to spatially align with the other scan.

Many longitudinal population studies use atlas-based registration. This involves registering each subject scan to an atlas, which is a reference or averaged volume. Both the deformation field and the warped image are used for anatomical analyses in these studies.

The deformation field provides a spatial measurement for anatomical changes that occur between patients or over time. This spatial measurement is able to capture anatomical differences that may not be identifiable when looking at intensity differences between scans.

The warped image itself is used in downstream clinical tasks such as segmentation. Classical segmentation methods often use the atlas and the registered (warped) image to propagate the atlas segmentations to the subject. This is often necessary because subject segmentations are expensive to obtain. Registration is also necessary for segmentation when clinicians only want to segment a particular region of the brain.

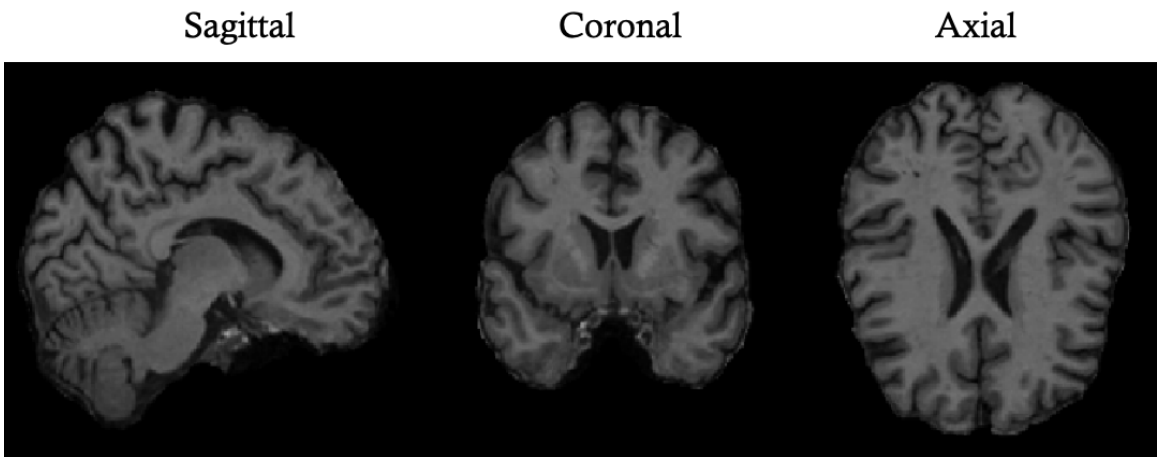


Figure 1-1: 2D MR slices from a research-quality scan in each of the three planes: sagittal, coronal, and axial.

For example, when segmenting brain scans of stroke patients, clinicians might only want to segment white matter hyperintensity within a particular region of the brain. If this region is only known in the atlas, then all subjects must be registered to the atlas to be properly segmented.

There has been extensive research on medical image registration. Classical registration methods require solving a high dimensional optimization on each new pair of scans. These methods are accurate but can take on the order of two hours to register a single pair of scans. Learning-based methods are often fast, but have lower accuracy when applied to clinical images.

Most of these existing registration methods focus on high-resolution research-quality scans (Figure 1-1), that take approximately an hour to acquire for MRI. In contrast, clinical settings often require limited scanning time because of patient safety and financial constraints. For 3D imaging modalities such as MRI, this often means a few 2D slices are acquired instead of a dense 3D volume. Each 2D slice can be high resolution (Fig. 1-2, left), but the 3D volume is often missing up to 85% of the slices that are typically available in a research-quality scan. The wide spacing between slices causes drastic discontinuities in anatomy between neighboring slices (Fig. 1-2, right). Furthermore, the anatomy captured by the slices is not consistent

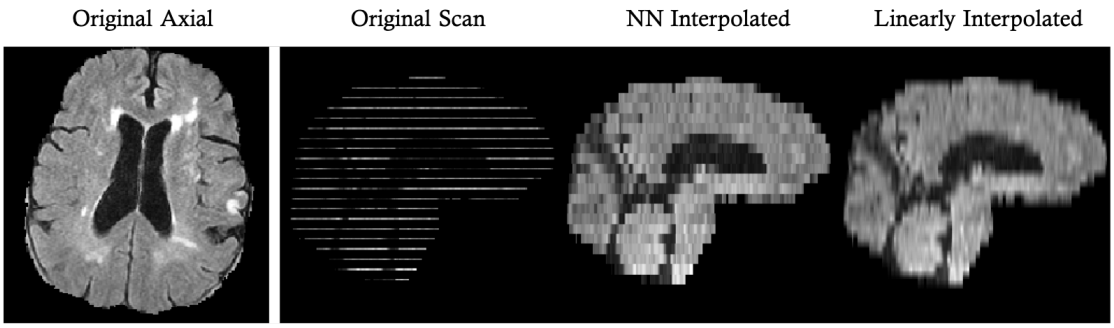


Figure 1-2: Example stroke subject scan with corresponding interpolated images. The subject scan has approximately 15% of the slices normally available in a full resolution research quality scan. Shown from left to right are an axial slice from the original scan, a sagittal slice from the original subject slices, a sagittal slice from the nearest neighbor (NN) interpolated image, and a sagittal slice from the linearly interpolated image.

across different scans. Depending on the orientation of the patient with respect to the scanner, different anatomy may be present in the acquired slices.

With existing registration methods, researchers and clinicians have to make a trade-off between speed and accuracy when registering clinical scans. The registration methods designed for clinical data take on the order of two hours to register a pair of scans, while the fast learning-based methods are not sufficiently accurate on clinical scans. In this thesis, we present a new registration method, SparseVM, that eliminates this trade-off. We build upon an open-source unsupervised learning-based method, VoxelMorph, to achieve very fast and accurate registration of both sparse clinical scans and research scans.

1.2 Contributions

In this section, we outline and describe our contributions. Our work

1. introduces a new generalized loss function, sparse local cross correlation (SLCC)
2. achieves state-of-the-art accuracy for registering clinical images
3. achieves state-of-the-art speed for registering clinical images
4. provides open source code, available at `voxelmorph.mit.edu`

Our new loss function, sparse local cross correlation (SLCC), generalizes beyond the clinical registration of MR brain scans shown in this thesis. SLCC takes in two masks and weights the contributions of each input in proportion to values in the respective masks. This loss function can therefore be used for in any application where sparsity (binary values) or confidence in observations (continuous values) can be defined by masks.

We evaluate SparseVM on a clinical study of stroke patients containing T2-FLAIR MR brain scans [34] and on a sparse T1 MR brain dataset simulated from the ADNI dataset [29]. We show that our method has statistically significant improvements over existing methods for both accuracy and speed. SparseVM is 1) more than 1000× faster than the best classical methods without losing accuracy and 2) more accurate than current learning-based methods. SparseVM therefore facilitates population analyses that were previously infeasible, and promises to extend to uses in the clinic.

1.3 Thesis Outline

The thesis layout is as follows. Chapter 2 discusses the problem and related work. Section 2.2 outlines the problem statement and terminology. Section 2.3 discusses the related work and provides a detailed overview the method on which we build, VoxelMorph. Section 2.4 presents our method, Sparse VoxelMorph (SparseVM). Chapter 3 discusses the data and experimental setup in Section 3.1, and the results in Sections 3.2 and 3.3. Chapter 4 summarizes the contributions of this thesis (Section 4.1) and the future work.

Chapter 2

Registration of 3D MR Clinical Scans

2.1 Magnetic Resonance Imaging Background

Magnetic Resonance Imaging (MRI) is an imaging modality that uses magnetic fields to acquire evenly spaced 2-dimensional anatomical slices. These 2-dimensional slices are stacked to form a 3-dimensional volume. MRI is commonly used to diagnose disease and is advantageous because it can be used to visualize internal anatomy without ionizing radiation. The physics of MRI can be adjusted depending on the type of tissue being imaged. The intensities of types of tissues appear differently depending on the type of MRI. In this work, we use two brain MR datasets (see subsection 3.1.2), which contain T1-weighted MR images and T2-FLAIR MR images respectively. High resolution brain MR images are often acquired for research purposes, but take on the order of an hour for full acquisition. These images, known as isotropic images, have sharply defined features and small spacing in between slices. MR images acquired in clinical settings, on the other hand, often require short acquisition times because of financial constraints and patient safety concerns. Short acquisition times lead to sparse sampling, which results in high resolution 2D slices that are spaced far apart. This leads to a sparse 3D scan that contains roughly 15% of the 2D slices of an isotropic scan. The anatomy present in the sparsely acquired slices differs across patients and differs across time for a single patient because of different patient orientations with respect to the scanner.

2.2 Problem Definition

2.2.1 Terminology

Let f, m (for fixed and moving) be two volumes defined over a n-D spatial domain $\Omega \subset \mathbb{R}^n$. In this thesis, we focus on 3D volumes (n=3). For the remainder of the work, we use image and volume interchangeably to refer to 3D volumes. For simplicity, we assume f and m contain grayscale data and are affinely aligned as a pre-processing step. We focus on atlas-based registration, in which a full resolution atlas, or a reference volume, m is registered to each clinical subject scan f; this is commonly used in population analysis [38]. For atlas-based registration, it is common practice to use linear interpolation to make the clinical scan the same resolution as the atlas.

2.2.2 Clinical Image Registration Task

Deformable registration establishes a dense, non-linear correspondence between a pair of n-D scans. As discussed in Chapter 1, deformable registration is a fundamental step in medical image analyses, such as population studies, and has been an active area of research for decades. Registration is important when measuring how changes in the brain correlate with factors such as age. One example of this is monitoring changes in the spatial distribution of white matter hyperintensity to study its correlation with stroke [34].

In this work, we focus on 3D MRI brain scans. For each clinical scan, f, our method aims to find a spatially smooth registration field ϕ such that the atlas warped by the registration field, $w_m = m \circ \phi$, is as similar to f as possible for some similarity metric $L_{sim}(f, w_m)$. We constrain ϕ to be spatially smooth according to $L_{smooth}(\phi)$, such that neighboring voxels move together.

$$L(f, m, \phi) = L_{sim}(f, m \circ \phi) + \lambda L_{smooth}(\phi), \tag{2.1}$$

In general, the registration field ϕ can take many forms. In this work, we model ϕ as a vector displacement field. This gives the vector movement for each voxel in m. Given

a linearly-interpolated clinical image f and an atlas m, our method aims to find the vector displacement field ϕ that minimizes the function 2.1 for a fixed regularization hyperparameter λ.

2.3 Related Work

2.3.1 Registration of Isotropic Images

There is extensive literature on 3D medical image registration. In this subsection, we provide a brief overview of classical non-learning-based registration methods and learning-based registration methods. Most of these registration methods are designed for full-resolution scans rather than sparse clinical scans.

Successful classical methods include elastic-type models [7, 36, 16], statistical parametric mapping [3], free-form deformations with b-splines [35], discrete methods [20], and Demons [39, 32]. In addition to these methods, there are also successful classical methods that use diffeomorphic transformations [2, 4, 10, 28, 41]. Diffeomorphic transforms are transforms that explicitly enforce topology constraints.

All of these non-learning based methods require solving a high dimensional optimization on each new pair of scans, and therefore have slow runtimes. Moreover, applying these methods to clinical scans is problematic because they were developed to work on high resolution scans and require spatial continuity for smooth image gradients, which are not available in clinical scans.

Learning-based registration methods use neural networks to learn a function that takes in two scans as input and outputs a deformation field. Learning-based methods are fast, but until now have been designed for high-resolution scans.

Many of these methods are supervised and require ground truth warp fields [33, 37, 42, 11]. While supervised learning-based methods are promising approaches to registration, ground truth warp fields can be time-consuming and difficult to obtain.

Unsupervised methods are therefore a desirable alternative. Two recent unsupervised methods [17, 27] use a convolutional neural network (CNN) and a spatial

transformation function [24] to register images. [17] is limited to small transformations and has only been demonstrated on a limited subset of images such as 2D slices. Similarly, [27] has only been demonstrated on 3D subregions.

VoxelMorph (VM) [8, 9, 13] is a more recent unsupervised learning-based method that achieves state-of-the-art accuracy on 3D isotropic scans. We build on VoxelMorph to accurately register clinical scans. VoxelMorph uses a convolutional neural network to learn a function, $g_\theta(f, m) = \phi$, to compute a deformation field ϕ for a pair of given scans $\{f, m\}$. At test time, VoxelMorph evaluates this function on two input scans and outputs the deformation field ϕ, as well as the registered image, $w_m = m \circ \phi$ (m warped by ϕ). VoxelMorph was developed for high resolution isotropic scans. In this thesis we present a method based on VoxelMorph to address the problem of sparsity in clinical scans.

2.3.2 Registration of Clinical Images

Clinical image registration has many challenges that are not present in isotropic image registration, as mentioned in section 2.1. Many of the existing registration methods perform worse on clinical images than on research images because of the wide spacing between slices in clinical images. This wide spacing creates discontinuities in anatomy from neighboring slices and violates smoothness assumptions needed for gradient-based methods [22]. Linear interpolation is usually used to create a full-resolution image, but the interpolated voxels often contain inaccurate information. Some methods handle this by hand-tuning parameters and pipelines to a particular database [38]; however since these methods do not generalize well, they are not easily adaptable to clinical applications.

Feature-based methods [25, 30, 40] and point set representations [19] can be used to extract representative features or sets of voxels; however, it is difficult to extract consistent features because of the wide spacing and inconsistencies in anatomy present between different scans.

Discrete registration methods minimize an energy function of a Markov Random Field (MRF) to find optimal registration, and are another alternative to gradient-

based methods [21, 23, 31]. A recent patch-based registration (PBR) method [15] builds on previous discrete registration methods and adapts them to sparse data by using 3-dimensional patches to measure image intensity similarities. This method incorporates a mask that weights voxel contributions to the loss function in proportion to confidence in those voxels. This differentiates the weighting of interpolated voxels from the weighting of voxels acquired from actual slices. While PBR is a flexible, accurate method for registering clinical methods, it takes on the order of two hours to register two images. This means using it to register a large database of clinical images is infeasible. In this thesis, we introduce a fast learning-based method that builds on the mask weighting idea in PBR.

2.3.3 Loss Functions

We first describe the loss functions used in VoxelMorph as well as other registration methods, and in the next section we propose loss functions designed for clinical data.

Most medical image registration methods optimize a loss function of the form:

$$L(f, m, \phi) = L_{sim}(f, m \circ \phi) + \lambda L_{smooth}(\phi), \tag{2.2}$$

where λ is a regularization hyperparameter. This loss function balances an image matching term L_{sim} with a regularization or spatial smoothness term L_{smooth}. The first term measures the similarity in appearance between the fixed image f and the warped image w_m. The second term encourages the displacement field ϕ to be smooth and anatomically realistic. This loss is used in learning-based methods to learn the global network parameter θ, where $\phi = g_\theta(\cdot, \cdot)$.

Intensity mean-squared error (MSE) and cross correlation (CC) are the most commonly used similarity measures. MSE compares intensities at each voxel:

$$MSE(f, w_m) = \frac{1}{|\Omega|} \sum_{p \in \Omega} [f(p) - w_m(p)]^2, \tag{2.3}$$

where the notation $f(p)$ indicates the value of image f at voxel location p.

Local cross correlation has been shown to be more invariant than MSE to differences in intensity distributions across scans [5]. For scans f and m, let $\mu_f(p)$ and $\mu_{w_m}(p)$ be the local neighborhood intensity mean around voxel p:

$$\mu_f(p) = \frac{1}{n^3} \sum_{p_i} f(p_i), \tag{2.4}$$

where n^3 is the neighborhood size and p_i iterates over the n^3 volume of voxels surrounding voxel p. The local cross correlation is then:

$$LCC(f, w_m) = \sum_{p \in \Omega} \frac{[\sum_{p_i}(f(p_i) - \mu_f(p))(w_m(p_i) - \mu_{w_m}(p))]^2}{[\sum_{p_i}(f(p_i) - \mu_f(p))^2][\sum_{p_i}(w_m(p_i) - \mu_{w_m}(p))^2]}. \tag{2.5}$$

The L_{smooth} loss penalizes the local spatial gradients of u, $\bigtriangledown u$, where u is the vector displacement such that $\phi = Id + u$ and Id is the identity transform:

$$L_{smooth}(\phi) = \sum_{p \in \Omega} || \bigtriangledown u(p)||^2. \tag{2.6}$$

The original VoxelMorph papers present several different architectures that trade off between accuracy and runtimes. All of the presented architectures are encoder-decoders with skip connections and consist of 3D convolutions followed by Leaky ReLu activation functions. The registration field warps m to $w_m = m \circ \phi$ using a spatial deformation function, which is described in detail in the original papers. In our experiments, we use their VM-1 architecture. The VM-1 architecture has encoding features of sizes $[16, 32, 32, 32]$ and decoding features of size $[32, 32, 32, 32, 8, 8, 3]$.

In the next section, we discuss how our method, SparseVM, builds on VoxelMorph to more accurately register sparse clinical images. VoxelMorph has only been evaluated on high resolution images, in which all voxels are equally reliable, and therefore contribute equally to the loss function during training. When used with clinical images which contains linearly interpolated voxels, VoxelMorph does not perform as accurately as with isotropic images. We hypothesize that this is because it does not differentiate

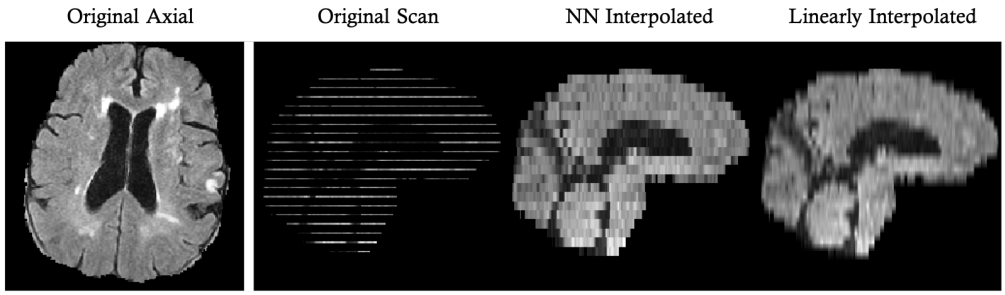


Figure 2-1: Example stroke subject scan with corresponding interpolated images. The subject scan has approximately 15% of the slices normally available in a full resolution research quality scan. Shown from left to right are an axial slice from the original scan, the original subject slices seen from a sagittal view, nearest neighbor (NN) interpolated image, and linearly interpolated image.

between unobserved and observed voxels. This discrepancy motivates our method, which differentiates interpolated voxels from acquired voxels.

2.4 Method

We propose two new image similarity losses that generalize Eq. (2.3) and Eq. (2.5), and lead to accurate registration on both clinical and high resolution scans. We upsample each clinical image to isotropic spatial resolution (equal resolution in all planes) by linearly interpolating between the acquired slices (Fig. 2-1, right).

Our new loss functions use masks, m_f and m_m, corresponding to the fixed image f and the moving image m, respectively. In its simplest form, each voxel in a mask has a value of 0 (interpolated voxel) or 1 (acquired voxel). We expand on other forms in Section 3.1.2. We let $m_c = m_f * m_m$ be a combined mask that indicates voxels that are observed in both scans, where $*$ denotes element-wise multiplication.

We define sparse mean squared error (SMSE) which generalizes Eq. (2.3) by calculating the loss over only observed voxels:

$$SMSE(f, m_c, w_m) = \frac{1}{\sum m_c(p)} \sum_{p \in \Omega} [f(p) - w_m(p)]^2 m_c(p). \tag{2.7}$$

The implementation of sparse local cross correlation (SLCC) is less straightforward, since it requires careful treatment of correlating the sparsely observed voxels. For each image, we first compute the local mean of each neighborhood over only the observed voxels:

$$\mu_f(p) = \frac{1}{\sum_{p \in n^3} m_f(p)} \sum_{p_i \in n^3} f(p_i) m_f(p_i). \tag{2.8}$$

We define $\hat{f}(p_i, p)$ at voxel p_i belonging to a neighborhood centered at p as the image with the local mean subtract out:

$$\hat{f}(p_i, p) = f(p_i) - \mu_f(p). \tag{2.9}$$

We then define the sparse local cross correlation as:

$$SLCC(f, m_c, w_m) = \sum_{p \in \Omega} \frac{[\sum_{p_i \in n^3} m_c(p_i) \hat{f}_i(p_i, p) \hat{w}_{mi}(p_i, p)]^2}{[\sum_{p_i \in n^3} m_c(p_i) \hat{f}_i(p_i, p)^2][\sum_{p_i \in n^3} m_c(p_i) \hat{w}_{mi}(p_i, p)^2]}, \tag{2.10}$$

where, as before, $i \in n^3$ indicate a voxel i in the neighborhood n^3 centered around p.

SLCC computes the local correlation in each neighborhood between voxels that are observed in both scans. Since SLCC does not use values over unobserved (interpolated) voxels, the neighborhoods are sparse. As in regular local cross correlation, the neighborhood size n^3 is a hyperparameter. In SLCC, special care must be taken to ensure the neighborhoods are large enough to contain enough non-zero voxels, but small enough to capture specific regions of the scans.

The smoothness constraint on the deformation field remains the same (Eq.2.6) regardless of whether the data is sparse or not.

Chapter 3

Experiments and Results

3.1 Experiments

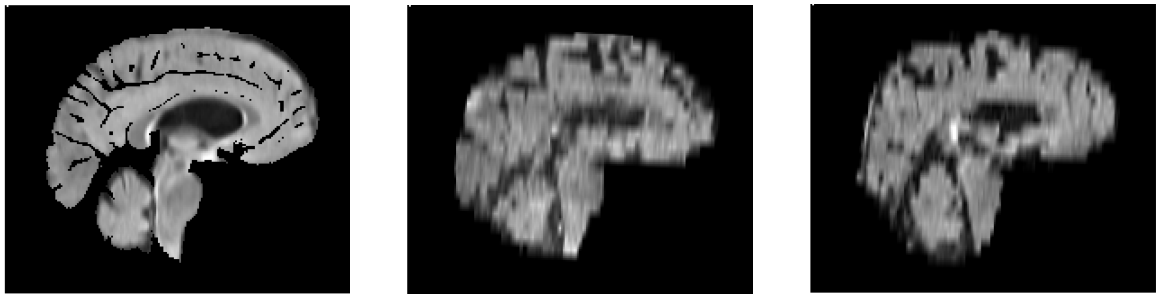
3.1.1 Evaluation Metric

We evaluate our performance using the Dice metric [18]. Dice measures the volume overlap of anatomical segmentations, with 0 representing no overlap and 1 representing perfect overlap. We evaluate Dice in the original subject space; this means the Dice score is only calculated over acquired, not interpolated, voxels.

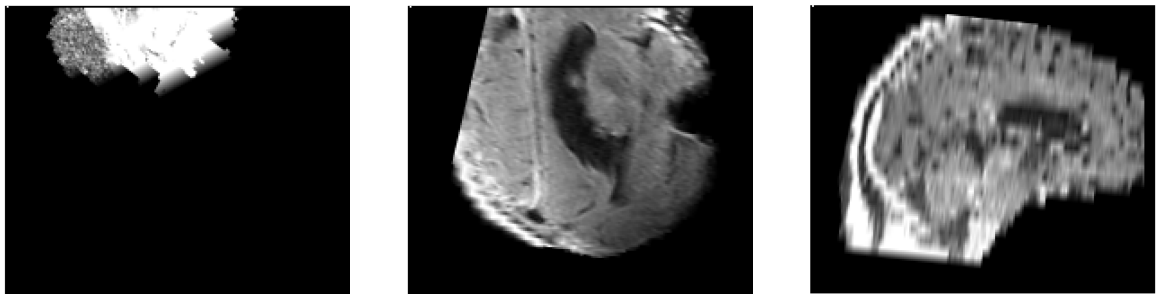
3.1.2 Datasets

We evaluate our method on atlas-based registration for two different MRI datasets. For each, we use a full-resolution atlas computed from an external dataset [38]. During pre-processing the scans were affinely-aligned to the atlas, and skull-stripped using an existing toolbox [38].

The first dataset is a clinically acquired stroke dataset. These T2-FLAIR images were acquired within 48 hours of stroke onset, and were used in a study aiming to quantify white matter disease burden [34]. The in-plane resolution is 0.85 mm and the spacing between slices is 5-7mm. The images were padded to be 160x192x160. The dataset contains manual segmentations of the lateral ventricles for 104 of the subjects. The 237 subjects without manual segmentations were used for training, and



(a) Atlas (left) and two linearly interpolated stroke scans with good affine alignment (center and right)



(b) Three examples of linearly interpolated stroke scans that failed affine alignment or skull-stripping pre-processing

Figure 3-1: Examples of good and bad pre-processing for linearly interpolated stroke scans. The wide spacing between acquired scans makes it more difficult to use existing pre-processing tools for the stroke dataset.

the 104 subjects with manual segmentations were split into validation (29 subjects) and test (75 subjects) sets. For evaluation, Dice is averaged over these large ventricle structures for each subject in the test set.

Because of the low resolution of clinical scans, many images failed one or both parts of pre-processing to varying extents (Fig. 3-1). Subjects that fail skull-stripping are defined as subjects that still have part or all of the skull remaining after pre-processing. Subjects that fail affine alignment are defined as subjects that have a low affine Dice score for the ventricles. We kept the training subjects with the top 70% of affine Dice scores. This was chosen because it brought the average affine Dice score over the dataset to approximately the same average as the isotropic dataset in the original VoxelMorph papers. We evaluate on all test subjects regardless of affine Dice score.

The atlas m in our atlas-based registration experiments is isotropic, and therefore m_m consists of value 1 at each voxel. In contrast, f has approximately 15% of the

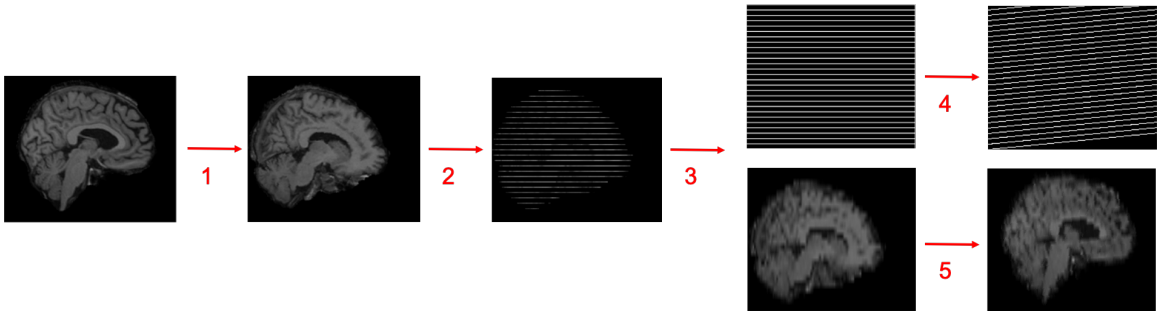


Figure 3-2: Steps to create simulated sparse ADNI dataset: 1) Apply random affine transform, 2) Delete 6 out of every 7 slices, 3) linearly interpolate between slices and create mask indicating acquired voxels from interpolated voxels, 4) and 5) apply inverse of affine transform to the mask and interpolated image.

number of acquired slices of an isotropic research scan, and we linearly interpolate the remaining slices to achieve the same resolution as the atlas. The subject mask $m_f(p)$ is one for voxels p in the acquired slices and zero for voxels p in the interpolated slices. To perform affine alignment, both f and m_f are resampled, and voxels are therefore linearly interpolated. The linear interpolation results in continuous values in the range $[0, 1]$ for the mask m_f. We interpret these values as weightings of confidence where voxels with low values contain less information from the acquired slices.

The second dataset contains isotropic T1-weighted brain scans from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) [29] dataset. We simulate realistic sparse clinical scans as shown in Figure 3-2. The resulting images are 160x192x224, and contain approximately 14% of the acquired voxels in the isotropic image. The number of slices deleted was chosen to emulate the wide spacing in the clinically acquired stroke dataset. This provides us with a realistic, challenging simulated clinical scan dataset with the advantages of having ground truth isotropic 3D scans and segmentations of 30 different structures in the brain. This allows us to analyze our performance on anatomy of different sizes.

3.1.3 Baselines

For the clinically acquired stroke dataset, we compare our results against VoxelMorph, PBR, and the Advanced Normalization Tools software package (ANTs) [6]. PBR, as

mentioned in Section 2.3.2, is a patch-based discrete registration method that explicitly accounts for sparsity in clinical data. This method is the current state-of-the-art for sparse clinical image registration. We compare to ANTs since it is a popular registration method and is used as a baseline in the original VM and PBR papers. For all baseline methods, we used the optimal parameters given in the original publications.

For the ADNI dataset, we only compare against VoxelMorph since it is the only baseline that runs in a feasible amount of time when registering large datasets. This comparison between SparseVM and VoxelMorph on the simulated data allows us to directly compare the performance of our sparse versions of mean squared error and cross correlation against the typical versions of these loss functions. By analyzing the performance of these two learning-based methods, we can better understand when the sparse versions of these loss functions are beneficial relative to their isotropic counterparts.

3.1.4 Implementation

We implement SparseVM using Keras [12] with a TensorFlow backend [1]. We use the ADAM optimizer [26] with an epsilon value of $5e{-}5$ and a learning rate of $5e{-}4$. We use the architecture proposed in the original VoxelMorph paper [8], which is publicly available online. For SLCC, we use sparse convolutional layers [14].

For the smoothness regularization hyperparamter λ, we use the value used in the original VoxelMorph papers to guide our parameter search. We then use the validation set to choose the best parameter value. We implement SLCC with a neighborhood size $n{=}15$, which we found to perform well given the sparsity in our data. This is larger than the value ($n{=}9$) used in the VoxelMorph paper.

Method	Average (SD)	Median (MAD)	GPU sec	CPU sec
ANTs (optimized)	0.696 (0.070)	0.720 (0.050)	-	9059 (2023)
PBR (best)	0.732 (0.071)	0.749 (0.052)	-	9269 (5134)
VoxelMorph CC15	0.723 (0.077)	0.748 (0.060)	0.283 (0.028)	37.1 (2.78)
SparseVM SMSE	0.719 (0.075)	0.740 (0.058)	0.303 (0.047)	41 (0.584)
SparseVM SLCC	**0.743 (0.076)**	**0.768 (0.060)**	**0.241 (0.030)**	**32.7 (0.452)**

Table 3.1: Dice scores and test runtimes for ANTs, PBR, VoxelMorph, and SparseVM on the stroke dataset. The average Dice score and median Dice score are computed for all test subjects on the left and right lateral ventricle labels. The standard deviation and median absolute deviation are shown in parentheses. The runtimes are calculated after pre-processing. SparseVM using SLCC is both more accurate and faster than the baselines.

3.2 Stroke Results

3.2.1 Quantitative Comparison with Baselines

Table 3.1 summarizes the average and median Dice scores for all test subjects on the ventricles, and the average test runtimes of each method. SparseVM has 1) the highest mean and the highest median Dice scores and 2) the fastest runtime on both CPUs and GPUs.

Both ANTs and PBR take on the order of two hours on a CPU, while VoxelMorph and SparseVM take less than a minute on a single-threaded CPU and less than a second on a GPU. SparseVM's improvement in runtime is statistically significant over the most accurate baseline, PBR, with a p-value of $3.76e{-}17$ using a paired t-test. There are no GPU implementations of ANTs or PBR. SparseVM's improvements over VoxelMorph in runtime on both a CPU and a GPU is statistically significant with p-values of 1.88e-23 and 1.24e-10 respectively.

Since the registration results for different subjects exhibit high variance, as shown by the standard deviations, we also show the Dice for each subject in Figure 3-3. SparseVM using SLCC outperforms ANTs on 86.7% of test subjects, PBR on 69.3% of test subjects, and VoxelMorph on 90.7% of test subjects. We perform additional analyses on the learning-based registration methods VM and SparseVM, which are the only methods feasible for large-scale analyses involving hundreds or thousands of

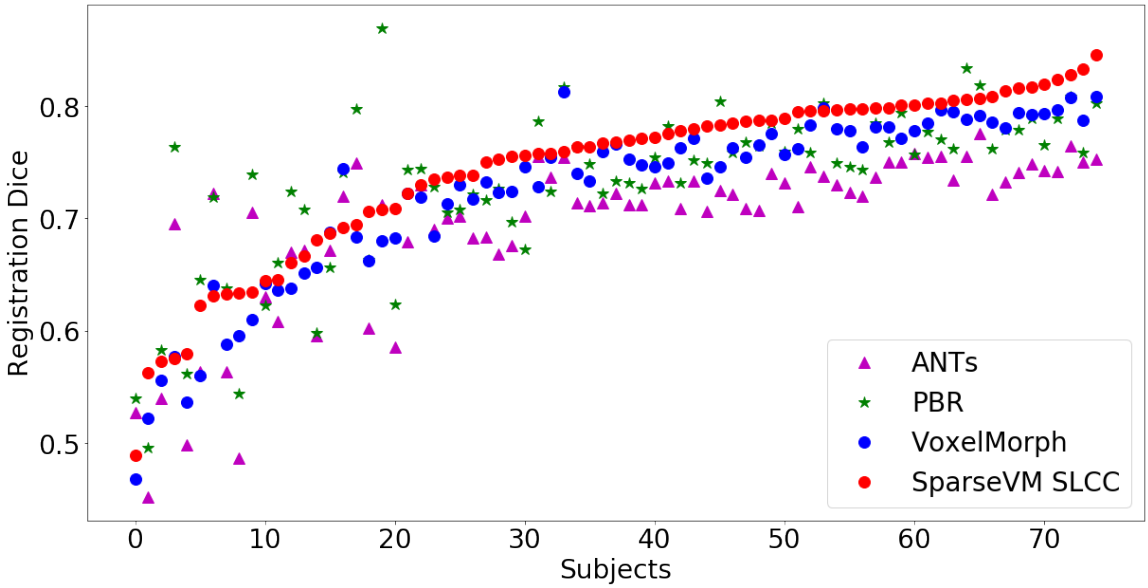


Figure 3-3: Dice score for each subject sorted by SparseVM performance. SparseVM using SLCC is better than ANTs on 86.7% of test subjects, better than PBR on 69.3% of test subjects, and better than VoxelMorph on 90.7% of test subjects.

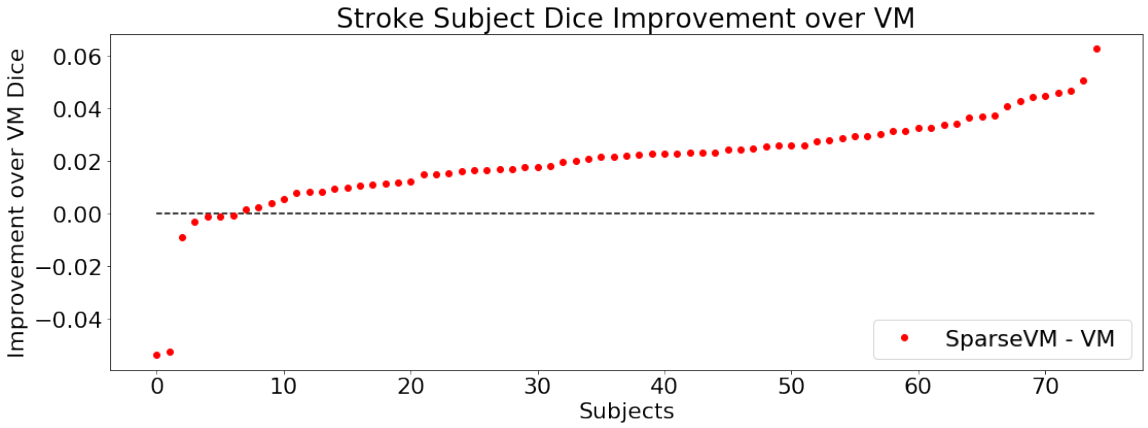


Figure 3-4: Dice improvement of SparseVM results over VoxelMorph results per subject sorted in increasing order of improvement. SparseVM is on average 2 Dice better than VoxelMorph. Upon visual inspection, we find that the first two subjects were aggressively skull-stripped such that part of the brain was removed.

scans. We plot the subject-wise difference in Dice accuracy between SparseVM and VoxelMorph in Figure 3-4. The pairwise improvement in accuracy of SparseVM is statistically significant with a p-value of $1.60e{-}14$ using a paired t-test. SparseVM yields an average improvement of two Dice points over VoxelMorph.

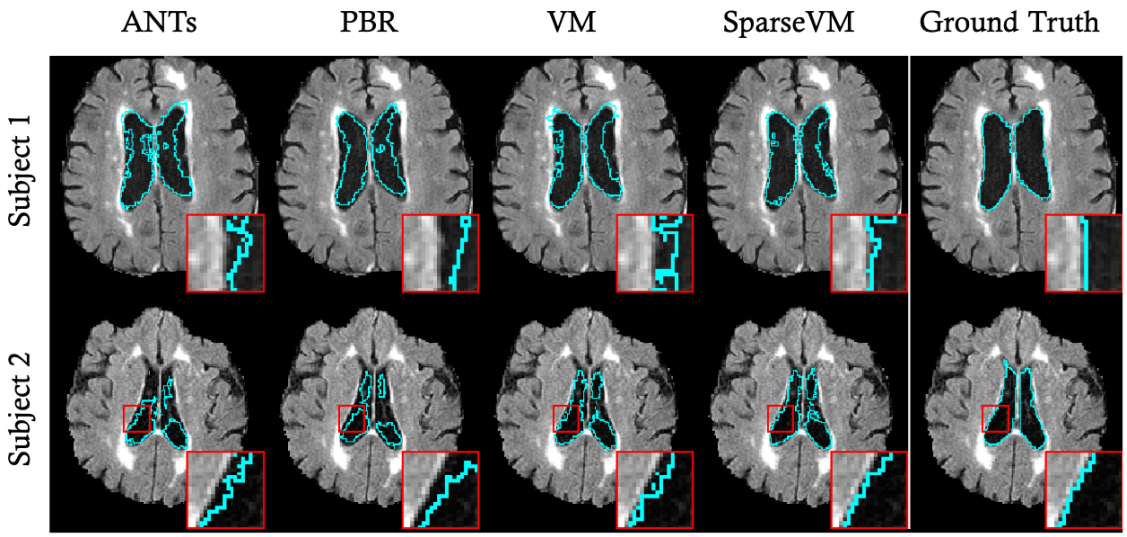


Figure 3-5: Example results. Warped atlas segmentations overlaid on the subject scan for subjects with different sized ventricles.

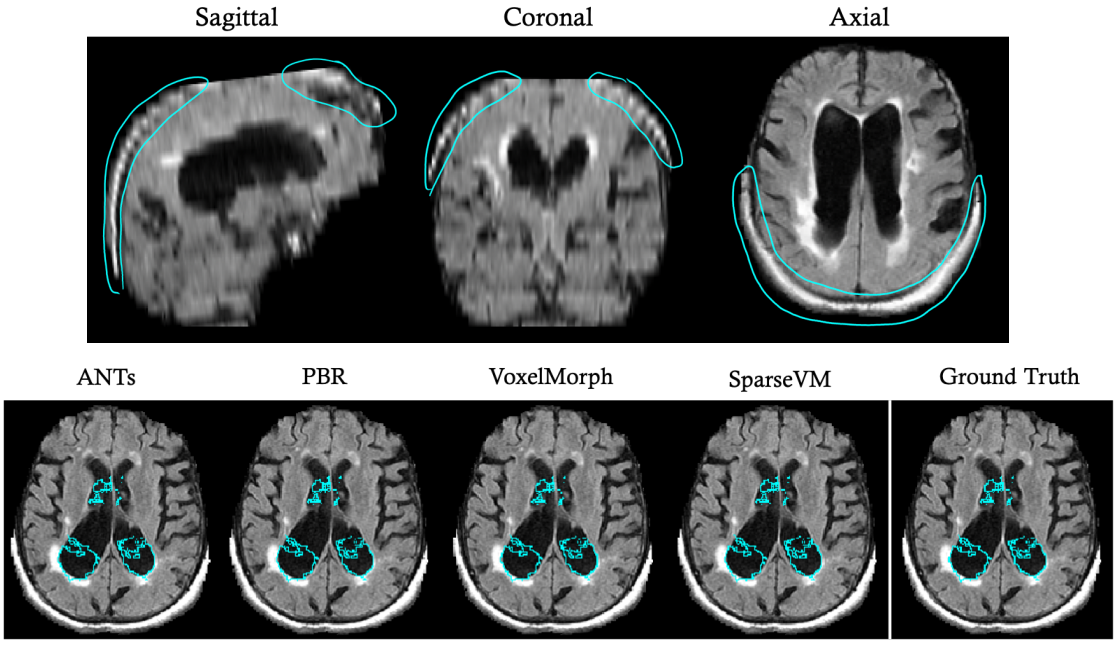


Figure 3-6: Example test subject that was not properly skull stripped during preprocessing. The top row highlights examples of areas where skull is present in the scan. The bottom row shows the warped atlas segmentations for each method. The PBR segmentations include most of the ventricles, while the other methods only segment part of the ventricles.

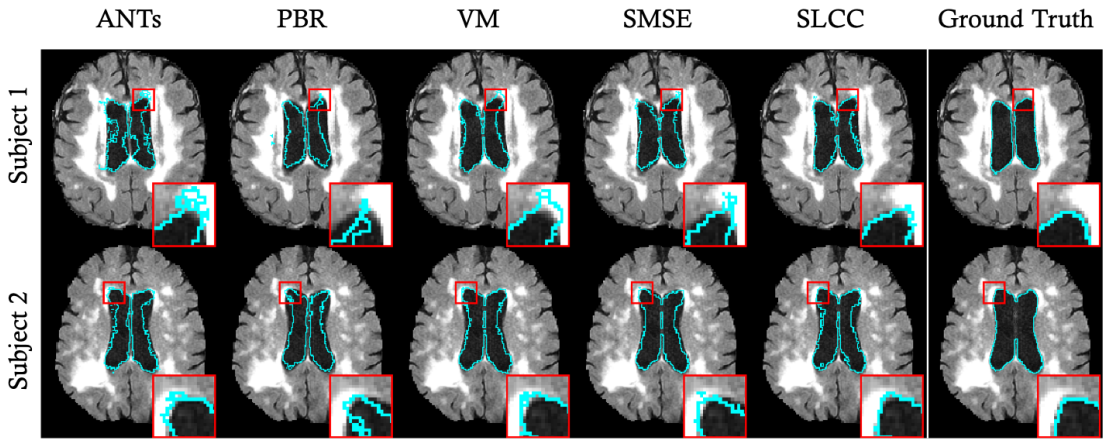


Figure 3-7: Example results. Warped atlas segmentations overlaid on the subject scan for subjects with high WMH burden.

3.2.2 Qualitative Analysis

Figure 3-5 shows registration results for several test subjects. This visualization enables a closer examination of the results. By evaluating specific regions of the brain, we observe that SparseVM provides registration that enables, for example, the ventricle segmentations to better snap to the actual anatomical boundaries than the other methods' segmentations. Consistent with our quantitative results, the close-up regions show that warped atlas segmentations using PBR, VM, and SparseVM outperform the popular toolbox ANTS, with SparseVM (SLCC) performing best among these methods.

We also examine subjects for which SparseVM using SLCC performs worse than a baseline. By visualizing such subjects in a manner similar to Figure 3-5, we find that SparseVM using SLCC tends to perform worse than PBR on subjects for which skull-stripping or affinely alignment fails during pre-processing. Figure 3-6 shows an example of a test subject that has not been properly skull stripped. Better affine alignment and skull-stripping algorithms for clinical data are current areas of research.

The original study that resulted in the stroke dataset investigated white matter hyperintensity (WMH) burden in stroke patients [34]. WMH is an important predictor of future cardiovascular health and is visible in the T2 FLAIR scans. However, differences in intensities caused by WMH add to the difficulty of clinical registration.

Method	Average (SD)	Median (MAD)
VoxelMorph CC15	0.714 (0.034)	0.723 (0.019)
SparseVM SLCC	**0.721 (0.023)**	**0.728 (0.017)**

Table 3.2: Dice scores for VoxelMorph and SparseVM on the simulated sparse ADNI dataset. The average Dice score and median Dice score are computed for all test subjects over 30 different structures. The standard deviation and median absolute deviation are shown in parentheses.

Therefore we also visually analyze the registration performance of SparseVM near regions of significant WMH burden. We find that SparseVM using SLCC leads to better performance than SparseVM using SMSE for subjects with high WMH, likely because of the robustness of local cross correlation to intensity heterogeneity. Figure 3-7 shows example subjects with high WMH burden. In most areas of the brain, SparseVM performs comparably or much better than the other methods.

3.3 ADNI Results

3.3.1 Dice Comparison with VoxelMorph

Table 3.2 shows the average and median Dice scores for both VoxelMorph and SparseVM with the cross correlation loss function. SparseVM outperforms VoxelMorph in both mean and median Dice scores. While the statistics for the stroke dataset shown in Table 3.1 were computed over the ventricle structures, the statistics for the simulated ADNI dataset were computed over 30 different anatomical structures. In both cases, we observe a similar trend in which the sparse cross correlation function outperforms the normal cross correlation function.

We believe that the average Dice for the simulated ADNI dataset is lower than that of the stroke dataset because the simulated ADNI dataset Dice is averaged over several smaller anatomical structures in addition to the larger ventricle structures. Figure 3-8 shows a boxplot of the average Dice per label for both SparseVM and VoxelMorph. SparseVM has an higher average for 14 of the 17 structures. VoxelMorph performs better on three of the smallest structures: amgydala, choroid plexus, and the third

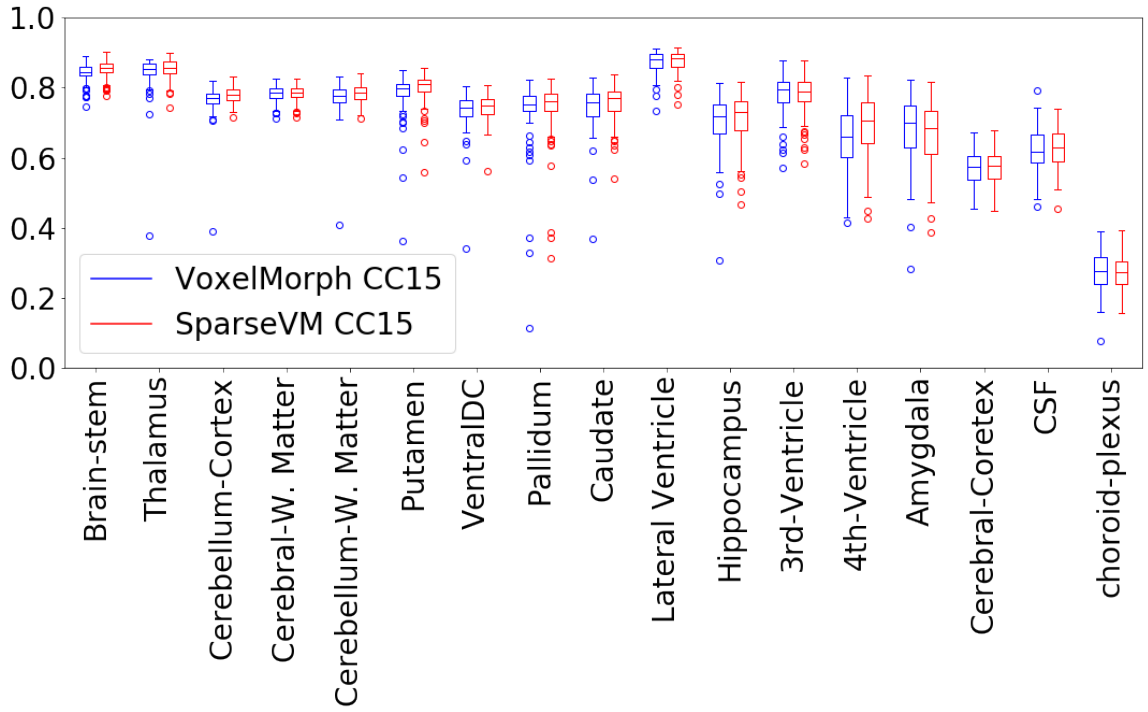


Figure 3-8: Boxplot of the Dice for all 17 different anatomical structures. For easier visualization, the left and right side of each anatomical structure (e.g. the left and right lateral ventricles) were averaged together into one boxplot. SparseVM has a higher average for 14 out of the 17 anatomical structures.

ventricle. Since these structures are significantly smaller than the other structures in terms of number of voxels present, we predict that there is not enough information from the acquired slices to inform SparseVM. Interpolated voxels, even if they are not completely accurate, may help guide the model in the case of small structures that have few voxels observed in both images f and m. The improvement of VoxelMorph is statistically significant only for the amygdala when using a paired t-test (p-value is 0.011). The improvement of SparseVM over VoxelMorph is statistically significant for 12 of the 17 anatomical structures.

Figure 3-9 shows the average Dice per subject and is sorted in increasing order of the SparseVM Dice. SparseVM performs better than VoxelMorph on 88% of test subjects. The improvement is statistically significant when using a paired t-test (p-value is 0.002). Figure 3-10 shows our method's Dice improvement over VoxelMorph. SparseVM is on average 0.007 Dice points higher than VoxelMorph on each subject.

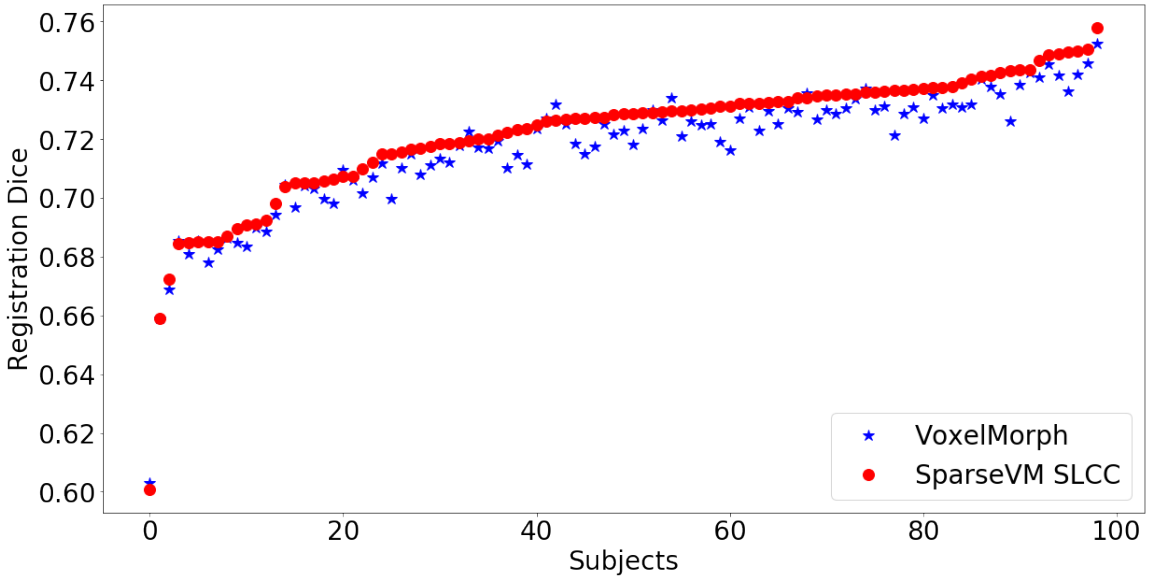


Figure 3-9: Per subject Dice averaged over all 30 structures. The subjects are in increasing order of the SparseVM Dice.

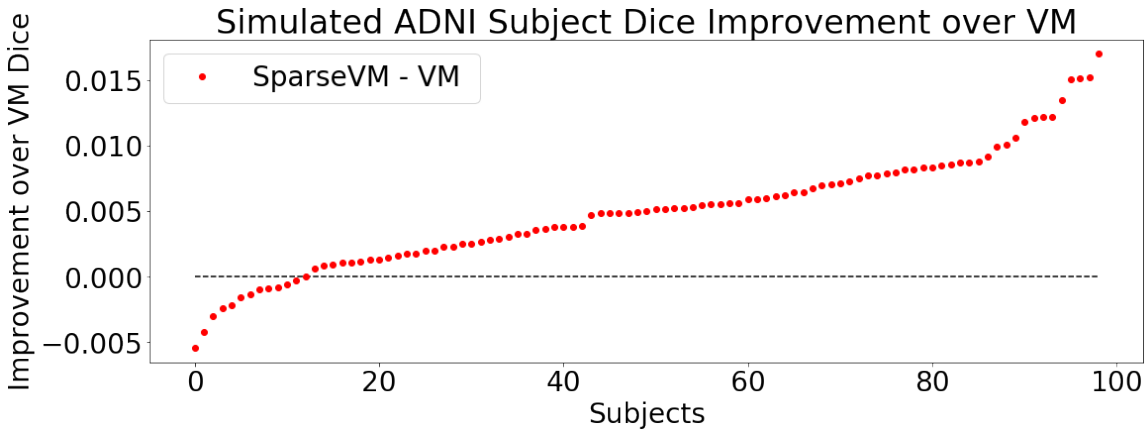


Figure 3-10: SparseVM's average Dice improvement over VoxelMorph. The Dice scores are averaged over all 30 anatomical structures for each subject.

Figure 3-11 shows one outlier subject in the test set. This subject contains an artifact that was accidentally created during the simulation process of the sparse ADNI dataset. For better visualization, this subject was removed from the plots in Figure 3-10 and Figure 3-11. The Dice for this subject was 0.451 when registering with VoxelMorph and 0.680 when registering with SparseVM.

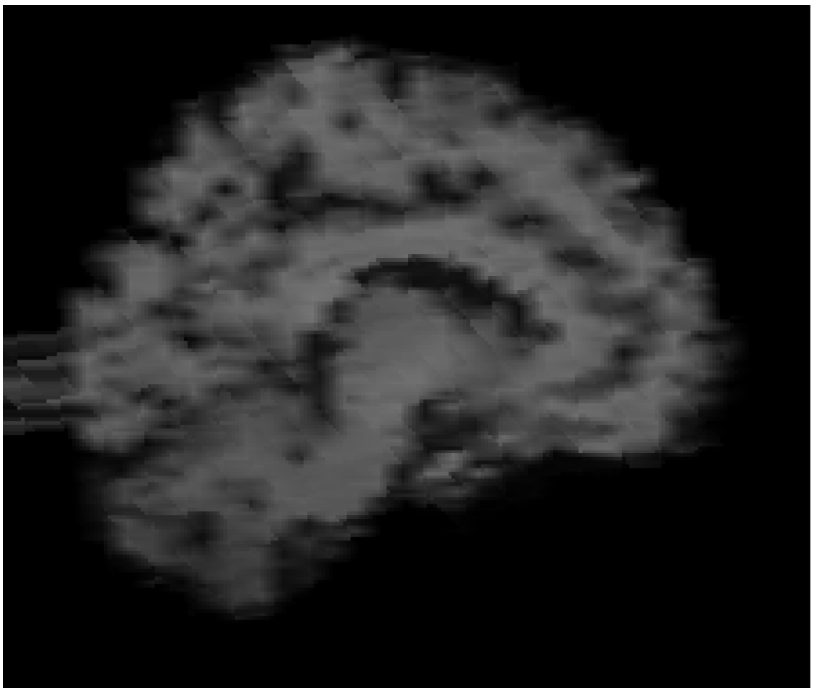


Figure 3-11: Outlier test subject. The artifact was accidentally created during the simulation process of the sparse ADNI dataset.

Chapter 4

Summary and Conclusion

Clinical image registration is important for many clinical analyses including population studies. In this work we focus on atlas-based registration of 3D MRI scans for a simulated sparse dataset and a clinically acquired ischemic stroke dataset. These scans contain many difficulties, such as sparsity and artifacts, that are not present in isotropic research-quality images. Current registration methods are either too slow to be feasibly used in clinical analyses or do not have sufficiently accurate performance on clinical data.

The work in this thesis has revealed several directions for future work. There are many pre-processing steps that we apply to the clinical data before registering. We believe our results would be further improved if the pre-processing steps were adapted for better performance on clinical images. Three future work directions in this vain would be to improve skull-stripping, affine alignment, and imputation.

4.1 Contributions

The main takeaway from this work is that clinical images can now be accurately registered in significantly less time than before. This means it is now feasible to register large clinical datasets. We introduce SparseVM, a fast unsupervised method that can accurately register sparse, low-resolution clinical scans. SparseVM significantly and meaningfully improves performance accuracy over VoxelMorph while maintaining

comparable speed. SparseVM is over 1000× faster than other methods that achieve state-of-the-art accuracy for clinical image registration, while achieving better accuracy. Since SparseVM can accurately register pairs of clinical scans in under a second on the GPU, it enables clinical image analyses that were not previously possible. For example, many medical population analyses involving tens of thousands of subjects were previously infeasible or would take months of computation. With our method, these analyses are now possible in minutes or hours. We are actively working on porting SparseVM to the clinical setting to be used in large clinical studies.

Bibliography

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

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

[3] J. Ashburner and K. Friston. Voxel-based morphometry-the methods. *Neuroimage*, 11:805–821, 2000.

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

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

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

[7] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46:1–21, 1989.

[8] Guha Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. An unsupervised learning model for deformable medical image registration. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[9] Guha Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *arXiv preprint arXiv:1809.05231*, 2018.

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

[11] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen. Deformable image registration based on similarity-steered cnn regression. pages 300–308. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017.

[12] F. Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[13] A. V. Dalca, Guha Balakrishnan, John Guttag, and M. R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. volume 11070, pages 729–738. MICCAI: Int. Conf. on Medical Image Computing and Computer Assisted Intervention., LNCS, 2018.

[14] A. V. Dalca, J. Guttag, and M. R. Sabuncu. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. pages 9290–9299. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[15] Adrian V. Dalca, Andreea Bobu, Natalia S. Rost, and Polina Golland. Patch-based discrete registration of clinical brain images. pages 60–67. Patch-Based Techniques in Medical Imaging: Second International Workshop, Patch-MI 2016, Held in Conjunction with MICCAI 2016, 2016.

[16] C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Computer Vision and Image Understanding*, 66(2):207–222, 1997.

[17] B. de Vos, F. F. Berendsen, Viergever M. A., M. Staring, and I. Ivsgum. End-to-end unsupervised deformable image registration with a convolutional neural network. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 204–212, 2017.

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

[19] Yi Gao and Allen Tannenbaum. Image processing and registration in a point set representation. In *Medical Imaging 2010: Image Processing*, volume 7623, page 762308. International Society for Optics and Photonics, 2010.

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

[21] Ben Glocker, Aristeidis Sotiras, Nikos Komodakis, and Nikos Paragios. Deformable medical image registration: setting the state of the art with discrete methods. *Annual review of biomedical engineering*, 13, 2011.

[22] Joseph V Hajnal and Derek LG Hill. *Medical image registration*. CRC press, 2001.

[23] Mattias P Heinrich, Mark Jenkinson, Michael Brady, and Julia A Schnabel. Mrf-based deformable registration and ventilation estimation of lung ct. *IEEE transactions on medical imaging*, 32(7):1239–1248, 2013.

[24] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.

[25] Hans J Johnson and Gary E Christensen. Consistent landmark and intensity-based image registration. *IEEE transactions on medical imaging*, 21(5):450–461, 2002.

[26] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[27] H. Li and Y. Fan. Non-rigid image registration using fully convolutional networks with deep self-supervision. *arXiv*, 2017.

[28] M. I. Miller, M. F. Beg, C. Ceritoglu, and C. Stark. Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *Proceedings of the National Academy of Sciences*, 102(27):9685–9690, 2005.

[29] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. R. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett. Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni). *Alzheimer's & Dementia*, 1(1):55–66, 2005.

[30] Yangming Ou, Aristeidis Sotiras, Nikos Paragios, and Christos Davatzikos. Dramms: Deformable registration via attribute matching and mutual-saliency weighting. *Medical image analysis*, 15(4):622–639, 2011.

[31] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[32] X. Pennec, P. Cachier, and N. Ayache. Understanding the "demon's algorithm": 3d non-rigid registration by gradient descent. pages 597–605, 1999.

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

[34] NS Rost, K. Fitzpatrick, A. Biffi, A. Kanakis, W. Devan, C. D. Anderson, L. Cortellini, K. L. Furie, and J. Rosand. White matter hyperintensity burden and susceptibility to cerebral ischemia. *Stroke*, 42(12):2807–2811, 2010.

[35] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformation: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[36] D. Shen and C. Davatzikos. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.

[37] H. Sokooti, B. de Vos, F. Berendsen, B. P. Levieveldt, I. Ivsgum, and M. Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. pages 232–239. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2017.

[38] R. Sridharan, A. V. Dalca, K. M. Fitzpatrick, L. Cloonan, A. Kanakis, O. Wu, K. L. Furie, J. Rosand, N. S. Rost, , and P. Golland. Quantification and analysis of large multimodal clinical image studies: Application to stroke. pages 18–30. International Workshop on Multimodal Brain Image Analysis, Springer, 2013.

[39] J. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[40] Matthew Toews, Lilla Zöllei, and William M Wells. Feature-based alignment of volumetric multi-modal images. In *International Conference on Information Processing in Medical Imaging*, pages 25–36. Springer, 2013.

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

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