Genetic, Clinical and Population Priors for Brain Images

by

Adrian Vasile Dalca

B.Sc., Department of Computer Science, University of Toronto, 2008
M.Sc., Department of Computer Science, University of Toronto, 2010
S.M., Electrical Engineering and Computer Science, M.I.T., 2012

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

Doctor of Philosophy
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

September 2016

Signature of Author: _____

Department of Electrical Engineering and Computer Science
August 31, 2016

Certified by: _____

Polina Golland
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: _____

Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Committee for Graduate Students

Genetic, Clinical and Population Priors for Brain Images

Adrian Vasile Dalca

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

Abstract

We develop mathematical models that exploit external information to improve analysis of a medical scan. Medical images enable visualization of the human body, and are central in clinical practice and many large-scale scientific studies. Medical image analysis uses computational models to interpret these scans towards the clinical or research goals. For example, in this thesis we are motivated by a clinical study of ischemic stroke, which aims to quantify cerebrovascular disease burden as observed in medical scans, along with its population trends and genetic predisposition.

In most analyses, anatomical information is extracted from images to provide insight into a problem, facilitating understanding of genetic variants, clinical variables and population trends. In contrast, this thesis investigates what these external factors tell us about the human anatomy and the medical scans themselves. First, we show how genetic and clinical indicators can be used to predict MRI scans of anatomical change through a semi-parametric generative model. Second, we demonstrate that a cohort of subjects with cerebrovascular disease can help identify the spatially complex pathology in a new subject through a generative computational model. Third, we use large collections of clinical images to dramatically improve the resolution of a new scan and recover fine-scale anatomy. We also present an approach for rapid interactive visualization of images in large studies. Bringing our methods together in large scale analyses of stroke and dementia subjects, we demonstrate new avenues of research enabled by these contributions.

Thesis Supervisor: Polina Golland
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

THIS thesis summarizes the scientific contributions of my graduate journey at MIT. However, alongside the research, this thriving environment has enabled more personal and professional growth than I could have ever imagined. With these next few paragraphs, I want to express my gratitude to the many people who have made this an exceptional experience that I will forever cherish. These few sentences do not do justice to the impact they've had on me, but I hope to pay forward the technical depth and warm kindness I've been taught.

I am deeply grateful to Polina Golland, an incredible mentor in all aspects of my graduate development. She is genuinely devoted to her students, making sure we learn to be good researchers while enjoying our PhD experience. She has encouraged me to follow ideas that motivate me, even those outside the scope of medical image analysis. Her constant endeavor to tease out meaningful insight has led me to strive for a deeper understanding of problems and methods. I have learned to look for meaningful directions and understand project realities thanks to Polina's regular reminders of the bigger picture and the clinical usefulness in our projects. Her optimism has helped me handle difficult periods and remain motivated even when there were plenty of reasons to feel discouraged. I've always walked away from meetings with Polina hopeful and motivated for the next steps. Polina's commitment to polished papers and presentations has dramatically improved my writing, editing and presentation skills. She is closely involved in our projects and contributes her insight while encouraging us to own and drive the projects ourselves. She manages an incredibly connected, motivated group and leads by example of what it means to be a great scholar and mentor. The lessons I've learned from Polina will forever shape who I am. As I continue on this academic path, I can't imagine a higher bar to aim for in being an academic supervisor.

I am fortunate to have had caring committee members who have offered genuine support and guidance throughout my PhD. I want to thank Mert Sabuncu, Natalia Rost and John Guttag for their commitment to my scholarly and personal development - I truly felt that they were invested in my development and well-being, which is rare and valuable to have. I inherited Mert's desk when I joined the lab, and I already knew this meant that I had big shoes to fill. Mert can provide the clearest insight and explanations of any problem or method. A short discussion with him always elucidates appropriate research decisions and next steps. His friendly, casual demeanor makes him a pleasure to work with, and I look forward to a thriving collaboration going forward. Natalia is our clinical guidance throughout the stroke projects. She has helped make sure our work focuses on realistic clinical needs, and through her efforts I am confident that the work in this thesis will enable downstream clinical discoveries. She has always constructively shared different viewpoints, broadening my understanding of clinical collaborations and scientific insight. I am thankful for her steady encouragement and optimism, both on a scholarly and personal level. Natalia's research group, and especially Lisa Cloonan, Cathy Zhang and Markus Schirmer, have been instrumental in motivating and enabling the work in this thesis. I want to thank John Guttag for his thorough attention to my work and thesis, and his ongoing support. He always takes the time to listen and give scholarly and professional guidance and his advice has really helped me think more clearly about my future.

Most of the work presented in this thesis was done in close collaboration with two wonderful colleagues, officemates and friends, Ramesh Sridharan and Katie Bouman. After long productive days full of whiteboard brainstorming sessions and mathematical debates, we still want to grab dinner and share stories. I believe these are truly rare connections. Ramesh has been closely involved with many of my projects. His incredible technical knowledge enabled the projects to move forward at a significant pace (otherwise I would even now be stuck on that first Variational EM derivation!), while his cheerful, carefree and encouraging demeanor made the research a delight to work on. With Katie, we've shared TAships, worked on projects from synthesizing video to improving medical data, and shared coffee or dinner at just about every place in Kendall Square. I could not ask for a more caring and attentive friend, and a more driven collaborator. Katie consistently surprises me with unique points of view that broaden my thinking and propel our research. All the way through the most tiring deadline crunches and post-deadline celebrations (irrespective of the outcome), Katie and Ramesh have made this an incredible journey.

I am thankful to the many people that have created the wonderful, lively, and productive environment I've had the pleasure to work in. My life is enriched by the many daily outings, passionate discussions, and group gatherings with members of the Medical Vision Lab: Ramesh Sridharan, George Chen, Danielle Pace, Christian Wachinger, Andreea Bobu, Archana Venkataraman, Ruizhi Liao, Polina Binder, Miaomiao Zhang, Kayhan Batmanghelich, Owen Hsu, Tammy Riklin-Raviv, Bjeorn Menze, Georg Langs, Danial Lashkari, Andrew Sweet, Michal Depa, Gabriel Robon, and officemates Miki Rubinstein, Giorgos Papachristoudis and Zoran Dzunic. I'm appreciative to all the many friends and colleagues from the Vision Graphics Neighbourhood, CSAIL, and beyond. I am grateful to Zoya Bylinskii for her friendship and unwavering encouragement. Zoya is a kind and cheerful close friend with wonderful research visions and passion for science, who's always reminded me of the importance of good science and dedication. I hope to always have a chance to look forward to evenings spent with Zoya and Alexei. I offer my gratitude to Fredo Durand for his teaching, his sharing of insight into photography, and his generosity. Finally, a big thank you to David Gifford, who has offered me amazing opportunities in videography and timelapses, working with video drones and exceptional dancers.

My parents Ana-Maria and Adrian Dalca Sr, and my sister, Ioana Dalca, have offered tireless support and encouragement. I know that their efforts and sacrifices lead to an opportunity for me to study and thrive at MIT. They raised me to be passionate and dedicated to fair and good work, and I will be eternally grateful for everything they've done for me. I am very happy to have enthusiastic conversations with my extended family, especially my inspirational grandparents. I am also amazed by the ongoing support offered by my wife's family Iolanda, Romeo and Eva Stanciu, and will always be thankful to them.

My wife, Monica Stanciu, is my home, my sanity, my partner in everything I do. She's carried me through my PhD, from the small things – the infinitely many decisions and hiccups – to the big challenges. She is my pillar of support in all aspects of life, my source of clinical clarification and scientific wisdom. I always appreciate the ability to have rich scientific conversations of all sorts. I am grateful for her love, understanding and unconditional support, and for helping me improve into the scientist I am today. Her kindness, cuteness, curiosity and excitement will always be my inspiration and drive in life.

Thank you all for helping to shape me at such an important time in my life.

Contents

Abstract 3

Acknowledgments 4

List of Figures 13

1 Introduction 21

1.1 Motivating Clinical Study . . . 22
1.2 Genetic Impact on Anatomical Imaging . . . 24
1.3 Spatial Pattern of Pathology . . . 27
1.4 Anatomical Structure in Clinical Images . . . 29
1.5 Contributions . . . 30

2 Prediction of Anatomical Change 33

2.1 Technical Background . . . 33
2.2 Prediction Model . . . 35
2.2.1 Subject-Specific Longitudinal Change . . . 35
2.2.2 Learning . . . 37
2.2.3 Prediction . . . 37
2.3 Model Instantiation for Anatomical Predictions . . . 38
2.3.1 Anatomical Phenotype . . . 38
2.3.2 Health Similarities . . . 39
2.4 Experiments . . . 40
2.4.1 Volumetric Predictions . . . 41
2.4.2 Anatomical Prediction . . . 42
2.5 Example Application . . . 43

2.6 Discussion and Extensions . . . 43
2.6.1 Summary . . . 45

3 Segmentation of Cerebrovascular Pathology 47
3.1 Background . . . 47
3.2 Generative Model . . . 49
3.2.1 Spatial Distribution Prior for Leukoaraiosis . . . 51
3.3 Inference . . . 52
3.4 Implementation . . . 53
3.5 Data . . . 54
3.6 Preprocessing . . . 55
3.7 Results . . . 58
3.8 Discussion . . . 59

4 Population Based Image Restoration 61
4.1 Technical Background . . . 62
4.2 Patch Mixture Model . . . 64
Model . . . 66
Restoration . . . 68
4.3 Patch Mixture Models with missing Data . . . 69
4.3.1 Latent Missing Data . . . 70
Learning . . . 70
Restoration . . . 72
4.3.2 Latent Low-Dimensional Patches . . . 72
Learning . . . 73
Restoration . . . 74
4.3.3 Implementation . . . 74
4.4 Results . . . 74
4.5 Discussion . . . 77

5 Visualization of Large-Scale Results 81
5.1 Background . . . 82
5.2 Design and Features . . . 83
Display Dimensions versus Navigation Dimensions . . . 84
Implementation and Features . . . 84
5.3 Evaluations . . . 86

Outlier Detection . . . 86
Pattern Identification . . . 87
5.4 Discussion . . . 89

6 Discussion and Conclusions **91**
6.1 Technical Directions . . . 91
6.1.1 Genetic and Clinical Factors . . . 91
6.1.2 Collection of Subjects with Spatial Disease Burden . . . 92
6.1.3 Collections of Clinical Scans . . . 92
6.1.4 Rapid Visualization of Image Collections . . . 92
6.2 Clinical Analyses . . . 93
6.2.1 Multi Cohort Comparison . . . 93
6.2.2 Volume Progression with Age . . . 93
6.2.3 Distribution Progression with Age . . . 95
6.2.4 Prediction of Patient-Specific Growth . . . 97
6.2.5 Genetic Discovery . . . 97
6.3 Conclusions . . . 99

A Prediction of Anatomical Change: Parameter Updates **101**

B Mixture of Principal Components with Missing Data **103**
B.1 Preamble . . . 103
B.2 Single Multivariate Normal . . . 104
B.2.1 Multivariate Normal with Missing Data . . . 104
B.2.2 Multivariate Normal with Transformed Patches with Missing Voxels 106
B.2.3 PCA with Missing Voxels . . . 110
B.2.4 ML-PCA for Missing Data . . . 110
B.2.5 ML-PCA with Transformed Patches with Missing Voxels . . . 113
B.2.6 PPCA with Missing Data . . . 114
B.3 Mixture of Multivariate Normals . . . 117
B.3.1 Gaussian Mixture with Missing Data . . . 117
B.3.2 Mixture Model of Principal Analysers for Transformed Patches with Missing Voxels . . . 120
B.3.3 Mixture Model of Probabilistic Principal Analysers for Transformed Patches with Missing Voxels . . . 122

Bibliography **125**

List of Figures

1.1 Example of imaging used in stroke patients. Diffusion weighted imaging (DWI, left) visualizes acute infarct (red outline), T2-FLAIR MRI (middle) shows the white matter hyperintensity, or *leukoaraiosis* (green), and computed tomography angiogram (CTA, right) captures the cerebrovascular tree and the blood flow stoppage. 23

1.2 Example clinical images. Top row (orange box): research high resolution scan. Next three rows: images from a patient in the stroke study: T1 (top), T2-FLAIR (middle), and DWI (bottom). Three orthogonal slices are shown for each modality; in plane slices are highlighted in blue. Note the large slice thickness in clinical scans. 25

1.3 Significant ventricle expansion (orange outlines) due to brain atrophy (shrinkage). Corresponding axial slices shown for each age. These brain scans cover four years for an ADNI subject with an APOE gene mutation and mild cognitive impairment (MCI). Large initiatives study the potential correlation between genetic variants, structural brain changes, and diagnosis. 26

1.4 Average spatial patterns (yellow) of while matter disease, or leukoaraiosis, in three population clusters, overlayed on top of a template brain. . 28

1.5 Brain extraction example. Top: a typical brain extraction result for a research quality high resolution image. Bottom: a typical brain extraction result for a clinical low resolution scan. 29

2.1 A graphical representation of our generative model. Circles indicate random variable and rounded squares represent parameters. The plates indicate replication. I is the predicted image, and y_c are the phenotypes that help us determine the image. Phenotypes are affected by h, which are drawn from Gaussian processes and capture among-subject similarity through genetic, clinical and imaging factors. 35

2.2 Overview of anatomical prediction. Arrows represent a displacement vector at each voxel from the common atlas space to the subject space. We predict the change in a displacement field from the baseline (orange) to the follow-up timepoint (purple) using external factors. A follow-up image can then be formed by propagating voxels from the baseline, to the atlas, and finally to the follow-up image locations. 38

2.3 Relative error (lower is better) of volume prediction for seven structures for subjects in the top decile of volume change. For intracranial volume, whole brain volume, ventricles, hippocampus, entorhinal cortex, fusiform gyrus, and middle temporal gyrus, we report relative change between the baseline and the follow-up measurement (red), relative error in prediction using a population model (green), and the complete model (blue). . . . 40

2.4 Prediction results. Dice scores of labels propagated through three methods for several structures implicatd in AD in subjects with the most volume change for each structure. We report the prediction based on the registration of the actual follow-up scan to the atlas as an upper bound for warp-based prediction accuracy (black), predictions based on the population-wide linear regression model (green), and the full model (blue). 41

2.5 Example predicted volume, in axial, sagittal and coronal views. Qualitatively, the scan presents plausible anatomy, and would be challenging to discern from an actual scan. 42

2.6 Example comparison of predicted follow-up scans. This subjects presents with significant ventricle expansion due to brain atrophy from the baseline image to the follow-up observation. The predicted scan using the full model is able to capture significantly more of the ventricle expansion compared to the age-regression based follow-up. Qualitatively, both predicted follow-up scans present as plausible anatomical images. . . . 43

2.7 Comparison of an acquired and predicted anatomical image for a patient diagnosed with AD using a healthy model. Ventricles are outlines in red. The third image shows the predicted follow-up with a color overlay, indicating the squared magnitude of the difference in predicted versus observed deformation field, showing that the ventricles expand significantly less in the predicted scan. The figures indicate a significantly different expansion trajectory of the ventricles when compared to the observed follow-up. 44

3.1 Left: T2-FLAIR axial slice. Stroke (blue outline) can appear anywhere in the brain, can vary dramatically in shape, and is hyperintense. Leukoaraiosis (yellow outline) is generally peri-ventricular, has a more predictable spatial distribution than stroke lesions, and is usually roughly symmetric across hemispheres. Imaging artifacts (red) that can interfere with automatic segmentation of pathologies. Right: Examples of stroke lesions manually outlined in green. 48

3.2 Left: A graphical representation of our generative model. Circles indicate random variable and rounded squares represent parameters. Shaded circles represent observed quantities and the plates indicate replication. I is the acquired image. The image intensities are generated from a normal distribution parametrized by μ_c and σ_c for each tissue class l in the label map $\boldsymbol{L}$. Priors for the tissue classes are controlled by the distribution $\boldsymbol{M}$, which is driven by the weights $\{\alpha_k\}$ and principal axes $\{\boldsymbol{M}_k\}$ of the K-component Gaussian shape model, and spatial parameters β_x that define the prior probability of stroke in non-leukoaraiosis tissue. Right: Schematic of the generative process. The generation of the spatial extent of the leukoaraiosis map M is controlled by the weights $\{\alpha_k\}$, resulting in the classification of each pixel in leukoaraiosis (red), healthy (gray) or chronic lesion (blue). A T2-FLAIR image I is generated based on the intensity model of each label. 50

3.3 Distances from patient images to median image for T2-FLAIR and DWI modalities, ranked in descending order. Outliers are shown in red, and are removed from subsequent analysis. 54

3.4 The first four modes of leukoaraiosis covariation in the stroke dataset. The spatial maps are shown in color, overlayed on a template brain. The distributions have been slightly smoothed and intensity normalized for visualization purposes. Lateral symmetry, anterior-posterior symmetry, and periventricular co-variation stand out as dominant patterns learned by the model. 56

3.5 Illustrative automatic segmentation result for two slices (top and bottom). Left: The original (skull-stripped) flair image. Middle: hyperintensity segmentation (red) using a threshold-based method tuned to analyzing T2-FLAIR images in stroke patients [124]. Right: our automatic segmentation of leukoaraiosis (red) and chronic lesion (blue). Our method excludes chronic lesions from leukoaraiosis segmentation, as well as other spurious noisy segmentations seen in the results of thresholding. 57

3.6 Segmentation in difficult images. Automatic segmentation on two difficult subjects with significant chronic lesions showing lesion segmentation (blue outlines) and leukoaraiosis (yellow outlines). 57

3.7 Leukoaraiosis segmentation evaluation. Left: Volume measurements based on the automatic segmentation (orange) are within the range of experts (blue circles). Right: Volume estimates based on the automatic segmentation of leukoaraiosis against volume estimates based on the manual segmentations; the correlation coefficient is $r = 0.82$. 58

3.8 Leukoaraiosis segmentation comparison with ground truth for the 417 ISGC subjects. Volumes automatically segmented with our method (orange) have a significantly stronger correlation than those computed with a previous threshold-based segmentation method (blue) specifically designed for stroke patients [124]. 59

4.1 Problem setup with close up images including the medial longitudinal fissure. Top row: isotropic image that we seek to recover is sampled according to the sampling mask W leading to the observed image I. Due to affine registration, each mask is not binary, but includes continuous values between 0 and 1. Bottom row: nearest neighbour, linear interpolation, and our restoration. The most dramatic improvement can be seen in restorations of the skull, dura matter, and ventricles. 62

4.2 Top row: close up images of observed scan planes in the common atlas coordinate system. Bottom row: averages of an increasing number of sampling masks (images indicating where known voxels are present). With 21 images, most of the volume is covered. Due to affine registration, each mask is not binary, but includes continuous values between 0 and 1. 64

4.3 Schematic of our image restoration approach. To capture the structure of the training set, we learn a Gaussian Mixture of patches, which involves modelling missing data and low-dimensional embeddings of the patches. For clarity, the training images are shown as high-resolution in this schematic, but our application only involves sparse images, even for training. We illustrate the method with 2D patches and images, but the algorithms operate on 3D volumes and 3D patches. Given a new image, we infer which cluster each image patch belongs to, and restore the missing data using the learned model and the known voxels. We quilt the final volume from overlapping restored patches. 65

4.4 Covariance example. The red and orange voxels are never observed together in the same patch, but we should be able to learn their correlation from the sparse patch observations shown here, using the red-blue and orange-blue correlations. Modelling low dimensional patch representations $\boldsymbol{x}$ helps achieve this. 69

4.5 Restoration in variable images. Three affinely-aligned volumes illustrate the typical variability in the scans. While most anatomy is grossly aligned, the methods must be robust against significant local variability. The images also have some slightly different intensity profiles, which are hard to normalize due to the quality of the scans. Nonetheless, the proposed method restores the images accurately. 75

4.6 Representative restorations. A representative original high resolution image, a reconstruction by our method as well as three standard methods: nearest neighbor, NLM, and linear interpolation. Our method reconstructs more anatomically plausible substructures, as can be seen in both Axial and Saggital planes. 76

4.7 Errors for image restorations compared to ground truth isotropic scans. Due to high variability among subject images, we illustrate improvements of NLM upsampling, linear interpolation, and our method over nearest neighbor interpolation. Mean squared error (lower is better), and PSNR (higher is better) illustrate that our method recovers significantly better signal. 77

5.1 User interface that implements our approach. Elements in orange were added to this screenshot for illustration purposes. The position of the cursor, shown as x and y, controls which 2D image is displayed on the canvas. The load matrix on the right information bar (green) offers a visual indication of the position of the current image in the entire dataset. 83

5.2 Example uses of tipiX on a dataset of 3D volumes with many subjects. Left: one axial slice of a subject is shown to the user at a time. The vertical position of the cursor controls axial slice location in the same subject, and the horizontal position of the cursor controls the subject index in the study. Right: exploring 2D sagittal views in the same dataset. 85

5.3 Outlier detection. Left panel: a typical brain scan after registration to an atlas and skullstripping, a poorly registered subject, and a simulated hypointense lesion on a subject (yellow arrow). Right plots: boxplots of time taken to identify two outliers in mis-registration and lesion detection tasks. 86

5.4 Detection of inclusion threshold. Left panel: a typical well-aligned subject image shown next to a typical poorly registered image, with the template brain and ventricle boundaries shown in green. Right plot: the average threshold versus the time taken by users. 87

5.5 A screenshot of two images shown to users in the parameter sweep study. Axial slices from four separate subjects are shown with segmentation contours (orange) implied by the aligned atlas. Moving the cursor left to right sweeps the registration parameters; moving the cursor up and down goes through different axial slices. Left: results from the median parameter choice. Right: a parameter setting chosen by a user who noted that the top-left subject is otherwise mis-registered. 88

6.1 Average leukoaraiosis maps for four ISGC centers and ADNI. While the distribution patterns among the centers is consistent, the ADNI cohort, which does not include stroke cases, has a significantly weaker or more constrained spread. 94

6.2 Top: the two-component regression mixture model clusters the patients into those with high leukoaraiosis growth as a function of age (red) and those with low leukoaraiosis growth as a function of age (blue). The lines show a kernel regression of leukoaraiosis volume as a function of age in each cluster. The representative images shown are obtained via kernel regression of the leukoaraiosis label maps as a function of age. Bottom: the two sets of representative images in more detail. 96

6.3 Preliminary prediction of leukoaraiosis distribution. For three follow-up scans of one subject, the observed, predicted and error maps of leukoaraiosis distribution are projected onto the spatial statistical model of leukoaraiosis. The prediction error is multiplied by 10 for visualization. While the spatial change within one subject is minimal compared to population changes, the method can predict the slow growth over time. 98

Chapter 1

Introduction

MEDICAL images, such as Ultrasound (US), Magnetic Resonance (MR) and Computed Tomography (CT) facilitate visualization of the human body, and are a central component of clinical practice. For example, MRI and CT scans are part of the standard of care in diagnosing and evaluating stroke [20, 70, 114], a leading cause of death in the United States [86]. Large-scale scientific studies also employ medical imaging in examining populations for anatomical variability, functional response, correlation with clinical factors, and diagnostic prediction. For instance, the Alzheimer's Disease Neuroimaging Initiative (ADNI) uses MRI and PET imaging together with health measures to study the progression of Alzheimer's disease [63, 87].

Large collections of medical scans can provide powerful avenues of analysis. However, gathering good quality data and making it available for collaborative research is a costly and complex undertaking. In contrast, millions of diagnostic scans are acquired every day as part of standard medical practice, and are often tuned to a particular medical condition. These scans can be accompanied by other clinical factors such as basic patient information and blood tests, constituting a rich data source for analysis. Through collaborative efforts and changing public policy, these heterogeneous clinical scans are being made available, opening new avenues for research, while presenting novel challenges in dealing with widely heterogeneous, high dimensional data.

Medical image analysis uses computational models to help interpret medical images and aid in achieving clinical or research goals. Most often, meaninful information is extracted from images to provide insight to a problem. For example, spinal vertebra and nerve roots [22, 91, 146], tumors [15, 82], and strokes [49, 73] can be identified, delineated and analysed to aid in diagnoses and treatment planning. Similar techniques can be used in scientific studies to drive discovery, or elucidate population-wide patterns. In addition to medical scans, recent availability of external data, such as clinical variables and genotyping, are facilitating better patient-specific understanding and population

analyses [129, 142]

Medical image analysis has also started to play an important role in informing genetic studies. Population genetics investigates genetic variation with respect to phenotypes or diagnoses of interest. For example, the ADNI study aims to discover and understand the role of genetic covariates in Alzheimer's disease. Medical images offer a rich phenotype where derived features represent meaningful traits to study by themselves, or can help untangle complex correlations between genetics and disease.

The anatomical information gained from analysis of medical images is therefore a gateway to improvements in clinical practice and scientific study. They have been used to help diagnose, treat, explain disease and understand medical variables. But what can these non-imaging factors, in turn, tell us about the human anatomy and medical images themselves?

Preview of Contributions

In this thesis, we study how factors external to a new medical scan, such as genotypes, medical charts and images of other subjects, can help characterize anatomy in the new patient. First, we show how genetic and clinical indicators can be used to predict MRI scans of anatomical change. Second, we demonstrate that a cohort of subjects with white matter pathology can help predict the spatially complex disease in a new subject. Third, we use large collections of low quality clinical scans to dramatically improve the resolution of a new subject and recover fine-scale anatomy. We also present an approach for rapid interactive visualization of imaging data in large studies. Bringing these methods together in large scale analyses and evaluation, we outline new avenues of research enabled by our contributions.

■ 1.1 Motivating Clinical Study

Throughout this thesis, we will work with two motivating clinical problems and associated image collections: ischemic stroke and Alzheimer's disease.

Ischemic Stroke

Stroke is a leading cause of death and debilitating injury in the United States [86], leading to significant clinical and scientific study. Medically, stroke refers to brain cell death due to improper blood flow, and ischemic stroke specifically implies an occlusion of a blood vessel. Most stroke episodes require urgent treatment, and lasting symptoms

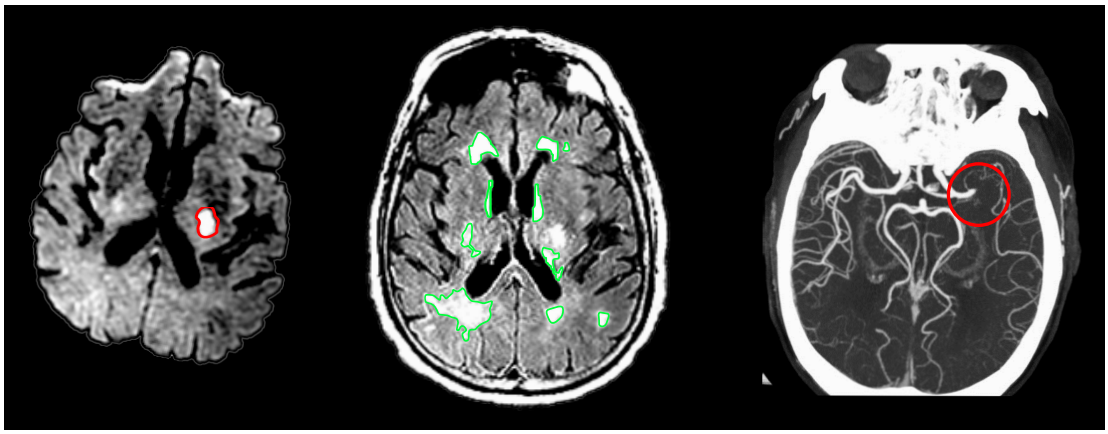

Figure 1.1. Example of imaging used in stroke patients. Diffusion weighted imaging (DWI, left) visualizes acute infarct (red outline), T2-FLAIR MRI (middle) shows the white matter hyperintensity, or *leukoaraiosis* (green), and computed tomography angiogram (CTA, right) captures the cerebrovascular tree and the blood flow stoppage.

vary widely and are often debilitating for survivors. Prevention, and many scientific studies, concentrate on the risk factors, such as smoking, hypertension, and heredity [59, 101]. Despite significant research commitment to the study of stroke, many of its causes and repercussions are largely not explained by known risk or genetic factors [56, 101].

We concentrate on a clinical stroke study within the International Stroke Genetics Consortium (ISGC) [83]. The ISGC is a large collaboration that aims to understand genetic influences in stroke, in close collaboration with other large consortia, such as CHARGE [98] and METASTROKE [132]. The effort includes a large number of data sources spanning genotypes, imaging, clinical characteristics, stroke clinical scores, patient chart data, and outcomes [101].

The particular study we tackle aims to quantify white matter disease burden as observed in medical scans, along with its population trends and genetic predisposition in the context of ischemic stroke patients [109]. Diagnostic imaging, specifically CT and MRI, are most often acquired within 48 hours of stroke onset for each subject as part of clinical practice, leading to a large collection of images of stroke patients. As part of routine care of ischemic stroke, computed tomography angiogram (CTA) is used to help identify the vascular blockage location, diffusion imaging reveals tissue affected by the stroke, and sometimes T2-FLAIR imaging is acquired to capture the small vessel disease through hyperintense voxels (Figure 1.1). Our motivating stroke study involves twelve acquisition sites including more than 3000 patients with MRI scans, genotype

data, and health factors such as hypertension and smoking.

The stroke study aims to extract information from these clinically acquired images to gain insight into stroke burden and genetic predispositions. The clinical images are of dramatically lower quality than scans available in research studies, and the imaging phenotypes of interest are complex. We show that using factors external to a subject's own scan can help us infer spatial pathology. The stroke dataset is therefore an ideal problem and challenge, while providing an avenue for our methods to contribute to clinical understanding of a devastating disease.

Alzheimer's Disease

Alzheimer's Disease (AD) is the most common subtype of dementia, affecting more than five million adults in the United States, and is characterized symptomatically by loss of memory and cognitive abilities [2, 12, 36]. Recent focus on Alzheimer's Disease research has provided advancement in early diagnosis, genetic predisposition, risk factors and relevant imaging biomarkers, but no treatment or cure has been widely accepted [2, 10, 33, 103, 106, 128]. These efforts have given rise to several large studies that include AD patients, healthy volunteers and subjects with Mild Cognitive Impairment (MCI) - an intermediate stage between normal cognitive decline and dementia.

We concentrate on the Alzheimer's Disease Neuroimaging Initiative (ADNI) [63, 87], which has enrolled hundreds of AD patients, MCI subjects and healthy volunteers between the ages of 55 and 95. The study includes basic and cognitive patient information, longitudinal MRI scans, genotypes, and circulating biomarkers. Specifically, the images include high resolution T1 MRI scans as well as lower resolution T2-FLAIR images, the same modality prevalent in the stroke study which enables visualization of white matter disease burden. With ADNI, we demonstrate that using factors external to a subject's scan can help us infer fine-scale anatomy in a low-resolution image, and predict entire follow-up scans. In Chapter 6 we examine how our methods can be applied to both the stroke and ADNI cohorts, leading to new avenues of clinical research.

■ 1.2 Genetic Impact on Anatomical Imaging

Human genetics studies heritability and the effect of genes on inherited traits. Studies have analysed genetic effects on a diverse list of traits, from height and skin color [11, 104, 117], to very specific clinical diagnoses [19, 90]. With both genotyping and medical imaging becoming more accessible, some studies aim to obtain both in the same

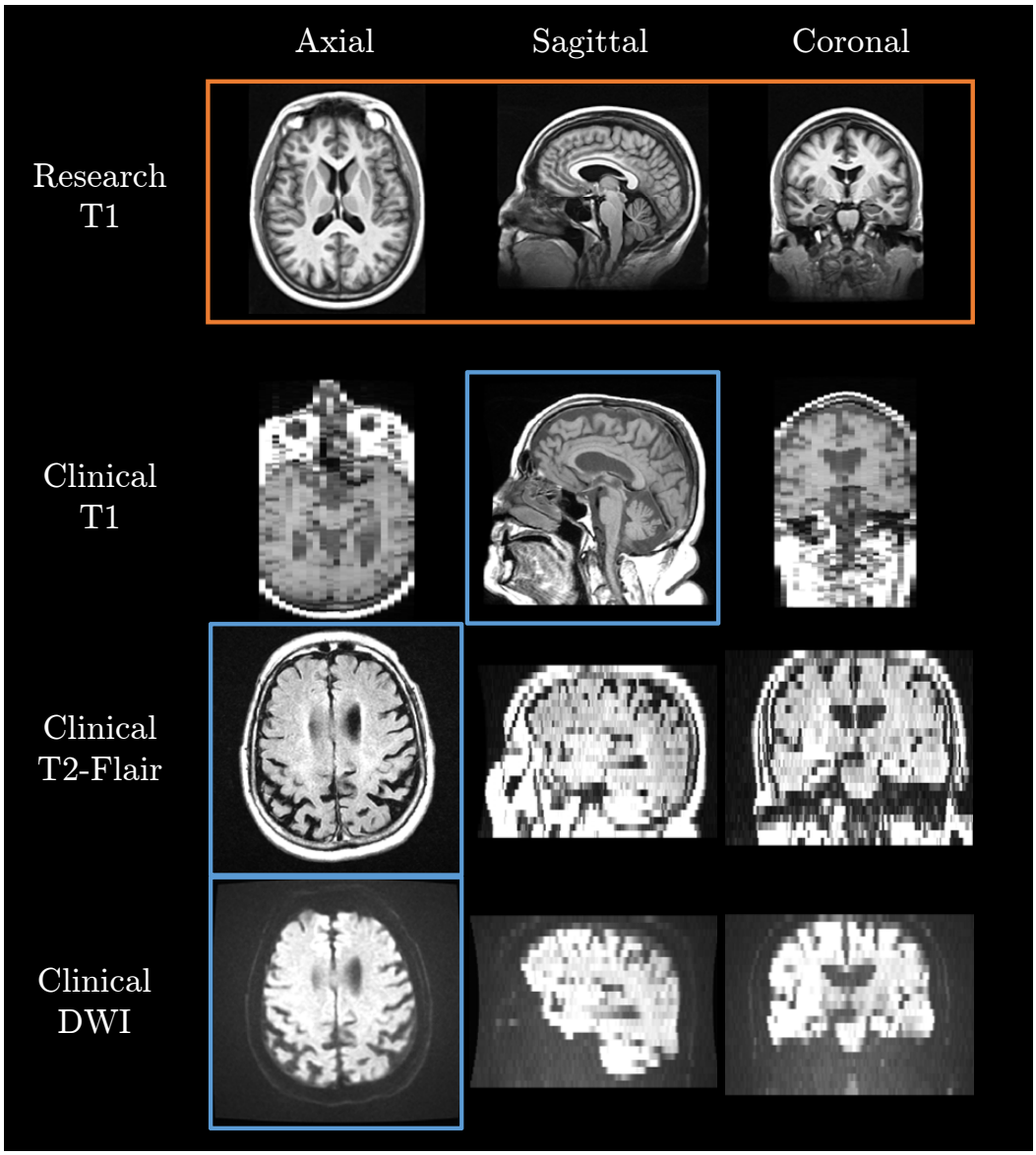

Figure 1.2. Example clinical images. Top row (orange box): research high resolution scan. Next three rows: images from a patient in the stroke study: T1 (top), T2-FLAIR (middle), and DWI (bottom). Three orthogonal slices are shown for each modality; in plane slices are highlighted in blue. Note the large slice thickness in clinical scans.

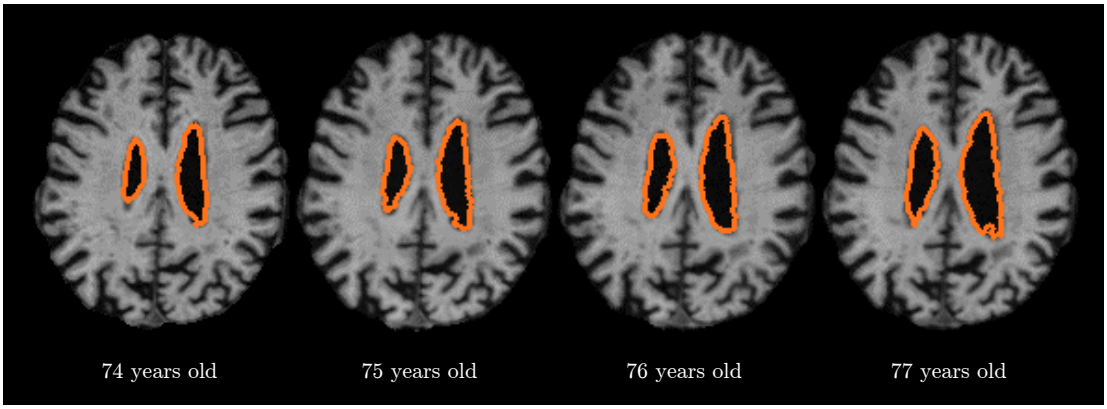

Figure 1.3. Significant ventricle expansion (orange outlines) due to brain atrophy (shrinkage). Corresponding axial slices shown for each age. These brain scans cover four years for an ADNI subject with an APOE gene mutation and mild cognitive impairment (MCI). Large initiatives study the potential correlation between genetic variants, structural brain changes, and diagnosis.

subjects wherever possible [63, 83, 87, 102]. Here, we introduce the assumptions of the developing field of imaging genetics, and offer a novel perspective on using genetics to help characterize human anatomy.

The search for genetic variants that increase the risk of a particular disorder is a central challenge in medical research, and has been traditionally performed via genome-wide association studies (GWAS). Such studies examine each genetic marker and its correlation with disease prevalence, usually independently of all other genetic variants in the study. While such analyses have been successful in discovering strong univariate genetic effects, complex genetic mechanisms leading to multifaceted diseases have proven more difficult to study. Similarly, genetic variants with important but weak effects have been challenging to identify. Traditionally, these problems are addressed by gathering larger cohorts of subjects to increase statistical power, leading to tens or hundreds of thousands of subjects in large collaborative studies [98]. However, such studies are rarely possible.

Imaging genetics studies the relationships between genetic variation and measurements from imaging data, often in the context of a disorder. While traditional genetic analyses are successful in deciphering simple genetic traits, imaging genetics can aid in understanding the underlying complex genetic mechanisms of multifaceted phenotypes. Specifically, imaging-based biomarkers are often used as an intermediate phenotype that provides a rich quantitative characterization of disease [5, 6, 113, 137]. For ex-

ample, hippocampal volume and brain atrophy have been linked to particular genetic variants which correlate strongly with Alzheimer's disease prognosis [8, 97, 106, 119, 122] (Figure 1.3). Since statistical population analysis is a central topic in medical image analysis, appearance regression models [29, 107], linear mixed effects models [27, 115] and longitudinal atlases [84, 93, 106] offer powerful analysis of biomarkers or images across populations. This classical correlation-based analysis has also yielded important characterization of relationships between imaging data and independent clinical variables [29, 84, 93, 106, 107]. Population genetics studies that typically require large patient cohorts are also making increased use of imaging data [84, 102].

Although there is a growing body of work aimed at using imaging phenotypes to drive genetic discovery, there has been limited modeling of the anatomical effects of genetic variants, and using these models to predict anatomy in new subjects. The established correlations between genetic variants and imaging-based biomarkers, such as hippocampal volume or brain atrophy, suggest a useful causality from genotype to anatomy. It is plausible to use genetic variation to predict anatomical change and drive discovery. We therefore reason that standard medical image analysis tasks stand to benefit from using genetic information. For example, in multi-atlas segmentation frameworks, one can choose atlases based on the anatomically relevant genetic similarity to a new patient, and genotype correlations with specific anatomical structures can be used to regularize image registration.

In this thesis we present a method for predicting anatomical change of a patient with the help of information external to their scan, including genetic and clinical indicators. Specifically, given only a first scan of a new subject in a longitudinal study, we predict anatomical changes and generate subsequent images by leveraging a longitudinal image collection, subject-specific genotypes, and medical factors. We discuss how predictive modeling facilitates novel analyses in both voxel-level studies and longitudinal biomarker evaluation, and demonstrate a novel application.

■ 1.3 Spatial Pattern of Pathology

The stroke study aims to characterize the burden of leukoaraiosis, or small-vessel disease, in stroke patients. Identifying and differentiating cerebrovascular pathologies as observed in brain MRI is critical for understanding cerebral ischemia. Leukoaraiosis describes non-specific changes in the white matter characterized by periventricular (near-ventricle) hyperintense voxels in T2-FLAIR MRI acquisitions (Figure 1.1). To

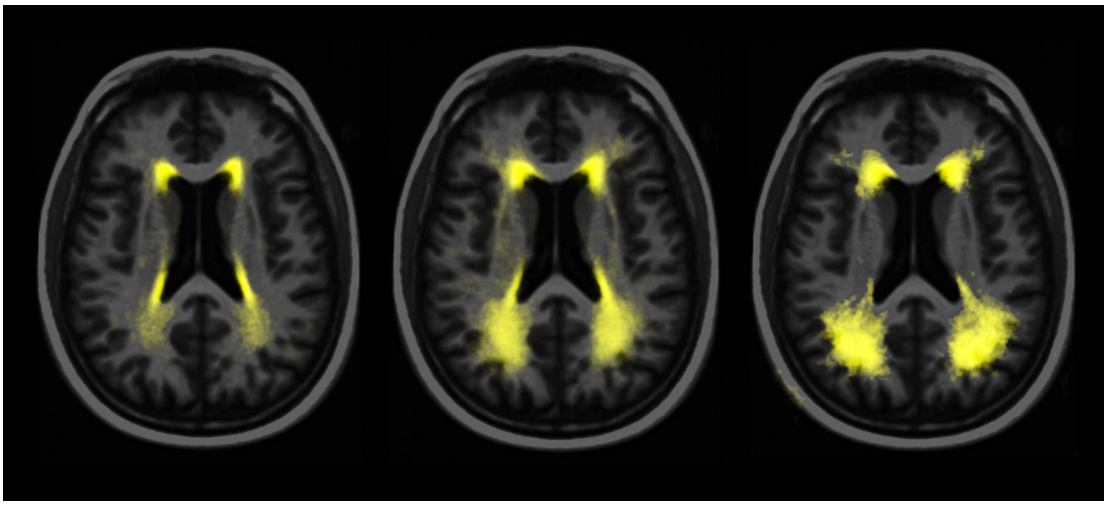

Figure 1.4. Average spatial patterns (yellow) of while matter disease, or leukoaraiosis, in three population clusters, overlayed on top of a template brain.

understand susceptibility to cerebral ischemia and associated risk factors, clinicians manually outline and analyze vascular pathologies, focusing on leukoaraiosis and separating it from stroke and other lesions. Using this approach, leukoaraiosis burden has been shown to be lower in patients with transient ischemic attacks compared to patients with more damaging cerebral infarcts [109]. Manual delineation of leukoaraiosis and stroke takes up to 30 minutes per patient. With thousands of patients already enrolled in the study and an increasing recruitment rate, automatic methods of analysis are directly needed.

While commonly found in stroke patients, leukoaraiosis can also be often seen in healthy subjects, especially in the elderly population. As such, T2-FLAIR images with leukoaraiosis are available in other studies, such as ADNI, offering the opportunity for cross-dataset and longitudinal analysis.

Unfortunately, different lesion types, such as leukoaraiosis and stroke, cannot be distinguished purely based on intensities in T2-FLAIR or location. Clinicians use anatomical and other medical knowledge to categorize and delineate pathology. In essence, they use their clinical insight gathered from numerous patients to infer the distribution of leukoaraiosis in a new patient. Drawing on the clinican's experience, we model intensity, shape, and spatial distribution of pathologies, learned from a population, to capture this anatomical knowledge of variability of pathology. By leveraging the entire population, we can successfully annotate clinical brain scans in stroke patients. Here,

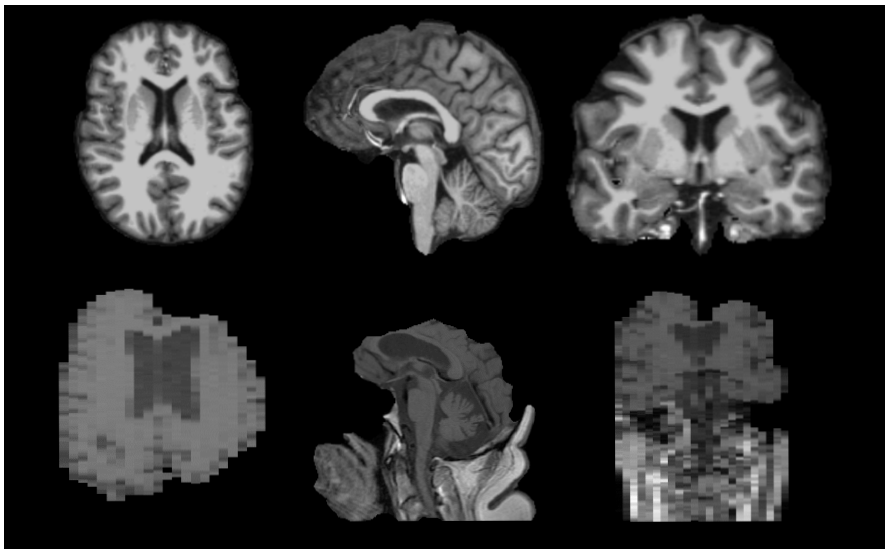

Figure 1.5. Brain extraction example. Top: a typical brain extraction result for a research quality high resolution image. Bottom: a typical brain extraction result for a clinical low resolution scan.

we focus on segmenting leukoaraiosis and separating it from stroke lesions and imaging artifacts. We discuss large-scale applications and describe novel clinical analyses facilitated by our method in a later Chapter.

■ 1.4 Anatomical Structure in Clinical Images

Enabled by increasingly affordable imaging technology and collaborative efforts like the ISGC [57, 102], the trend of large scale multi-site clinical studies involving imaging is likely to continue. In contrast to high quality research scans in studies that commonly motivate method development, such as ADNI [140] and Predict Huntington Disease (Predict-HD) [89], clinical studies present significantly lower quality images. For instance, in the stroke study, the brain scans are acquired within a few hours of stroke onset, limiting scanning time and resulting in MR scans with slice thickness of 6mm, compared to the usual $1-2$mm in research scans (Figure 1.2). Recently, the same scan modality is acquired on ADNI patients, enabling the study of white matter disease burden. Despite this limitation, the wealth of information in clinical data that is acquired through routine clinical care should be embraced and used to understand the disease.

Thick slices, patient motion, intensity artifacts, and cropped fields of view cause standard image processing algorithms and downstream analyses to fail on clinical im-

ages. Scans of different modalities are not only low resolution, but are also anisotropic in different directions, as illustrated in Figure 1.2. Even distinguishing between white and gray matter is challenging in the resulting T1 images due to poor tissue contrast. Analyses of such medical image collections are often hindered by the image quality, presenting significant challenges for basic tasks such as image registration (alignment) or segmentation (delineation of structures of interest) [109, 123]. For example, a standard first step in brain image analysis is the isolation of the brain voxels from skull and other structures. In research-quality images, skull stripping or brain mask extraction algorithms often assume that the brain consists of a single connected component separated from the skull and dura by cerebrospinal fluid [121]. Unfortunately, the validity of this assumption depends on image quality, and requires high resolution and reliable contrast. As a result, methods developed for analysis of research scans often fail when applied to clinical images, as illustrated in Figure 1.5.

We demonstrate image analysis techniques that use information across an entire collection of low quality images to infer missing anatomical structure in a low resolution image of a new subject. We take advantage of the fact that local fine-scale structure is intrinsically shared in a population, and each low-resolution scan captures some aspect of this structure. We present a novel method for constructing high resolution anatomically plausible volumetric images consistent with low resolution scans. Our method does not require high resolution scans or expert annotations, which would limit its usability, but can instead build the missing structure by learning from collections of clinical scans of similar quality to that of the input clinical image. The results are anatomically plausible restored scans, enabling subsequent analysis of the clinical scans with existing analysis algorithms. This approach is broadly applicable to many clinical problems, and we demonstrate it in the ADNI study where high-resolutoin data enables validation of the results. To the best of our knowledge, this is the first demonstration that a collection of low resolution images can predict fine structure in a new subject.

■ 1.5 Contributions

The main contribution in this thesis comprises mathematical models that exploit available genetic information, clinical factors and large image collections to improve analysis and interpretation of images of a new patient. This thesis builds on several publications that first introduced the methods [6, 23, 24, 25, 124, 125].

Prediction of anatomical appearance from genetic and clinical factors

We demonstrate that external factors, including the subject's genotype and health factors, have a significant impact on the subject's anatomy as seen on a medical scan. In Chapter 2, we develop a semi-parametric generative model for predicting anatomy of a patient in subsequent brain scans following a baseline image. We capture anatomical change through a combination of population-wide regression and a non-parametric model of the subject's external health based on individual genetic and clinical indicators.

Identifying spatial pathologies from distributions across patients

We demonstrate that a collection of patients can help identify complex spatial pathology in a new patient. In Chapter 3, we model intensity, shape, and spatial distribution of pathologies to capture anatomical knowledge of disease variability from a population of subjects, and successfully annotate clinical brain scans in new stroke patients. In Chapter 6 we discuss directions of clinical research enabled by this model, and show preliminary results comparing disease burden in a larger stroke cohort and ADNI.

Restoring anatomy in low-resolution images from clinical population structure

We demonstrate that a collection of low-resolution images can aid anatomical image restoration in a new subject. In Chapter 4, we develop a general method for constructing high resolution anatomically plausible volumetric images consistent with low resolution scans. Our method makes use of fine-scale structural similarities across a population to restore clinical scans to the full resolution typical of research images. We demonstrate our method on the ADNI cohort, where available high-resolution scans enable validation of our method. We facilitate analyses of clinical images by improving image resolution while preserving anatomical structure.

Rapid Visualization of Image Collections

In Chapter 5, we present an approach that enables rapid interactive visualization of high dimensional image data. The method facilitates rapid assessment of large scale computational analysis of the studies tackled in this thesis.

Large-scale processing

In Chapter 6 we use the image analysis methods introduced in this thesis to extract and analyse cerebrovascular pathologies in both the stroke cohort and ADNI, totalling over 4000 subjects. We present insights learned from large scale processing and discuss clinical directions facilitated by this thesis. We also discuss future and related technical directions, with a focus on how anatomical insight can be gained from data external to an individual scan.

Software

For the methods discussed in this thesis we include open-sourced, accessible code. Additionally, we release several useful large-scale MATLAB libraries, with particular focus to working with image patches.

Chapter 2

Prediction of Anatomical Change

In this chapter, we explore how genetic and clinical factors affect anatomical phenotypes relevant to a disease. Specifically, we use patient-specific medical factors, along with patterns learned from a collection of MR images, in order to predict follow-up scans for new subjects. Such voxel-wise prediction opens up several new areas of analysis, enabling novel investigations both at the voxel level and at the level of derivative biomarker measures.

Given only a single baseline scan of a new subject in a longitudinal study, our model predicts anatomical changes and generates a subsequent image by leveraging subject-specific genetic and clinical information. We capture anatomical change through a combination of population-wide regression and a non-parametric model of the subject's health based on individual indicators. In contrast to classical correlation and longitudinal analysis, we focus on predicting new observations from a single subject observation. We demonstrate prediction of follow-up anatomical scans in the ADNI cohort, and illustrate a novel analysis approach that compares a patient's scans to the predicted subject-specific healthy anatomical trajectory. At the end of the chapter, we discuss insights gained from genetic prediction of anatomy, discuss a wide variety of applications and extensions, and propose research directions where genetics can be used to guide medical image analyses.

■ 2.1 Technical Background

Our model describes the change from a single (or *baseline*) medical scan in terms of population trends and subject-specific external information. The first scan of a subject in a longitudinal study serves as the baseline for the purposes of prediction. We model how anatomical appearance changes with age on average in a population, as well as deviations from the population average using a person's *health profile*. We characterize

such profiles non-parametrically based on the genotype, clinical information, and the baseline image. Other external sources can easily be incorporated in the model depending on data availability. Subject-specific change is constructed from the similarity of health profiles in the cohort, using a Gaussian process parametrized by a population health covariance. Given the predicted change, we synthesize new images through an appearance model.

Statistical population analysis is one of the central topics in medical image computing. The classical correlation-based analysis has yielded important characterization of relationships within imaging data and with independent clinical variables [29, 84, 93, 107]. Regression models of object appearance have been previously used for atlas construction and population analysis [29, 107]. These methods characterize population trends with respect to external variables, such as age or gender, and construct clinically relevant population averages.

Longitudinal analyses also characterize subject-specific temporal effects, usually in terms of changes in the biomarkers of interest. Longitudinal cohorts and studies promise to provide crucial insights into aging and disease [84, 93]. However, prior work in this area typically involve statistical comparisons methods, and do not attempt to predict relevant features. Mixed effects models have been shown to improve estimation of subject-specific longitudinal trends by using inter-population similarity [37, 115]. While these approaches offer a powerful basis for analysis of biomarkers or images in a population, they require multiple observations for any subject, and do not aim to provide subject-specific predictions given a single observation. The parameters of the models are examined for potential scientific insight, but they are not tested for predictive power.

Recently, several papers have proposed prediction of medical images or specific anatomical structures by extrapolating at least two time points (e.g. a baseline and a follow-up image) to simulate a later follow-up measurement [9, 41]. These models, however, necessitate observations of at least two time points to make predictions which are difficult to acquire for every new patient. They are therefore not applicable to most patients, not applicable during a first visit, and are sensitive to noise in these few measurements. Recent methods have also used stratified population average deformations to make follow-up predictions from a baseline image [85]. While classified by population subgroups, these predictions are not adapted to a particular subject's health or environment.

In contrast, we define the problem of population analysis as predicting anatomical

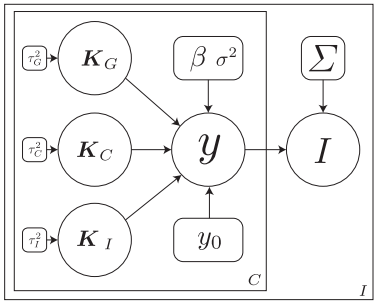

Figure 2.1. A graphical representation of our generative model. Circles indicate random variable and rounded squares represent parameters. The plates indicate replication. I is the predicted image, and y_c are the phenotypes that help us determine the image. Phenotypes are affected by h, which are drawn from Gaussian processes and capture among-subject similarity through genetic, clinical and imaging factors.

changes for individual subjects. Our generative model incorporates a population trend and uses subject-specific genetic and clinical information, along with the baseline image, to generate subsequent anatomical images. This prediction-oriented approach provides avenues for novel analysis, as illustrated by our experimental results. We are able to predict full subsequent anatomical images, facilitating novel clinical and population analyses.

■ 2.2 Prediction Model

Given a dataset of patients with longitudinal data, and a single baseline image for a new patient, we predict follow-up anatomical states for the patient. We model anatomy as a phenotype y that captures the underlying structure of interest. For example, y can be a low-dimensional descriptor of the anatomy at each voxel. We assume we only have a measurement of our phenotype at baseline y_b for a new subject. Our goal is to predict the phenotype y_t at a later time t. We let x_t be the subject age at time t, and define $\Delta x_t = x_t - x_b$ and $\Delta y_t = y_t - y_b$. We model the change in phenotype y_t using linear regression:

$$\Delta y_t = \Delta x_t \beta + \epsilon, \tag{2.1}$$

where β is the subject-specific regression coefficient, and noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is sampled from zero-mean Gaussian distribution with variance σ^2.

■ 2.2.1 Subject-Specific Longitudinal Change

To model subject-specific effects, we define $\beta = \bar{\beta} + H(\boldsymbol{g}, \boldsymbol{c}, f_b)$, where $\bar{\beta}$ is a global regression coefficient shared by the entire population, and H captures a deviation from this coefficient based on the subject's genetic variants g, clinical information c, and

baseline image features f_b.

We assume that patients' genetic variants and clinical indicators affect their anatomical appearance, and that subjects with similar health profiles exhibit similar patterns of anatomical change. We let $h_G(\cdot)$, $h_C(\cdot)$, $h_I(\cdot)$ be functions that capture genetic, clinical and imaging effects on the regression coefficients:

$$H(\boldsymbol{g}, \boldsymbol{c}, \boldsymbol{I}_b) = h_G(\boldsymbol{g}) + h_C(\boldsymbol{c}) + h_I(f_b). \tag{2.2}$$

Combining with (2.1), we arrive at the full model

$$\Delta y_t = \Delta x_t \bar{\beta} + \Delta x_t \big(h_G(\boldsymbol{g}) + h_C(\boldsymbol{c}) + h_I(f_b)\big) + \epsilon, \tag{2.3}$$

which captures the population trend $\bar{\beta}$, as well as the subject-specific deviations $[h_G(\cdot), h_C(\cdot), h_I(\cdot)]$. Together with a description of an image from phenotype y in the next section, equations (2.5) and (2.3) define a generative probabilistic interpretation of our model, which we show graphically in Figure 2.1.

For a longitudinal cohort of N subjects, we group all T_i observations for subject i to form $\Delta \boldsymbol{y}_i = [y_{i_1}, y_{i_2}, ...y_{i_{T_i}}]$. We then form the global vector $\Delta \boldsymbol{y} = [\Delta \boldsymbol{y}_1, \Delta \boldsymbol{y}_2, ..., \Delta \boldsymbol{y}_N]$. We similarly form vectors $\Delta \boldsymbol{x}$, $\boldsymbol{h}_G$, $\boldsymbol{h}_C$, $\boldsymbol{h}_I$, $\boldsymbol{g}$, $\boldsymbol{c}$, $\boldsymbol{f_b}$ and $\boldsymbol{\epsilon}$, to build the full regression model:

$$\Delta \boldsymbol{y} = \Delta \boldsymbol{x} \bar{\beta} + \Delta \boldsymbol{x} \odot (\boldsymbol{h}_G(\boldsymbol{g}) + \boldsymbol{h}_C(\boldsymbol{c}) + \boldsymbol{h}_I(\boldsymbol{f}_b)) + \boldsymbol{\epsilon}, \tag{2.4}$$

where $\odot$ is the Hadamard, or element-wise product. This formulation is mathematically equivalent to a General Linear Model (GLM) [80] in terms of the health profile predictors $[\boldsymbol{h}_G, \boldsymbol{h}_C, \boldsymbol{h}_I]$.

We employ Gaussian process priors to model the health functions:

$$h_D(\cdot) \sim GP(\boldsymbol{0}, \tau_D^2 K_D(\cdot, \cdot)), \tag{2.5}$$

where the covariance kernel function $\tau_D^2 K_D(z_i, z_j)$ captures the similarity between subjects i and j using feature vectors $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ for $D \in \{G, C, I\}$. We discuss the particular form of $K(\cdot, \cdot)$ used in Section 2.3.

■ 2.2.2 Learning

The Bayesian formulation in (2.4) and (2.5) can be interpreted as a linear mixed effects model (LMM) [81] or a least squares kernel machine (LSKM) regression model [44, 75]. We use the LMM interpretation to learn the parameters of our model, and the LSKM interpretation to perform final phenotype predictions.

Specifically, we treat $\bar{\beta}$ as the coefficient vector of fixed effects and $\boldsymbol{h}_G, \boldsymbol{h}_C$, and $\boldsymbol{h}_I$ as independent random effects. We seek the maximum likelihood estimates of parameters $\bar{\beta}$ and $\boldsymbol{\theta} = (\tau_G^2, \tau_C^2, \tau_I^2, \sigma^2)$ by adapting standard procedures for LMMs [44, 75]. As standard LMM solutions become computationally expensive for thousands of observations, we take advantage of the fact that while the entire genetic and the image phenotype data is large, the use of kernels on baseline data reduces the model size substantially. We obtain intuitive iterative updates that project the residuals at each step onto the expected rate of change in likelihood, and update $\bar{\beta}$ using the best linear unbiased predictor. Complete updates are listed in Appendix A.

■ 2.2.3 Prediction

Under the LSKM interpretation, the terms $h(\cdot)$ are estimated by minimizing a penalized squared-error loss function, which leads to the following solution [44, 68, 75, 138]:

$$h(z_i) = \sum_{j=1}^{N} \alpha_j K(\boldsymbol{z}_i, \boldsymbol{z}_j) \quad \text{or} \quad \boldsymbol{h} = \boldsymbol{\alpha}^T \boldsymbol{K} \tag{2.6}$$

for some vector $\boldsymbol{\alpha}$. Combining with the definitions of the LMM, we estimate coefficients vectors $\boldsymbol{\alpha}_G, \boldsymbol{\alpha}_C$ and $\boldsymbol{\alpha}_I$ from a linear system of equations that involves our estimates of $\hat{\beta}$ and θ (see Appendix A). We can then re-write (2.4) as

$$\Delta \boldsymbol{y} = \Delta \boldsymbol{x} \bar{\beta} + \Delta \boldsymbol{x} \left(\boldsymbol{\alpha}_G^T \boldsymbol{K}_G + \boldsymbol{\alpha}_C^T \boldsymbol{K}_C + \boldsymbol{\alpha}_I^T \boldsymbol{K}_I \right) \tag{2.7}$$

and predict a phenotype at time t for a new subject i:

$$y_t = y_b + \Delta x_t \left[\bar{\beta} + \sum_{j=1}^{N} \alpha_{G,j} K_G(\boldsymbol{g}_i, \boldsymbol{g}_j) + \alpha_{C,j} K_C(\boldsymbol{c}_i, \boldsymbol{c}_j) + \alpha_{I,j} K_I(f_i, f_j) \right]. \tag{2.8}$$

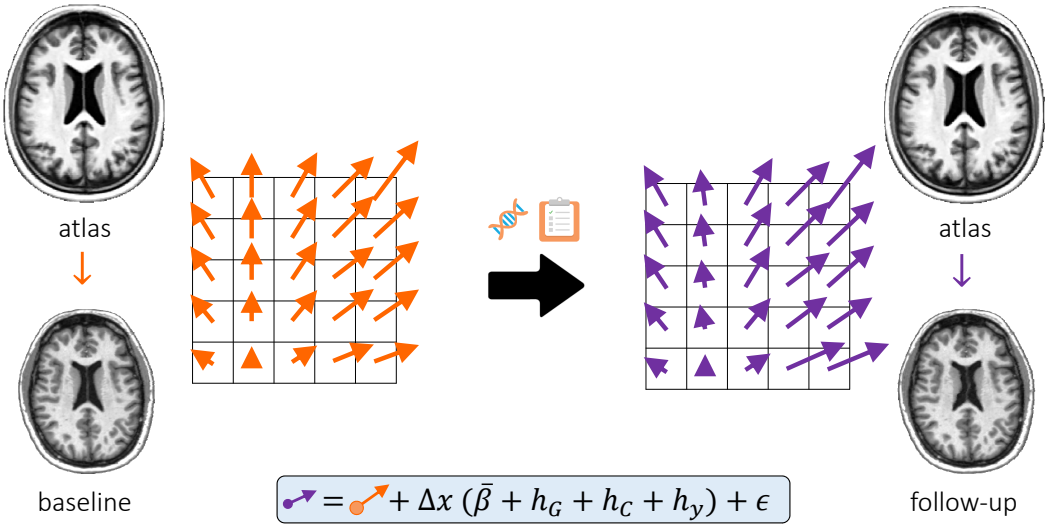

Figure 2.2. Overview of anatomical prediction. Arrows represent a displacement vector at each voxel from the common atlas space to the subject space. We predict the change in a displacement field from the baseline (orange) to the follow-up timepoint (purple) using external factors. A follow-up image can then be formed by propagating voxels from the baseline, to the atlas, and finally to the follow-up image locations.

■ 2.3 Model Instantiation for Anatomical Predictions

The full model (2.3) can be used with many reasonable phenotype definitions. Here, we describe the phenotype model we use for anatomical predictions and specify the similarity kernels of the health profile. In the Chapter 5, we explore prediction of other anatomical models for prediction of spatial disease signatures.

■ 2.3.1 Anatomical Phenotype

We define a voxel-wise phenotype that enables us to predict entire anatomical images. Let Ω be the set of all spatial locations v (voxels) in an image, and $I_b = \{I_b(v)\}_{v\in\Omega}$ be the acquired baseline image. We similarly define $A = \{A(v)\}_{v\in\Omega}$, to be the population atlas template. We assume each image I is generated through a deformation field Φ_{AI}^{-1} parametrized by the corresponding displacements $\{\boldsymbol{u}(v)\}_{v\in\Omega}$ from the common atlas to the subject-specific coordinate frame [107], such that $I(v) = A(v + \boldsymbol{u}(v))$. We further define a follow-up image I_t as a deformation Φ_{Bt} from the baseline image I_b, which can be composed to yield an overall deformation from the atlas to the follow-up scan

via $\Phi_{At}^{-1} = \Phi_{AB}^{-1} \circ \Phi_{Bt}^{-1} = \{\boldsymbol{u}'(v)\}_{v\in\Omega}$:

$$I_t(v) = A(v + \boldsymbol{u}'(v)). \tag{2.9}$$

Using displacements $\boldsymbol{u}'(v)$ as the phenotype of interest in (2.1) captures the necessary information for predicting new images, but leads to very high dimensional descriptors. To regularize the transformation and to improve efficiency, we define a low-dimensional embedding of $\boldsymbol{u}'(v)$. Specifically, we assume that the atlas provides a parcellation of the space into L anatomical labels $\mathcal{L} = \{\Psi\}_{l=1}^{L}$. We build a low-dimensional embedding of the transformation vectors $\boldsymbol{u}(v)$ within each label using principal component analysis. We define the relevant phenotypes $\{y_{l,c}\}$ as the coefficients associated with the first C principal components of the model that capture 95% of the variance in each label, for $l = 1 \dots L$.

We predict the phenotypes using (2.8). To construct a follow-up image I_t given phenotype y_t, we first form a deformation field $\widehat{\Phi}_{At}^{-1}$ by reconstruction from the estimated phenotype y_t, and use $\widehat{\Phi}_{At}$ assuming an invertible transformation. Using the baseline image, we predict a subsequent image via $\Phi_{Bt} = \widehat{\Phi}_{At} \circ \Phi_{AB}^{-1}$. Note that we do not directly model changes in image intensity. While population models necessitate capturing such changes, we predict changes from a baseline image. We also assume that affine transformations are not part of the deformations of interest, and thus all images are affinely registered to the atlas.

Using this appearance model, we use equation (2.8) to predict the anatomical appearance at every voxel, as illustrated in Figure 2.2.

■ 2.3.2 Health Similarities

To fully define the health similarity term $H(\cdot,\cdot,\cdot)$, we need to specify the forms of the kernel functions $K_G(\cdot,\cdot)$, $K_C(\cdot,\cdot)$, and $K_I(\cdot,\cdot)$.

For genetic data, we employ the identical by state (IBS) kernel often used in genetic analysis [100]. Given a vector of genetic variants $\boldsymbol{g}$ of length S, each genetic locus is encoded as $\boldsymbol{g}(s) \in \{0,1,2\}$, and

$$K_G(\boldsymbol{g}_i, \boldsymbol{g}_j) = \frac{1}{2S}\sum_{s=1}^{S}(2 - |\boldsymbol{g}_i(s) - \boldsymbol{g}_j(s)|). \tag{2.10}$$

To capture similarity of clinical indicators $\boldsymbol{c}$, we form the kernel function

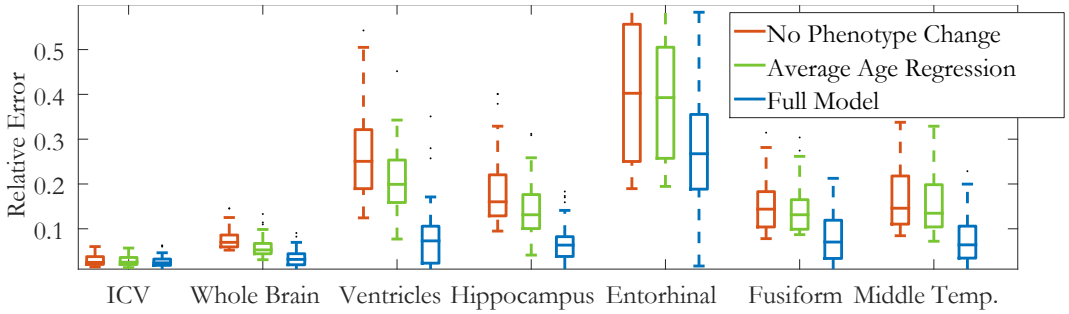

Figure 2.3. Relative error (lower is better) of volume prediction for seven structures for subjects in the top decile of volume change. For intracranial volume, whole brain volume, ventricles, hippocampus, entorhinal cortex, fusiform gyrus, and middle temporal gyrus, we report relative change between the baseline and the follow-up measurement (red), relative error in prediction using a population model (green), and the complete model (blue).

$$K_C(\boldsymbol{c}_i, \boldsymbol{c}_j) = \exp\left(-\frac{1}{\sigma_C^2}(\boldsymbol{c}_i - \boldsymbol{c}_j)^T \boldsymbol{W}(\boldsymbol{c}_i - \boldsymbol{c}_j)\right), \tag{2.11}$$

where diagonal weight matrix $\boldsymbol{W}$ captures the effect size of each clinical indicator on the phenotype, and σ_C^2 is the variance of the clinical factors.

We define the image feature vectors f_b as the set of all PCA coefficients defined above for the baseline image. We define the image kernel matrix as

$$K_I(f_{b,i}, f_{b,j}) = \exp\left(-\frac{1}{\sigma_I^2}||f_{b,i} - f_{b,j}||_2^2\right), \tag{2.12}$$

where σ_I^2 is the variance of the image features.

We emphasize that other data sources can easily be incorporated in the model by simply adding appropriate kernels for that data.

■ 2.4 Experiments

We illustrate our approach by predicting image-based phenotypes based on genetic, clinical and imaging data in the ADNI longitudinal study [63] that includes two to ten follow-up scans acquired $0.5-7$ years after the baseline scan. We use affine registration to align all subjects to a template constructed from 145 randomly chosen subjects, and compute non-linear registration warps Φ_{AI} for each image using ANTs [4]. We utilize a list of 21 genetic loci associated with Alzheimer's disease (AD) as the genetic vector

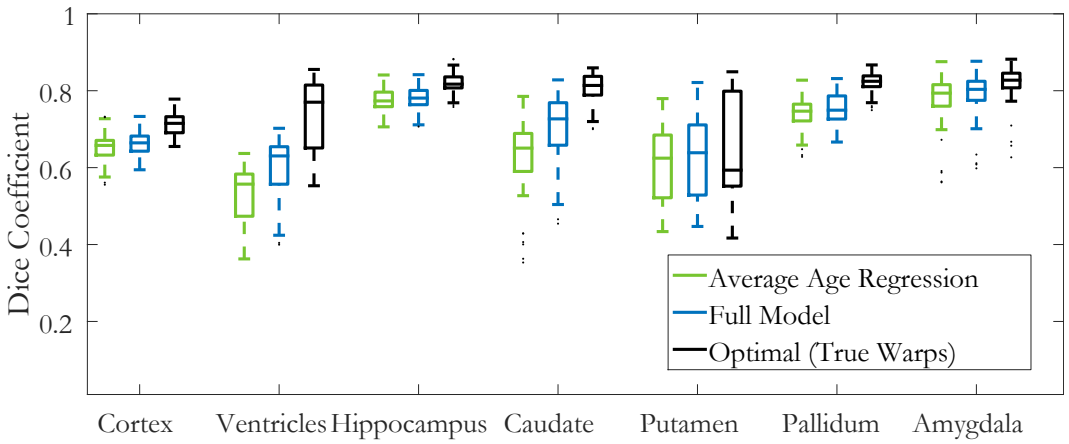

Figure 2.4. Prediction results. Dice scores of labels propagated through three methods for several structures implicatd in AD in subjects with the most volume change for each structure. We report the prediction based on the registration of the actual follow-up scan to the atlas as an upper bound for warp-based prediction accuracy (black), predictions based on the population-wide linear regression model (green), and the full model (blue).

g, and the standard clinical factors including age, gender, marital status, education, disease diagnostic at baseline, and cognitive tests at baseline, as the clinical indicator vector c. We learn the model parameters from 341 randomly chosen subjects and predict follow-up volumes on a separate set of 100 subjects. To evaluate the advantages of the proposed predictive model, we compare its performance to a population-wide linear regression model that ignores the subject-specific health profiles (i.e., $H = 0$).

■ 2.4.1 Volumetric Predictions

In the first simplified experiment, we define phenotype y to be a vector of several scalar volume measurements obtained using FreeSurfer [40]. These include intracranial volume, whole brain volume, ventricles, hippocampus, entorhinal cortex, fusiform gyrus, and middle temporal gyrus. In addition to the population-wide linear regression model, we include a simple approach of using the baseline volume measurements as a predictor of the phenotype trajectory, effectively assuming no volume change with time. Since in many subjects, the volume differences are small, all three methods perform comparably when evaluated on the whole test set. To evaluate the differences between the methods, we focus on the subset of subjects with substantial volume changes, reported in Fig. 2.3. Our method consistently achieves smaller relative errors than the two baseline

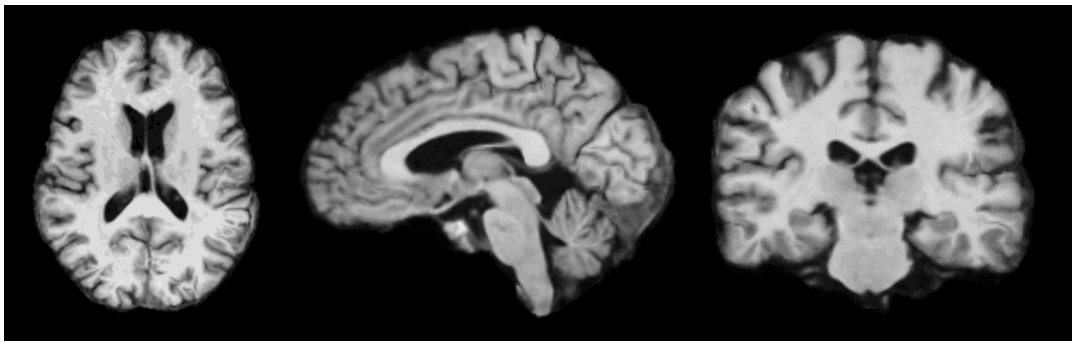

Figure 2.5. Example predicted volume, in axial, sagittal and coronal views. Qualitatively, the scan presents plausible anatomy, and would be challenging to discern from an actual scan.

approaches.

■ 2.4.2 Anatomical Prediction

We also evaluate the model for full anatomical scan prediction. To quantify prediction accuracy, we propagate segmentation labels of relevant anatomical structures from the baseline scan to the predicted scan using the predicted warps. We compare the predicted segmentation label maps with the actual segmentations of the follow-up scans. The warps computed based on the actual follow-up scans through the atlas provide an indication of the best accuracy the predictive model could achieve when using warps to represent images. Similar to the volumetric predictions, the full model offers modest improvements when evaluated on the entire test set, and substantial improvements in segmentation accuracy when evaluated in the subjects who exhibit large volume changes between the baseline scan and the follow-up scan, as reported in Fig. 2.4. In both experiments, all components h_g, h_c and h_I contributed significantly to the improved predictions.

Our experimental results suggest that the anatomical model depends on registration accuracy. In particular, we observe that directly registering the follow-up scan to the baseline scan leads to better alignment of segmentation labels than when transferring the labels through a composition of the transformations from the scans to the atlas space. This suggests that a different choice of appearance model may improve prediction accuracy, a promising direction for future work.

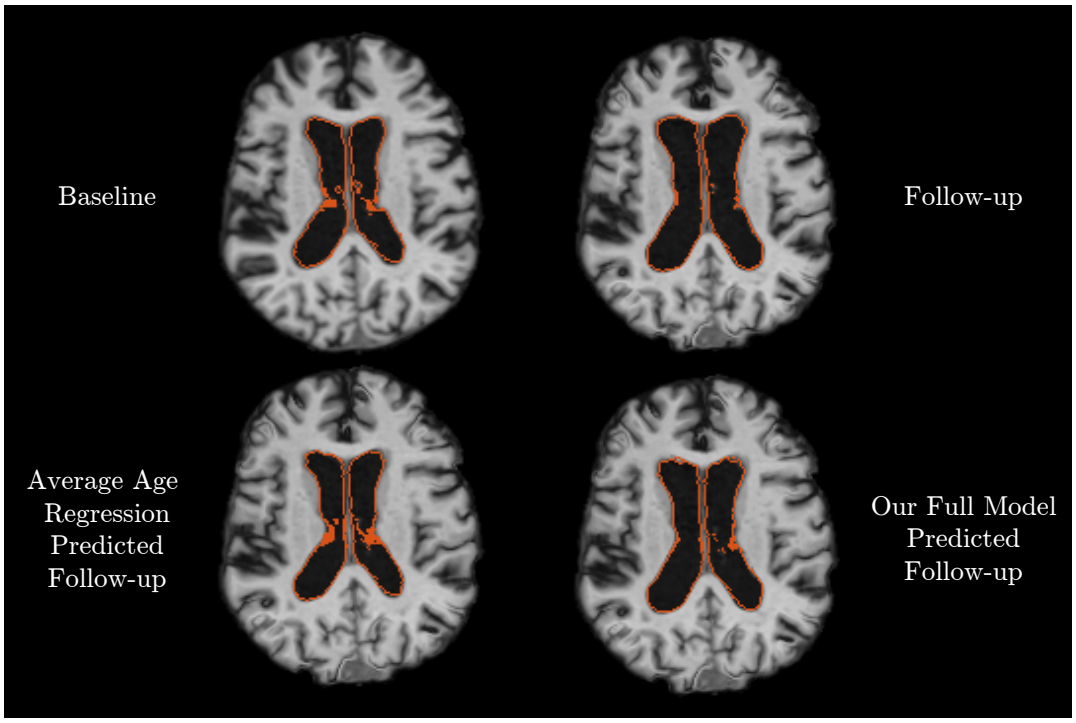

Figure 2.6. Example comparison of predicted follow-up scans. This subjects presents with significant ventricle expansion due to brain atrophy from the baseline image to the follow-up observation. The predicted scan using the full model is able to capture significantly more of the ventricle expansion compared to the age-regression based follow-up. Qualitatively, both predicted follow-up scans present as plausible anatomical images.

■ 2.5 Example Application

To demonstrate the potential of the anatomical prediction, we predict the follow-up scan of a patient diagnosed with dementia as if the patient were healthy. Specifically, we train our model using healthy subjects, and predict follow-up scans for AD patients. In Fig. 2.6 we illustrate an example result, comparing the areas of brain anatomy that differ from the observed follow-up in the predicted *healthy* brain of this AD patient. Our prediction indicates that ventricle expansion would be different if this patient had a healthy trajectory.

■ 2.6 Discussion and Extensions

Generally, the formulation presented in this chapter opens up several avenues of analysis, and can be applied in a variety of applications. For example, our model could

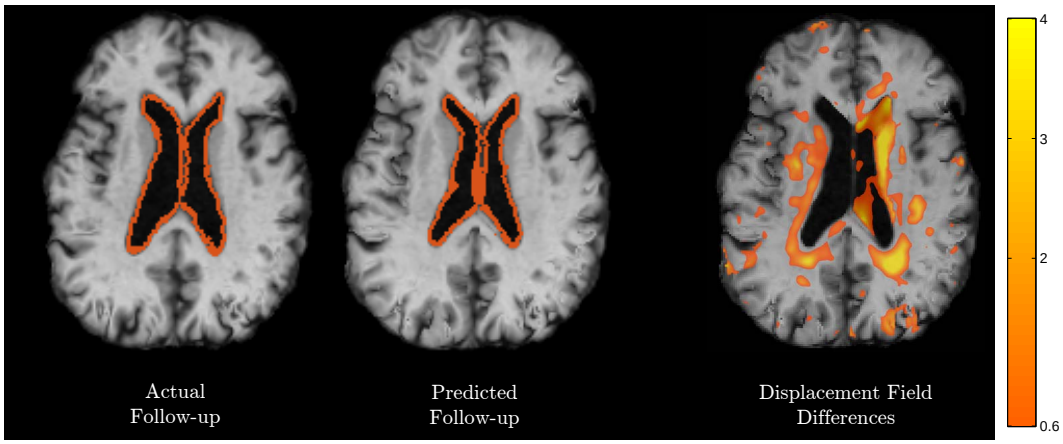

Figure 2.7. Comparison of an acquired and predicted anatomical image for a patient diagnosed with AD using a healthy model. Ventricles are outlines in red. The third image shows the predicted follow-up with a color overlay, indicating the squared magnitude of the difference in predicted versus observed deformation field, showing that the ventricles expand significantly less in the predicted scan. The figures indicate a significantly different expansion trajectory of the ventricles when compared to the observed follow-up.

predict changes in facial appearance of a person based on factors such as their genotype and environment, for instance learning from the FaceBase dataset which facilitates craniofacial research [53]. Changes can also be predicted on a different time scale, such as progression of a disease during a single hospital stay given patient-specific clinical measures, or long-scale musculoskeletal evolution of animals over centuries, given genus-specific and environmental information. When presented with other sources of data, researchers can design appropriate kernels and utilize them in the full model (2.7). Similarly, the model can accept any phenotype.

Several immediate model modifications will facilitate richer analyses. For example, phenotypes might respond non-linearly to the auxiliary data available, which can be handled by modifying the functional dependency in the main regression (2.7). While our model requires only a single baseline scan, other follow-up measurements, such as blood work or disease diagnostics, could help improve prediction of follow-up medical scans. Similarly, we aim to make predictions of anatomical or facial appearance from genotypes without even a baseline measurement $\boldsymbol{y}_B$. While this is a significantly harder challenge, the presented model provides a flexible framework from which to start.

■ 2.6.1 Summary

In this chapter, we demonstrated a model to predict the anatomy in patient follow-up images driven by genetic and clinical information. We validate our prediction method on scalar volumes and anatomical images, and show that it can be used as a powerful tool to illustrate how a subject-specific brain might differ if it were healthy. However, anatomical prediction is just one example of exploring insight from a subject's genotype in analysing medical scans. Given that genetic variants can help adapt subjects' temporal trajectories, it is likely that these factors can be used for other personalized application of medical imaging analyses. For example, segmentation of anatomical structures using the multi-atlas segmentation framework could use external factors for personalized atlas selection. Through this and other new applications, the insights present here illustrate a novel opportunity for the study of disease, anatomical development and image understanding.

Chapter 3

Segmentation of Cerebrovascular Pathology

IN this chapter, we demonstrate how a collection of images can help us infer complex pathology in the white matter of stroke patients. Specifically, we propose and demonstrate an inference algorithm for automatic segmentation of cerebrovascular pathologies in clinical MR images of the brain. Identifying and differentiating pathologies is important for understanding the underlying mechanisms and clinical outcomes of cerebral ischemia. Manual delineation of separate pathologies is infeasible in large studies of stroke that include thousands of patients. Unlike normal brain tissues and anatomical structures, the location and shape of the lesions vary across patients, presenting serious challenges for prior-driven segmentation. Our generative model captures spatial patterns and intensity properties associated with different cerebrovascular pathologies in stroke patients. We demonstrate the resulting segmentation algorithm on clinical images of the stroke patient cohort. We close the chapter with ongoing directions and insights about learning pathology from populations.

■ 3.1 Background

Variability in shape and location of white matter pathology is one of the main challenges in automatic segmentation of stroke scans. Leukoaraiosis appears hyperintense in T2-FLAIR, is found peri-ventricularly, has a widely variable extent, and is roughly bilaterally symmetric. While also hyperintense, strokes can happen nearly anywhere in the brain and vary dramatically in size and shape. While acute stroke (stroke that occurred in the last 48 hours) is visible on diffusion weighted MR (DWI), the same is not true for chronic stroke (stroke that occurred a long time before imaging). Additionally, DWI is often not available [143]. We concentrate on the more difficult task of

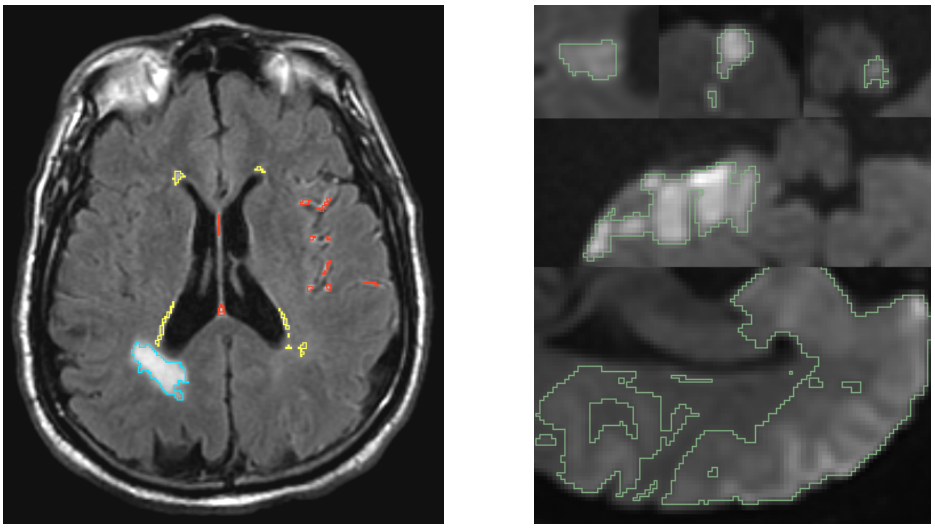

Figure 3.1. Left: T2-FLAIR axial slice. Stroke (blue outline) can appear anywhere in the brain, can vary dramatically in shape, and is hyperintense. Leukoaraiosis (yellow outline) is generally peri-ventricular, has a more predictable spatial distribution than stroke lesions, and is usually roughly symmetric across hemispheres. Imaging artifacts (red) that can interfere with automatic segmentation of pathologies. Right: Examples of stroke lesions manually outlined in green.

separating leukoaraiosis from stroke, both acute and chronic, in T2-FLAIR.

As discussed in the Introduction, another challenge is the low quality of images in the clinical setting due to the extremely limited scanning time. The resulting thick or sparse slices often have bright artifacts, that hinder registration and intensity equalization of clinical images and further complicate automatic segmentation. Representative images and segmentations are shown in Figure 3.1, illustrating our challenge.

To enable robust quantifications of pathology we introduce a generative probabilistic model of the effects of the cerebrovascular disease on the brain. The model integrates important aspects of each pathology, leading to an effective inference algorithm for segmentation and separation of different tissues in stroke patients. Specifically, we learn the spatial distribution and intensity profile of leukoaraiosis, as well as the intensity profile of stroke. We train the model on a dataset that was labeled by an expert rater and demonstrate that our modeling choices capture notions used by clinicians, such as symmetry and covariation of intensity patterns. To the best of our knowledge, this is the first comprehensive segmentation approach for different cerebrovascular pathologies.

Our model incorporates several approaches previously proposed for segmentation of

healthy anatomy that is consistent across individuals [16, 135, 141]. We combine these methods to accurately model pathology. Intensity-based lesion segmentation algorithms utilize tissue intensities to segment pathology [1, 66]. Spatial priors are sometimes added in a form of Markov Random Fields or spatial distributions [45, 116, 135]. The resulting methods are successful in delineating structures that are hyper- or hypointense compared to their surroundings, such as MS lesions or tumors. Unfortunately, these algorithms are not designed to differentiate between multiple hyperintense structures, such as leukoaraiosis, stroke, and certain artifacts, which share an intensity profile and can co-occur spatially. Clinicians use spatial features, such as the bilateral symmetry of leukoaraiosis, to tell them apart.

Shape-based methods generally model the shape of a structure, either via an explicit [16, 67, 127, 133] or implicit [48, 71, 96] representation. We utilize a shape model to capture the variability in spatial distribution of leukoaraiosis, which develops in a consistent pattern peri-ventricularly. In contrast, stroke can happen at random locations almost anywhere in the brain, and has no obvious shape or location profile (Figure 3.1).

We emulate clinician intuition in delineating leukoaraiosis by learning distribution and covariation patterns of the pathology from a collection of subjects. We demonstrate that combining intensity and spatial context for stroke and spatial distribution models for leukoaraiosis produces accurate segmentation. We validate the method on hundreds of stroke subjects with various pathologies and artifacts.

■ 3.2 Generative Model

We use a generative model to describe the spatial distribution, shape and appearance of healthy tissue and cerebrovascular pathology. Figure 3.2 provides a graphical representation of our model.

We let Ω be the set of all spatial locations (voxels) in an image, and $I = \{I_x\}_{x\in\Omega}$ be the acquired image. We assume image I is generated from a spatially varying label map $\boldsymbol{L} = \{\boldsymbol{L}_x\}_{x\in\Omega}$ that represents tissue classes. For each voxel x, $\boldsymbol{L}_x$ is a length-3 binary indicator vector that encodes three tissue labels – leukaraiosis (R), stroke (S) and healthy tissue (H). We use notation $L_x(l) = 1$ to mean that the tissue class at voxel x is l, for $l \in \{R, S, H\}$. Otherwise, $L_x(l) = 0$.

Given the label map $\boldsymbol{L}$, the intensity observations I_x are generated independently

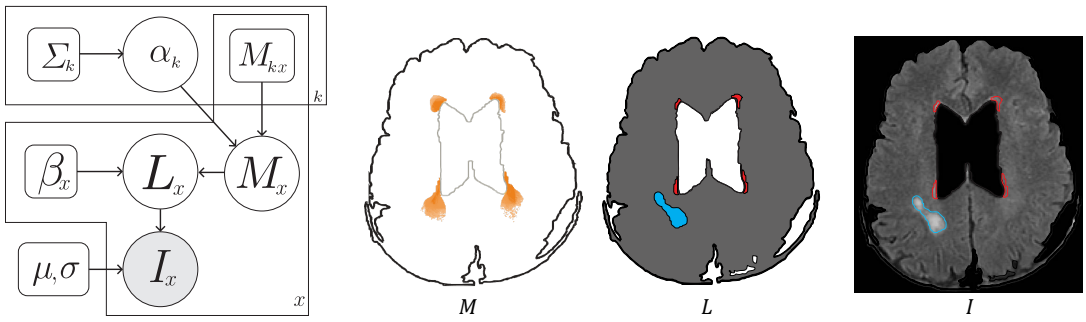

Figure 3.2. Left: A graphical representation of our generative model. Circles indicate random variable and rounded squares represent parameters. Shaded circles represent observed quantities and the plates indicate replication. I is the acquired image. The image intensities are generated from a normal distribution parametrized by μ_c and σ_c for each tissue class l in the label map $\boldsymbol{L}$. Priors for the tissue classes are controlled by the distribution $\boldsymbol{M}$, which is driven by the weights $\{\alpha_k\}$ and principal axes $\{\boldsymbol{M}_k\}$ of the K-component Gaussian shape model, and spatial parameters β_x that define the prior probability of stroke in non-leukoaraiosis tissue. Right: Schematic of the generative process. The generation of the spatial extent of the leukoaraiosis map M is controlled by the weights $\{\alpha_k\}$, resulting in the classification of each pixel in leukoaraiosis (red), healthy (gray) or chronic lesion (blue). A T2-FLAIR image I is generated based on the intensity model of each label.

from a Gaussian distribution:

$$P(\boldsymbol{I}|\boldsymbol{L};\boldsymbol{\mu},\boldsymbol{\sigma}) = \prod_{x\in\Omega}\prod_{l\in\mathcal{L}} \mathcal{N}(I_x;\mu_l,\sigma_l)^{L_x(l)}, \tag{3.1}$$

where $\mathcal{N}(\cdot;\mu,\sigma)$ is the normal distribution parametrized by mean μ and variance σ^2, $\mathcal{L} = \{R,S,H\}$, $\boldsymbol{\mu} = \{\mu_R,\mu_S,\mu_H\}$ and $\boldsymbol{\sigma} = \{\sigma_R,\sigma_H,\sigma_S\}$.

The prior for the tissue classes captures our knowledge about spatial distributions and shape of pathology. We assume that the spatial extent of leukoaraiosis depends on a spatial distribution $\boldsymbol{M} = \{M_x\}_{x\in\Omega}$, where M_x is a prior for leukoaraiosis for voxel x. As we describe later on, $\boldsymbol{M}$ will be parametrized by vector $\boldsymbol{\alpha}$. If voxel x is not assigned to leukoaraiosis, it is assigned to be stroke with spatially varying probability β_x, and to be healthy tissue with probability $(1-\beta_x)$. To encourage spatial contiguity, we incorporate a Markov Random Field (MRF) as a spatial prior. Formally,

$$P(\boldsymbol{L}|\boldsymbol{\alpha},\boldsymbol{\beta}) = \prod_x\prod_l \pi_x(l)^{L_x(l)} \prod_{y\in N(x)} \exp\left(\boldsymbol{L}_x^T A \boldsymbol{L}_y\right), \tag{3.2}$$

where

$$\boldsymbol{\pi}_x = [M_x(\boldsymbol{\alpha}),\ (1 - M_x(\boldsymbol{\alpha}))\beta_x,\ (1 - M_x(\boldsymbol{\alpha}))(1 - \beta_x)]^T \tag{3.3}$$

is a length-3 vector of prior probabilities for the three tissue classes as described above, $N(x)$ is the set of voxel locations neighboring x, and the 3×3 matrix A is chosen to encourage neighboring voxels to share the same tissue label. In our implementation, the MRF term penalizes interactions between stroke and other tissues more than leukoaraiosis bordering healthy tissue, as we find that stroke is generally more spatially contiguous than leukoaraiosis, which is more diffuse.

■ 3.2.1 Spatial Distribution Prior for Leukoaraiosis

We model the spatial extent of leukoariosis with a probabilistic atlas constructed by applying Principal Component Analysis (PCA) to a training set of manual leukoaraiosis binary segmentation maps. We let $\overline{\boldsymbol{M}}$ be the mean map, $\{M_k\}_{k=1}^K$ be the principal components that correspond to the K largest eigenvalues, and α_k be the weights (or loadings):

$$P(\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\alpha}; 0, \Sigma), \tag{3.4}$$

where Σ is the diagonal covariance matrix containing the K largest eigenvalues. Given α, the spatial prior $\boldsymbol{M} = \{M_x\}_{x \in \Omega}$ is deterministically defined:

$$\boldsymbol{M}(\boldsymbol{\alpha}) = \overline{\boldsymbol{M}} + \sum_k \alpha_k \boldsymbol{M}_k.$$

A simple average probability map representation fails to capture the covariation of leukoaraiosis distribution. As we increase the complexity of the model (k), it captures more detail but is more challenging to train without overfitting. We also experimented with LogOdds shape representation [96], often used for modeling normal anatomical variability. We found that the leukoaraiosis structures are in general too thin and variable in location to be properly captured by this representation.

3.3 Inference

Using (3.1), (3.2) and (3.3), we form the posterior distribution for the tissue classes:

$$\begin{aligned} P(\boldsymbol{L}|\boldsymbol{I};\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha},\boldsymbol{\beta}) &\propto P(\boldsymbol{I},\boldsymbol{L};\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha},\boldsymbol{\beta}) \\ &= P(\boldsymbol{I}|\boldsymbol{L};\boldsymbol{\mu},\boldsymbol{\sigma})P(\boldsymbol{L};\boldsymbol{\alpha},\boldsymbol{\beta}) \\ &= \prod_{x\in\Omega}\left(\prod_{l\in\mathcal{L}}[\pi_x(l)\mathcal{N}(I_x;\mu_l,\sigma_l)]^{L_x(l)}\prod_{y\in N(x)}\exp\left(\boldsymbol{L}_x^T A\boldsymbol{L}_y\right)\right) \end{aligned} \tag{3.5}$$

and we describe parameter estimation below. Given estimated parameters, we obtain the segmentations by performing MAP inference:

$$\hat{\boldsymbol{L}} = \arg\max_{\boldsymbol{L}} P(\boldsymbol{L}|\boldsymbol{I};\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha},\boldsymbol{\beta}). \tag{3.6}$$

Since exact computations become infeasible whenever the MRF weight matrix $\boldsymbol{A}$ is non-zero, we employ a variational EM approximation [62] to estimate the model parameters. Specifically, we approximate the posterior distribution $P(\boldsymbol{L}|\boldsymbol{I};\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha},\boldsymbol{\beta})$ with the fully factored distribution

$$q(\boldsymbol{L}) = \prod_{x\in\Omega} q(\boldsymbol{L}_x) = \prod_{x\in\Omega}\prod_{l\in\mathcal{L}} w_x(l)^{L_x(l)}, \tag{3.7}$$

where $\boldsymbol{w}_x$ is a vector of probabilities for the three tissue classes at voxel x. Because the prior for the loadings $P(\boldsymbol{\alpha})$ is not conjugate to the likelihood $P(\boldsymbol{L}|\boldsymbol{\alpha})$, we approximate the corresponding E-step computation with a regularized projection:

$$\begin{aligned} \boldsymbol{\alpha} &\leftarrow \arg\min_{\boldsymbol{a}} ||\boldsymbol{w}(R) - U\boldsymbol{a}||^2 + \lambda\boldsymbol{a}^T\Sigma^{-1}\boldsymbol{a} \\ &= \left(U^TU + \lambda\Sigma^{-1}\right)^{-1}U^T\boldsymbol{w}(R), \end{aligned} \tag{3.8}$$

where $U = [M_1, ..., M_k]$. We use clipping to force the resulting values in $\boldsymbol{M}(\boldsymbol{\alpha})$ to be between 0 and 1. Once the map of tissue priors $\boldsymbol{M}(\boldsymbol{\alpha})$ is computed, it determines the variational parameters $\boldsymbol{w}_x(l)$ for each $x \in \Omega$, $l \in \boldsymbol{L}$. The variational posterior parameters $\boldsymbol{w}_x$ are weighted by their agreement with the neighbors:

$$w_x(l) \leftarrow \pi_x(l)\mathcal{N}(I_x;\mu_l,\sigma_l)\prod_{y\in N(x)}\exp\left(\boldsymbol{w}_x^T A\boldsymbol{w}_y\right), \tag{3.9}$$

where $\pi_x(l)$ is defined in (3.3).

In the M-step, we update the parameters of the model. The updates are intuitive. The class mean and variance estimates are computed as weighted averages:

$$\mu_l \leftarrow \frac{\sum_x w_x(l) I_x}{\sum_x w_x(l)}, \tag{3.10}$$

$$\sigma_l^2 \leftarrow \frac{\sum_x w_x(l)(I_x - \mu_l)^2}{\sum_x w_x(l)}, \tag{3.11}$$

for $l \in \mathcal{L}$. Given large variability of intensity due to pathologies and severe artifacts, image inhomogeneity cannot be corrected through pre-processing steps. To address image inhomogeneity for the healthy tissue, we model the intensity mean estimate as spatially varying, and introduce a low pass filter G_H to enforce spatial smoothness, similar to the original EM-segmentation formulation [141]. Specifically the update becomes,

$$\mu_H \leftarrow G_H * (w_x(H) \cdot I)\,, \tag{3.12}$$

where $*$ denotes spatial convolution. The healthy tissue prior β_x is a fraction of current frequency estimates for stroke and healthy tissue probabilities:

$$\beta_x \leftarrow \frac{w_x(S)}{w_x(H) + w_x(S)}. \tag{3.13}$$

We iterate the updates until the parameter estimates converge.

■ 3.4 Implementation

The fixed parameters λ and A were chosen manually to optimize results in a single test example, not included in the results below. In particular, we use $\lambda = 250$, $A(l, l) = 100$ for $l \in \{R, H, S\}$, $A(R, H) = 97, A(S, R) = 1$, and $A(S, H) = 20$. This choice discourages stroke from neighbouring leukoaraiosis more than it neighbouring healthy tissue. We initialized the posterior estimates using a simple threshold classifier learned from the training subjects [124]. We named our algorithm Cerebro, and the implementation is publicly available[1].

[1]The open source implementation is available at https://github.com/adalca/cerebro

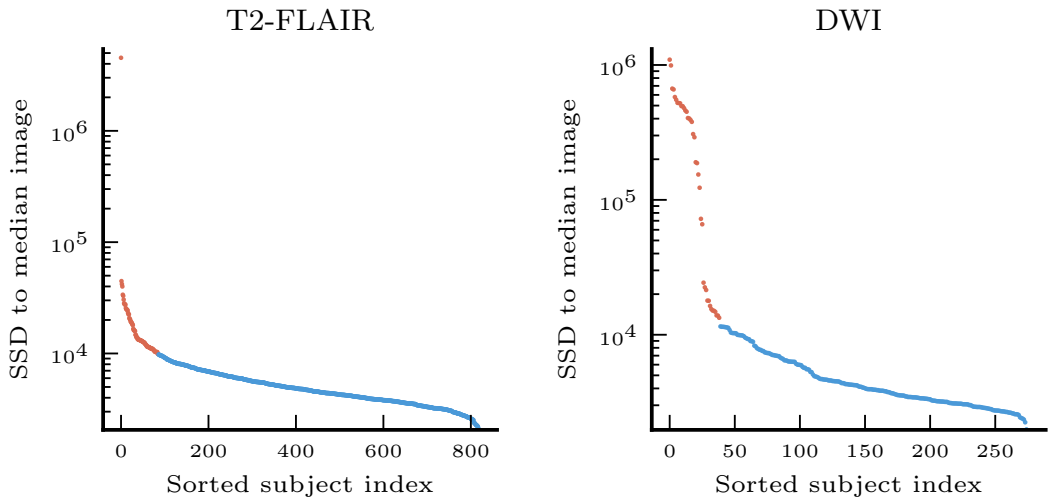

Figure 3.3. Distances from patient images to median image for T2-FLAIR and DWI modalities, ranked in descending order. Outliers are shown in red, and are removed from subsequent analysis.

■ 3.5 Data

We use T2-FLAIR stroke scans (1 × 1mm in-plane, slice thickness 5-7mm, PROPELLER sequence sometimes used if the patient moved). We train the PCA shape model $(\{M_k\}, \Sigma)$ on binary maps of manual leukoaraiosis segmentations in 42 held-out previously segmented scans.

We run our experiments on three groups of subjects. First six test volumes, each with manually delineated leukoaraiosis by *multiple* experts. Second, we use 100 test images with manually delineated leukoaraiosis, selected based on having completed successfully in the pre-processing steps, including registration to an atlas. Finally, we analyze 417 images of subjects from a larger ISGC study. This set underwent less stringent quality control, and includes subjects with significant movement artifacts, and higher variance in ground truth segmentation. While our algorithm was not designed to handle such heterogeneity of data quality, this more tumultuous situation represents a realistic clinical study scenario.

■ 3.6 Preprocessing

To tackle the challenging clinical data, we design robust pre-processing steps. In general, we focus on less accurate but more robust methods to avoid catastrophic failures. We then adjust downstream analyses to account for the less precise processing results [124].

When registering images of the same MR modality, we first compute an initial rigid registration. However, the heterogeneity of intensity distributions in images of the same modality in clinical datasets render the usual intra-modality metrics such as cross correlation (CC) and sum of squared differences (SSD) ineffective. Standard methods for matching intensity profiles, such as histogram equalization, cannot be used either, since they would be dominated by the neck and other extracranial structures. As a result, we employ mutual information (MI) in performing this rigid registration. We build on this initial registration to solve the problems of inconsistent field of view and intensity profiles described above.

We restrict the region where the registration metric is evaluated to the brain. In research-quality images, skull stripping or brain mask extraction assumes that the brain consists of a single connected component separated from the skull and dura by CSF [121]. Unfortunately, such techniques are highly dependent on image quality, and often fail when applied to clinical images. Instead, we propagate a brain mask from the atlas via the estimated rigid transformation. While not a perfect brain mask, it enables intensity correction and constrains the final nonrigid registration to a region that reasonably approximates the brain.

In our experiments, MI failed when used in nonrigid registration, and the images are too heterogeneous for intensity-based measures such as CC and SSD to be applied directly. Using the approximate brain mask, we match the intensity of the white matter, the largest structure in the brain, of each subject with that in the atlas. We estimate the mode of white matter intensity for each patient as the mode of the component with higher intensity in a two-component mixture model for intensity values within the brain mask. Histogram equalization still cannot be used due to the approximate nature of the brain mask and variable intensity profiles.

Following these steps, all subjects are non-rigidly registered to an atlas template using ANTs [4]. Following registration, we propagate anatomical labels, such as the white matter and CSF, from the atlas to each subject. In the experiments in the next section, we only run our algorithm inside the white matter where we expect to see most of the leukoaraiosis and strokes [124].

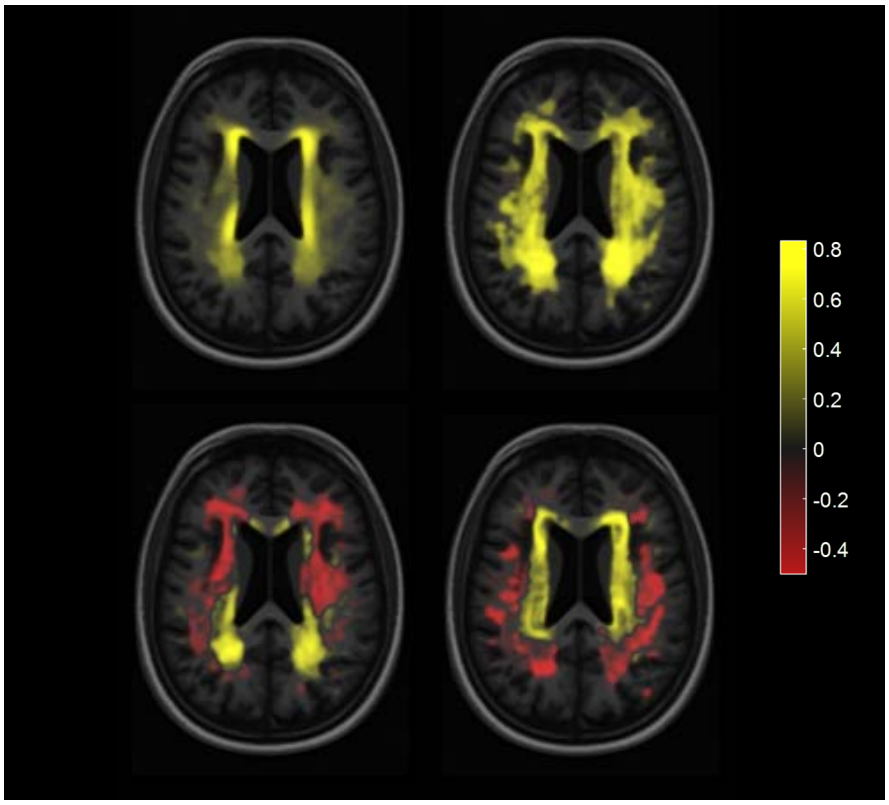

Figure 3.4. The first four modes of leukoaraiosis covariation in the stroke dataset. The spatial maps are shown in color, overlayed on a template brain. The distributions have been slightly smoothed and intensity normalized for visualization purposes. Lateral symmetry, anterior-posterior symmetry, and periventricular co-variation stand out as dominant patterns learned by the model.

In processing of large image collections, fast evaluation of various steps is required. Population statistics are often used to summarize intermediate or final results, to identify outliers, and to evaluate trends [40, 76, 126]. To evaluate registration, the step most prone to error, we devise a new measure of registration quality. We construct a voxel-wise median image of registered subjects in the common atlas space. We then compute the sum of squared differences of each patient image from this median image within the brain mask. Finally, we identify badly registered subjects as those with a difference significantly higher than the rest, via the Tukey fence, or more than 1.5 times the interquartile range above the third quartile [134]. Example statistics for the stroke cohort are shown in Figure 3.3.

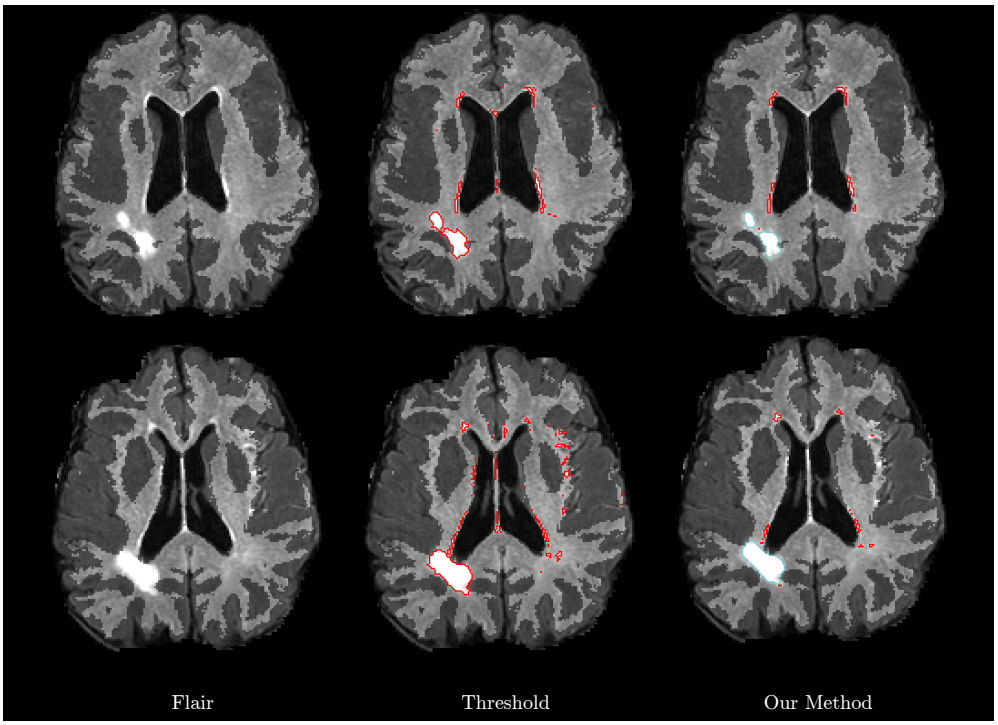

Figure 3.5. Illustrative automatic segmentation result for two slices (top and bottom). Left: The original (skull-stripped) flair image. Middle: hyperintensity segmentation (red) using a threshold-based method tuned to analyzing T2-FLAIR images in stroke patients [124]. Right: our automatic segmentation of leukoaraiosis (red) and chronic lesion (blue). Our method excludes chronic lesions from leukoaraiosis segmentation, as well as other spurious noisy segmentations seen in the results of thresholding.

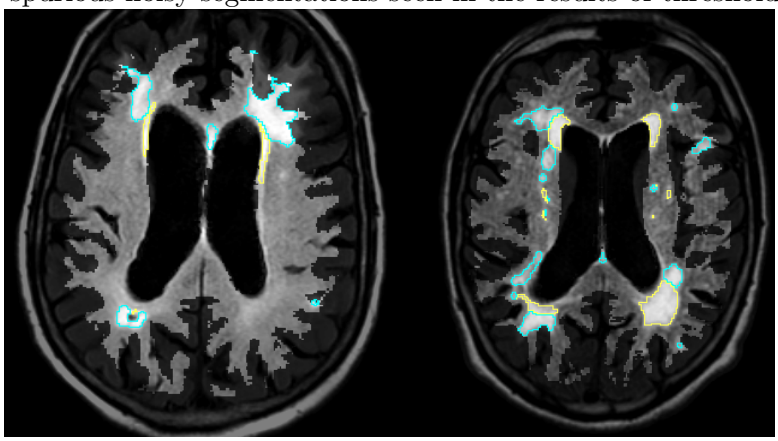

Figure 3.6. Segmentation in difficult images. Automatic segmentation on two difficult subjects with significant chronic lesions showing lesion segmentation (blue outlines) and leukoaraiosis (yellow outlines).

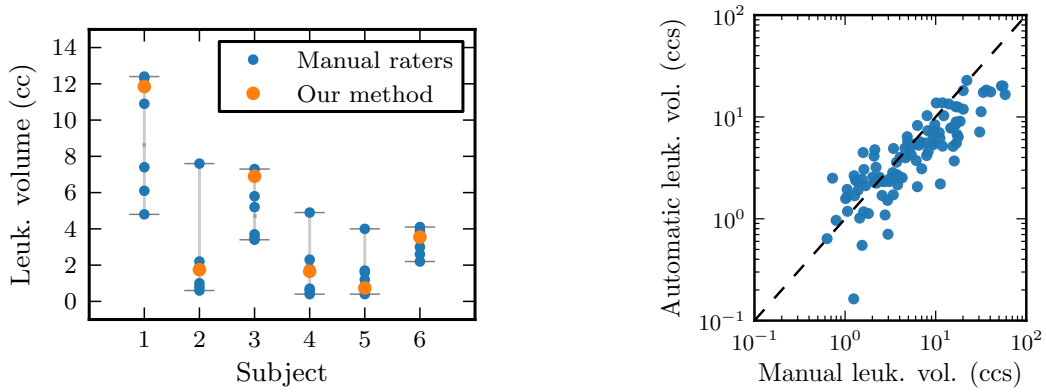

Figure 3.7. Leukoaraiosis segmentation evaluation. Left: Volume measurements based on the automatic segmentation (orange) are within the range of experts (blue circles). Right: Volume estimates based on the automatic segmentation of leukoaraiosis against volume estimates based on the manual segmentations; the correlation coefficient is $r = 0.82$.

■ 3.7 Results

Leukoaraiosis is the primary phenotype in many stroke studies, and thus its segmentation and delineation from stroke is our main focus in the experiments. In Figure 3.4, we show the first four modes of variation of the learned laukoaraiosis maps $\boldsymbol{M}_k$, illustrating the patterns of symmetry and covariation learned by the model. The maps mimic medical expertise, where lateral symmetry, anterior-posterior symmetry, and periventricular co-occurance of leukoaraiosis stand out as dominant co-variation patterns.

Figures 3.5 and 3.6 provides example segmentation results for subjects with leukoaraiosis and chronic stroke. The subject in Figure 3.5, used for parameter tuning, includes a stroke in areas where leukoaraiosis is often found near the ventricles and provides an intuitive illustration of Cerebro's behaviour. The first subject in Figure 3.6 illustrates a typical result, where we see that most of the stroke is accurately separated from leukoaraiosis. The second subject is an example of a more difficult pattern, where the separation of the two hyperintense pathologies is nearly impossible to define even by a clinical expert. In both subjects, our leukoaraiosis segmentation is conservative, which is likely caused by the regularized projection.

In Figure 3.7, we compare the volume of leukoariosis obtained by our method against expert delineations. The first graph demonstrates that the automatic approach is consistently within the range of inter-rater variability. The second graph compares the automatically computed leukoaraiosis burden to that based on manual segmentations

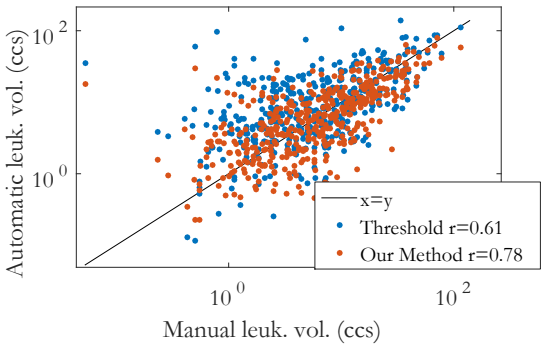

Figure 3.8. Leukoaraiosis segmentation comparison with ground truth for the 417 ISGC subjects. Volumes automatically segmented with our method (orange) have a significantly stronger correlation than those computed with a previous threshold-based segmentation method (blue) specifically designed for stroke patients [124].

in a test set of 100 subjects. Visual inspection of the outlier image reveals little to no apparent leukoaraiosis and a possible manual over-segmentation.

In Figure 3.7, we compare our method with a method tuned to analyzing T2-FLAIR images in stroke patients in the cohort analyzed as part of the ISGC analysis. Briefly, the method learns a hyperintensity threshold learned from manual delineations of ISGC stroke subject scans, taking into account the prior probabilities of healthy and hyperintense tissue. By delineating leukoaraiosis separately from other lesions, our method significantly improves the correlation with manual outlines, from a correlation coefficient of 0.61 to 0.78 in this data set. Visual inspection of the outlier image reveals dramatic motion artifacts and some manual over-segmentation.

■ 3.8 Discussion

In this chapter, we present an algorithm for segmentation of different cerebrovascular pathologies in brain MRI. Our algorithm is derived from a generative probabilistic model that captures experts' knowledge of the disease, learned from a population of subjects from a similar cohort. In essence, the distribution and covariation of pathology in this cohort of stroke patients instructs our algorithm of the patterns expected from such a pathology. By modeling the spatial distribution of leukoaraiosis, as well as the intensities of leukoaraiosis and stroke lesions, our method automatically segments tissues that are indistinguishable based on intensity alone. We presented our method on a study of stroke patients, and demonstrated strong agreement between our results and expert segmentation volumes.

Using our framework, an important extension is to include healthy gray matter and cerebrospinal fluid tissue classes, as well as allow for other modalities such as

diffusion weighted imaging. This will enable more precise exclusion of acute lesions and segmentation of the entire brain. As white matter disease is not specific to stroke and can be observed in the general population, especially in the elderly, the proposed method is broadly applicable beyond the stroke patient cohort that motivated its development. In Chapter 6 we discuss an application to ADNI and several other stroke centers. Learning spatial covariation of disease from a collection of images to infer disease in a new scan can be applied to a spectrum of clinical questions.

Chapter 4

Population Based Image Restoration

We present an algorithm to create high resolution anatomically plausible images consistent with acquired clinical low-resolution brain MRI scans. Although large databases of clinical scans contain a wealth of information, medical acquisition constraints result in scans of low quality and resolution, rendering computational analysis impractical [32, 63, 109, 123]. Even basic tasks, such as skull stripping and non-linear registration, present significant challenges as algorithms' underlying assumptions are violated in clinically acquired scans [52, 121, 124]. In images with wide slice spacing, the image is no longer smooth, and the anatomical structure may change dramatically between consecutive slices (Fig. 4.1). Aiming to take advantage of the dramatic increases in the number and size of patient cohorts in clinical archives, highly specialized or application-specific algorithms that explicitly handle sparse slice spacing or low-resolution promise to address some challenges of analyzing clinical scans. For example, we presented such specialized algorithms to tackle problems in the stroke dataset in Section 3.6. In contrast, here we reconstruct high resolution images that represent plausible anatomy from the low-resolution scans that can then be used with standard image analysis algorithms such as non-linear registration and skull stripping.

We use a large collection of low-resolution scans to help characterize and infer fine-scale anatomy of a particular subject. We introduce models that capture fine-scale structural similarity across subjects in large medical image collections and use it to fill in the missing data in a novel low-resolution scan. Our algorithm produces anatomically plausible volumetric images consistent with low-resolution scans, facilitating computational analysis of clinical scans with existing state of the art techniques. Our method does not require high resolution scans or expert annotations, but instead builds the missing structure by learning from collections of clinical scans of similar quality to that of the input scan. We validate our algorithm on brain images from the ADNI cohort, and find that the method significantly outperforms current upsampling methods.

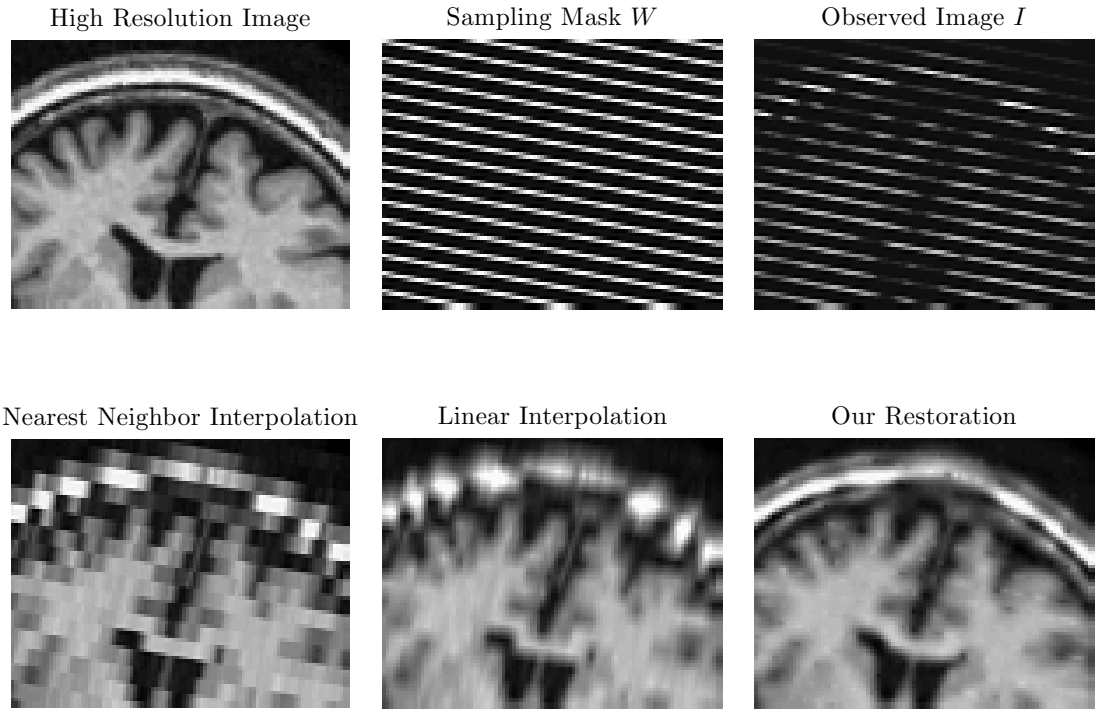

Figure 4.1. Problem setup with close up images including the medial longitudinal fissure. Top row: isotropic image that we seek to recover is sampled according to the sampling mask W leading to the observed image I. Due to affine registration, each mask is not binary, but includes continuous values between 0 and 1. Bottom row: nearest neighbour, linear interpolation, and our restoration. The most dramatic improvement can be seen in restorations of the skull, dura matter, and ventricles.

■ 4.1 Technical Background

Many image restoration techniques depend on having enough information in a single image to synthesize data. Traditional interpolation methods, such as linear, cubic or spline, assume a functional representation of the image. They treat the low resolution voxels as samples, or observations, and estimate function parameters to infer the rest of the voxels [118]. Clinical images are often characterized by sampling that is too sparse to adequately fit such representations and recover the lost detail. For example, 6mm slice spacing in our motivating example is far too high to accurately estimate realistic approximating functions without prior knowledge, as illustrated in Fig. 4.1.

Several recently proposed methods use additional data to synthesize better medical images. Superresolution algorithms often use multiple scans from the same subject to

synthesize a single volume [14, 65, 95]. For example, multiple low resolution acquisitions with small shift differences, or in perpendicular plane directions, might be used. Unfortunately, multiple similar acquisitions are not commonly available in the clinical environment.

Patch methods that instead use fine-scale redundancy within a single scan have been used to improve image quality [77, 78]. These methods have been shown to successfully upsample images by "hallucinating" fine-scale structure [77, 95]. This approach depends on having enough repetitive detail in a scan to capture and re-synthesize the high frequency information. Unfortunately, in many clinical datasets, the thickness or separation of the slices can be 5 to 7 times higher than the in-plane resolution. In such cases, a single image is unlikely to contain enough fine-scale information to provide anatomically plausible reconstruction in the direction of slice acquisition. For example, in the stroke imaging example, interpolated images provided by these methods do not improve downstream analysis.

Some recent methods for superresolution rely on an external dataset of high-resolution scans of the same modality as the low resolution scan. Non-parametric methods fill in missing data by matching a low-resolution patch from the scan with a high resolution patch from the training dataset [17, 60, 65, 69, 112]. Other non-parametric methods capture local similarity via patches in order to propagate segmentation of structures from manual segmentations, but require a ground truth dataset [17, 112]. Some non-parametric methods attempt to improve resolution from a collection of lower resolution images, but they can only tackle slice spacing of roughly three times the in-plane resolution [111].

Parametric local low-dimensional embedding of patches from high resolution images have been used to segment or classify scans [7]. These parametric methods are also often used for restoration of natural images. Recent patch-based algorithms model all natural image patches using a single Principal Component Analysis (PCA), Independent Component Analysis (ICA), or Gaussian Mixture Model (GMM) model. These generic models are not expressive enough to capture fine-scale structure present in low-quality clinical scans [30, 39, 47, 145, 147]. Additionally, high-resolution training datasets are not readily available for all image contrasts and scanners, are costly to obtain, and may not adequately represent pathology or other properties of clinical populations.

Our method takes advantage of the fact that local fine-scale structure is intrinsically shared in a population of medical images, and each low-resolution scan captures some aspect of this structure, as shown in Figure 4.2. Specifically, we model 3D patches from

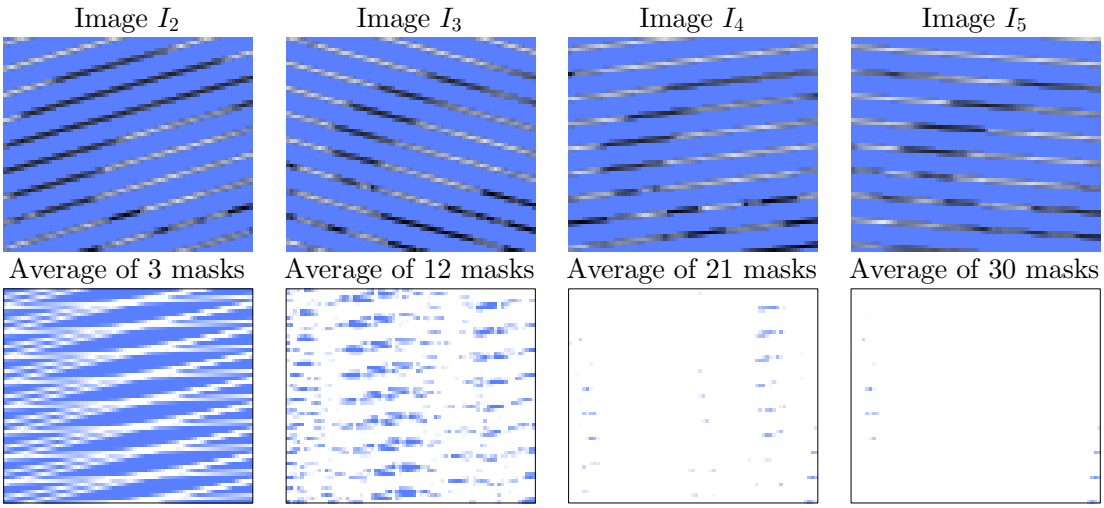

Figure 4.2. Top row: close up images of observed scan planes in the common atlas coordinate system. Bottom row: averages of an increasing number of sampling masks (images indicating where known voxels are present). With 21 images, most of the volume is covered. Due to affine registration, each mask is not binary, but includes continuous values between 0 and 1.

all volumetric images in a collection around a particular location as generated from a Gaussian Mixture Model (GMM), with partially missing data. The intuition is that locally, subgroups of subjects share similar anatomical structure. We investigate two different ways of modelling the missing data and present a model for restoration of local patches. We derive an iterative algorithm to learn the parameters of the model, and show how to use the model parameters to reconstruct entire anatomically consistent high-resolution images from clinical scans. The model can naturally be trained from datasets of various quality, from high resolution research datasets to just clinical images with sparse slices. We demonstrate our algorithm using scans from the ADNI cohort, including scans from subjects with Alzheimer's disease diagnosis. Although we are motivated by the example of sparse slices, the resulting method can be used for general image restoration from other types of sparse sampling patterns.

■ 4.2 Patch Mixture Model

For clarity, we first consider training a model from a collection of high-resolution scans. In the next section, we discuss the model changes given the realistic situation of training from a collection of low-resolution scans.

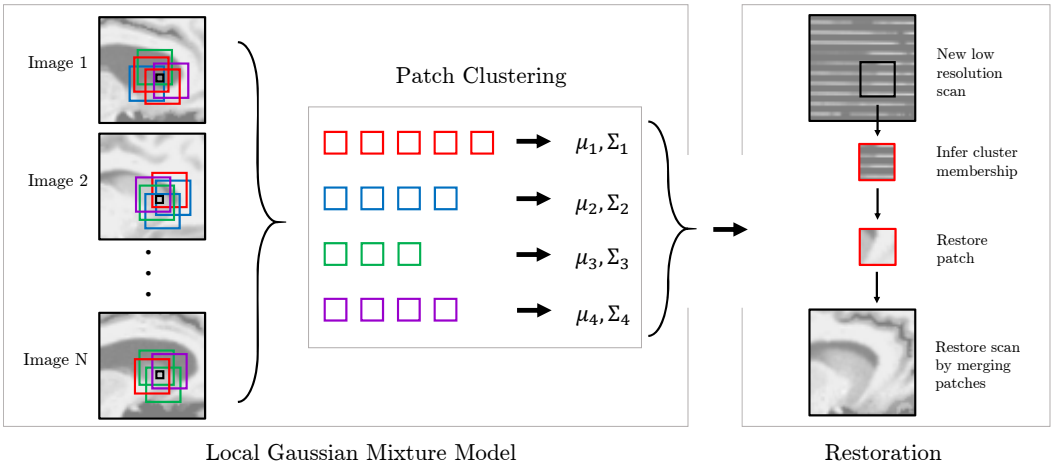

Figure 4.3. Schematic of our image restoration approach. To capture the structure of the training set, we learn a Gaussian Mixture of patches, which involves modelling missing data and low-dimensional embeddings of the patches. For clarity, the training images are shown as high-resolution in this schematic, but our application only involves sparse images, even for training. We illustrate the method with 2D patches and images, but the algorithms operate on 3D volumes and 3D patches. Given a new image, we infer which cluster each image patch belongs to, and restore the missing data using the learned model and the known voxels. We quilt the final volume from overlapping restored patches.

We capture locally similar structure across the collection using 3D image patches. Specifically, we think of all 3D patches across the entire image collection at a particular template location u as being drawn from a Gaussian Mixture Model (GMM). We further model the patch space of each cluster as a low-dimensional subspace using a reduced-dimensionality Gaussian distribution, with the intuition that image patches have a significantly smaller effective dimensionality than the number of voxels in the patch. Given this parametric representation of a patch at each location, we can directly reconstruct a full 3D patch at the same location from a new low-resolution scan. Given reconstructed patches, we merge them into a restored high-resolution volume. Our method is summarized schematically in Figure 4.3.

Appendix B contains detailed derivations of the algorithms described in this chapter.

Model

Let $\{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_N\}$ be an image collection roughly aligned (we use affine transformations in our experiments) into a coordinate atlas common space. Our method does not assume precise alignment and can handle misalignments produced by affine registration. For a new image $\boldsymbol{Y}_n$ where only a few slices are known, we seek to restore an anatomically plausible high-resolution image. We capture local structure using image patches. Throughout this chapter, we assume constant patch shape, and use $\boldsymbol{y}_i(u)$ to denote a d-length vector that contains voxels of the image patch centered at location u of image $\boldsymbol{Y}_i$. In all methods described here, we perform inference at each location independently, and drop u from our notation for simplicity.

We let $\boldsymbol{\Gamma}_i$ and $\boldsymbol{R}_i$ be the interpolation matrices that transform the original subject space patch $\tilde{\boldsymbol{y}}_i$ to the common coordinate atlas space patch $\boldsymbol{y}_i$, and vice versa:

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i \\ \tilde{\boldsymbol{y}}_i &= \boldsymbol{R}_i \boldsymbol{y}_i \end{aligned} \tag{4.1}$$

We therefore have $\boldsymbol{R}_i \boldsymbol{\Gamma}_i \approx \boldsymbol{I}_d$, and treat $\boldsymbol{R}_i$ as a generalized inverse of $\boldsymbol{\Gamma}_i$. For simplicity, we further assume $\boldsymbol{R}_i$ is square[1]. The distinction of the two reference spaces is important in restoration and modelling of the missing data.

We first model the set of common space patches $\{\boldsymbol{y}_i\}_{i=1}^N$ as drawn from a K-component multivariate Gaussian Mixture Model (GMM). Additionally, each patch $\boldsymbol{y}_i$ in cluster k is a high dimensional observation of a latent low dimensional patch $\boldsymbol{x}_i$ of dimensionality p:

$$\boldsymbol{y}_i = \boldsymbol{W}_k \boldsymbol{x}_i + \boldsymbol{\mu}_k + \boldsymbol{\epsilon}_i, \tag{4.2}$$

$$\boldsymbol{x}_i \sim \mathcal{N}(0, \boldsymbol{I}), \tag{4.3}$$

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}), \tag{4.4}$$

where $\boldsymbol{W}$ is a $d \times p$ matrix, $\boldsymbol{x}$ is a $p \times 1$ vector with $p < d$, and σ^2 is the noise variance. Each component is therefore distributed normally with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{C}_k \equiv \boldsymbol{W}_k^T \boldsymbol{W}_k + \sigma^2 \boldsymbol{I}$. Letting $\mathcal{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_n\}$, the likelihood of the observed data under

[1] Depending on the affine transformation model used, matrix $\boldsymbol{R}_i$ might not be square, but will usually be close.

our model is

$$p(\mathcal{Y}; \boldsymbol{\mu}, \boldsymbol{W}, \boldsymbol{\pi}, \sigma^2) = \prod_i p(\boldsymbol{y}_i; \boldsymbol{\mu}, \boldsymbol{W}, \boldsymbol{\pi}, \sigma^2)$$
$$= \prod_i \sum_k \pi_k \mathcal{N}(\boldsymbol{y}_i; \boldsymbol{\mu}_k, \boldsymbol{W}_k^T \boldsymbol{W}_k + \sigma_k^2 \boldsymbol{I}). \tag{4.5}$$

Learning

We employ Expectation-Maximization [31] to learn the optimal parameters $\{\boldsymbol{\mu}_k\}, \{\boldsymbol{W}_k\}, \sigma$ and $\boldsymbol{\pi}$ under likelihood (4.5). Here, we provide the resulting updates along with their interpretations, and include detailed derivations in Appendix B [130].

In the **expectation step**, we update the membership of patch i:

$$\gamma_{ik} \leftarrow \frac{\pi_k \mathcal{N}(\boldsymbol{y}_i; \boldsymbol{\mu}_k, \boldsymbol{C}_k)}{\sum_l \pi_l \mathcal{N}(\boldsymbol{y}_i; \boldsymbol{\mu}_l, \boldsymbol{C}_l)}. \tag{4.6}$$

In the **maximization step**, we update the standard GMM parameters given our expected memberships γ_{ik}. In addition to the means $\boldsymbol{\mu}_k$ and cluster priors π_k, we introduce the *empirical* covariance $\boldsymbol{\Sigma}_k$:

$$\boldsymbol{\mu}_k \leftarrow \frac{\sum_i \gamma_{ik} \boldsymbol{y}_i}{\sum_i \gamma_{ik}} \tag{4.7}$$

$$\pi_k \leftarrow \frac{1}{N} \sum_i \gamma_{ik} \tag{4.8}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{\sum_i \gamma_{ik}} \sum_i \gamma_{ik} \left[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T \right]. \tag{4.9}$$

Since each cluster is modelled as a low dimensional Gaussian distribution, we seek the low-dimensional approximation $\boldsymbol{C}$ to $\boldsymbol{\Sigma}$. Let $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T = \text{SVD}(\boldsymbol{\Sigma})$ be the singular value decomposition of $\boldsymbol{\Sigma}$ where we only keep the first p eigenvectors with p being the desired dimensionality of our embedded model (4.4), $\boldsymbol{U}$ is a $d \times p$ orthonormal matrix of eigenvectros, $\boldsymbol{\Lambda}$ is a $p \times p$ diagonal matrix of eigenvalues, and $\boldsymbol{V}$ is a $p \times p$ rotation matrix. We obtain the parameters of the low-dimensional decomposition:

$$\sigma_k^2 \leftarrow \frac{1}{d-q} \sum_{j=p+1}^{d} \boldsymbol{\Lambda}_k(j,j) \tag{4.10}$$

$$\boldsymbol{W}_k \leftarrow \boldsymbol{U}_k (\boldsymbol{\Lambda}_k - \sigma_k^2 \boldsymbol{I})^{1/2}, \tag{4.11}$$

where $\sigma_k^2 \boldsymbol{I}$ measures the variance lost by using a low-dimensional representation of the patch covariance and $\boldsymbol{W}_k$ is the current estimate of the principal axes of cluster k.

In summary, each iteration first updates the cluster memberships. The mean and covariance of each cluster are estimated using the standard update equations. The covariances are then projected to a low-dimensional space via SVD. The expectation and maximization steps are iterated until convergence of the likelihood (4.5).

In traditional probabilistic PCA [131], an alternative EM algorithm is proposed based on the modelling of $\{\boldsymbol{x}_i\}$ as latent variables. We avoid this modelling choice here, as the current setup provides a more useful introduction to the missing data framework of the next section. We will come back to this point in next section.

Restoration

Let $\tilde{\boldsymbol{Y}}$ be a sparse scan that we wish to restore. $\tilde{\boldsymbol{Y}}$ is affinely transformed into the common atlas space via matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{R}$ as defined in (4.1). The original image includes observed and missing voxels. In the transformed image $\boldsymbol{Y}$, the voxels are interpolated. Our goal is to restore the original image to high resolution. Let $\boldsymbol{y}$ indicate a vectorized patch at a particular location in the atlas space. We first estimate the cluster membership for this patch. Given the equivalences $\boldsymbol{y} = \boldsymbol{\Gamma}\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{y}} = \boldsymbol{R}\boldsymbol{y}$ (4.1), we can transform the mean and covariance of the selected cluster k from the common atlas space in which they are computed, to the subject's original space:

$$\tilde{\boldsymbol{\mu}}_k = \boldsymbol{R}\boldsymbol{\mu}_k \tag{4.12}$$

$$\tilde{\boldsymbol{C}}_k = \boldsymbol{R}\boldsymbol{C}_k\boldsymbol{R}^T. \tag{4.13}$$

We denote the missing voxels as $\tilde{\boldsymbol{y}}^{\mathcal{M}}$, and let $\tilde{\boldsymbol{y}}^{\mathcal{O}}$ be the observed entries. Similarly, $\boldsymbol{C}^{\mathcal{A},\mathcal{B}}$ is the sub-matrix of covariance $\boldsymbol{C}$ selecting rows in set $\mathcal{A}$ and columns in set $\mathcal{B}$, $\mathcal{A}, \mathcal{B} \in \{\mathcal{O}, \mathcal{M}\}$. As shown in Appendix B.1,

$$p(\tilde{\boldsymbol{y}}^{\mathcal{M}}|\tilde{\boldsymbol{y}}^{\mathcal{O}}) = \mathcal{N}(\boldsymbol{\mu}^{\mathcal{M}|\mathcal{O}}, \boldsymbol{C}^{\mathcal{M}|\mathcal{O}}) \tag{4.14}$$

where

$$\tilde{\boldsymbol{\mu}}^{\mathcal{M}|\mathcal{O}} = \tilde{\boldsymbol{\mu}}^{\mathcal{M}} + \tilde{\boldsymbol{C}}^{\mathcal{M}\mathcal{O}}(\tilde{\boldsymbol{C}}^{\mathcal{O}\mathcal{O}})^{-1}(\boldsymbol{y}^{\mathcal{O}} - \tilde{\boldsymbol{\mu}}^{\mathcal{O}}) \tag{4.15}$$

$$\tilde{\boldsymbol{C}}^{\mathcal{M}|\mathcal{O}} = \tilde{\boldsymbol{C}}^{\mathcal{M}\mathcal{M}} - (\tilde{\boldsymbol{C}}^{\mathcal{O}\mathcal{M}})^T(\tilde{\boldsymbol{C}}^{\mathcal{O}\mathcal{O}})^{-1}\tilde{\boldsymbol{C}}^{\mathcal{O}\mathcal{M}} \tag{4.16}$$

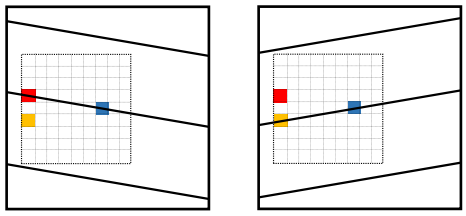

Figure 4.4. Covariance example. The red and orange voxels are never observed together in the same patch, but we should be able to learn their correlation from the sparse patch observations shown here, using the red-blue and orange-blue correlations. Modelling low dimensional patch representations $\boldsymbol{x}$ helps achieve this.

Therefore, we restore the given patch by setting

$$\tilde{\boldsymbol{y}}^{\mathcal{M}} = \tilde{\boldsymbol{\mu}}^{\mathcal{M}|\mathcal{O}} = \tilde{\boldsymbol{\mu}}^{\mathcal{M}} + \tilde{\boldsymbol{C}}^{\mathcal{M}\mathcal{O}}(\tilde{\boldsymbol{C}}^{\mathcal{O}\mathcal{O}})^{-1}(\boldsymbol{y}^{\mathcal{O}} - \tilde{\boldsymbol{\mu}}^{\mathcal{O}}). \tag{4.17}$$

Essentially, the missing values are imputed based on the known values, and the learned covariances between the missing and observed indices.

Finally, we use standard patch quilting techniques to combine overlapping patches [38].

■ 4.3 Patch Mixture Models with missing Data

Given a collection of only low-resolution images $\{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_N\}$, we explicitly model the missing data, learn a Gaussian Mixture Model, and use the same restoration strategy as the previous section. Intuitively, learning such a model with sparse data is possible because each image patch in the training set gives us slightly different voxels observations that contribute to the Gaussian mixtures mean and covariance estimation (Figure 4.2). Our models use low dimensional Gaussians to capture a low-dimensional representation of the patches. First, this is a more robust representation than full Gaussians, since image patches are spatially smooth and have significantly lower intrinsic dimensionality than the raw number of voxels. Secondly, for patches with sparse planes, it is likely that some voxel *pairs* are never observed in the same patch, as shown in Figure 4.4. This results in missing observations for that entry of the empirical covariance matrix. Representing covariance using low-rank approximations alleviates this effect.

We present two different approaches and discuss strengths and shortcomings of each. Appendix B provides derivations for both models of mixtures of principal subspaces given missing data. Both models assume that the data is missing at random, meaning that whether a voxels is observed is independent of the value of the patch voxels. Although the voxels missing in the patches of sparsely sampled images clearly form a

spatial pattern, we assume there is no correlation with the actual intensity of the voxels.

4.3.1 Latent Missing Data

As before, we aim to capture local structure across the collection using image patches. We model the atlas-space patches as coming from a Gaussian Mixture of Gaussian distributions:

$$\boldsymbol{y}_i \sim \sum_k \pi_k \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{C}_k), \tag{4.18}$$

where each mixture has a low dimensional representation, and as before $\boldsymbol{C}_k = \boldsymbol{W}_k^T \boldsymbol{W}_k + \sigma_k^2 \boldsymbol{I}$. In contrast to the previous section, we work directly with missing and observed voxels in the subject space:

$$\tilde{\boldsymbol{y}}_i \sim \sum_k \pi_k \mathcal{N}(\tilde{\boldsymbol{\mu}}_{ki}, \tilde{\boldsymbol{C}}_{ki}), \tag{4.19}$$

where

$$\tilde{\boldsymbol{\mu}}_{ki} = \boldsymbol{R}_i \boldsymbol{\mu}_k \tag{4.20}$$

$$\tilde{\boldsymbol{C}}_{ki} = \boldsymbol{R}_i \boldsymbol{C}_k \boldsymbol{R}_i^T, \tag{4.21}$$

are the GMM component parameters transformed to the subject space for each subject.

We use $\tilde{\boldsymbol{y}}_i^{\mathcal{M}_i}$ to denote missing voxels, and let $\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i}$ be the observed entries. We model the missing entries as latent random variables, and let $\tilde{\mathcal{Y}}^{\mathcal{O}} = \{\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i}\}_{i=1}^N$ and $\mathcal{R} = \{\boldsymbol{R}_i\}_{i=1}^N$.

The observed data likelihood is:

$$p(\tilde{\mathcal{Y}}^{\mathcal{O}}; \mathcal{R}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \prod_i \sum_k \pi_k \mathcal{N}(\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i}, \tilde{\boldsymbol{\mu}}_{ki}^{\mathcal{O}_i}; \tilde{\boldsymbol{C}}_{ki}^{\mathcal{O}_i \mathcal{O}_i}) \tag{4.22}$$

Learning

To estimate the likelihood in the presence of missing data, we use the Expectation Conditional Maximization (ECM) variant of the Generalized Expectation Maximization algorithm [74]. Here we summarize the intuitive updates, while the full derivation can be found in Appendix B.

The **expectation step** updates the class membership and the missing data statistics. The class membership is computed based on the known voxels in the original

subject space:

$$\gamma_{ik} \leftarrow \frac{\pi_k \mathcal{N}(\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i}; \tilde{\boldsymbol{\mu}}_{ki}^{\mathcal{O}_i}, \tilde{\boldsymbol{C}}_{ki}^{\mathcal{O}_i})}{\sum_k \pi_k \mathcal{N}(\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i}; \tilde{\boldsymbol{\mu}}_{ki}^{\mathcal{O}_i}, \tilde{\boldsymbol{C}}_{ki}^{\mathcal{O}_i})} \tag{4.23}$$

The statistics of the missing data are computed based on covariates of the known and unknown voxels in the original subject space:

$$\tilde{y}_{ij} \leftarrow \begin{cases} \tilde{y}_{ij} & \text{if } \tilde{y}_{ij} \text{ is observed} \\ \tilde{\mu}_{ij} + \tilde{\boldsymbol{C}}_{ki}^{j\mathcal{O}_i} (\tilde{\boldsymbol{C}}_{ki}^{\mathcal{O}_i\mathcal{O}_i})^{-1} (\tilde{\boldsymbol{y}}_{ki}^{\mathcal{O}_i} - \tilde{\boldsymbol{\mu}}_k^{\mathcal{O}_i}) & \text{otherwise} \end{cases} \tag{4.24}$$

$$\tilde{s}_{ijl} \leftarrow \begin{cases} 0 & \text{if } \tilde{y}_{ij} \text{ or } \tilde{y}_{ik} \text{ is observed} \\ \tilde{\boldsymbol{C}}_{ki}^{jk} - (\tilde{\boldsymbol{C}}_{ki}^{\mathcal{O}_i j})^T (\tilde{\boldsymbol{C}}_{ki}^{\mathcal{O}_i\mathcal{O}_i})^{-1} \tilde{\boldsymbol{C}}_{ki}^{\mathcal{O}_i l} & \text{otherwise} \end{cases} \tag{4.25}$$

where the correction in $\tilde{s}_{ijl}$ can be interpreted as the uncertainty in the covariance estimation due to the missing values.

Given estimates for the missing data, the **maximization step** proceeds very similarly to the full model from the previous section. While missing data is estimated in each subject's original space using transformed means and covariances, the GMM parameters are estimated in the common atlas space using $\boldsymbol{\Gamma}_i$:

$$\boldsymbol{\mu} \leftarrow \frac{1}{\sum_i \gamma_{ik}} \sum_i \gamma_{ik} \boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i \tag{4.26}$$

$$\boldsymbol{\Sigma} \leftarrow \frac{1}{\sum_i \gamma_{ik}} \sum_i \gamma_{ik} \left[(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})^T + \boldsymbol{\Gamma}_i \tilde{\boldsymbol{S}}_i \boldsymbol{\Gamma}_i^T \right] \tag{4.27}$$

$$\pi_k \leftarrow \frac{1}{N} \sum_i \gamma_{ik}. \tag{4.28}$$

where $\left[\tilde{\boldsymbol{S}}_i\right]_{jk} = \tilde{s}_{ijk}$. The mean update transforms each estimated subject patch to atlas space and obtains the average patch. The covariance update similarly uses transformed patches, with the additional covariance uncertainty term. As before, letting $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T = \text{SVD}(\boldsymbol{\Sigma})$ be the singular value decomposition of $\boldsymbol{\Sigma}$ leads to the low dimensional updates

$$\sigma_k^2 \leftarrow \frac{1}{d-q} \sum_{j=d+1}^{d} \boldsymbol{\Lambda}_k(j,j) \tag{4.29}$$

$$\boldsymbol{W}_k \leftarrow \boldsymbol{U}_k (\boldsymbol{\Lambda}_k - \sigma_k^2 \boldsymbol{I})^{1/2}. \tag{4.30}$$

Finally, we let $\boldsymbol{C}_k = \boldsymbol{W}_k^T \boldsymbol{W}_k + \sigma_k^2 \boldsymbol{I}$. The expectation (4.23), (4.25) and maximiza-

tion (4.28), (4.30) steps are iterated until convergence.

Note that the main difference between this model and the full model in Section 4.2 is the estimation of the missing data statistics in the expectation step. The ECM algorithm maintains the guarantee that the expectation of the likelihood increases[2] although the maximization steps updates are not closed form. The most notable difference between this formulation and simpler algorithms that iteratively fill in missing voxels and then update the covariance is the additional term in $\tilde{\boldsymbol{S}}_i$ (4.25), which captures the covariance of the missing and observed data. For a more detailed discussion on this topic, see [74], Chapter 8.

Restoration

Using the converged parameters π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{C}_k$, we use the same restoration strategy as presented in 4.2 to reconstruct high-resolution volumes. The image being restored could be from the low-resolution training collection itself or from a completely new subject.

■ 4.3.2 Latent Low-Dimensional Patches

In this section we model the low-dimensional patch representations $\boldsymbol{x}_i$ as latent variables, and do not explicitly model the missing voxels. Intuitively, this means that the low-dimensional mixtures will be learned using only the observed data, rather than imputing missing voxel statistics. In this section we work in the common atlas space only, and assume that voxels are either known or unknown in this space.

Within each mixture, we model the observed data as being sampled from a low-dimensional principal analyser:

$$\boldsymbol{y}_i^{\mathcal{O}_i} = \boldsymbol{W}_k^{\mathcal{O}_i}\boldsymbol{x}_i + \boldsymbol{\mu}_k^{\mathcal{O}_i} + \boldsymbol{\epsilon}_i \tag{4.31}$$

$$\boldsymbol{x}_i \sim \mathcal{N}(0, \boldsymbol{I}) \tag{4.32}$$

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma_k^2 \boldsymbol{I}) \tag{4.33}$$

where the notation $\boldsymbol{W}_k^{\mathcal{O}_i}$ extracts the rows of $\boldsymbol{W}_k$ corresponding to the observed values

[2] $Q(\boldsymbol{\theta}|\boldsymbol{\theta}) \geqslant Q(\boldsymbol{\theta}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ includes all the parameters of the model, and $Q(\boldsymbol{\theta}|\boldsymbol{\theta})$ is the expectation of the log likelihood under the estimated posterior.

of $\boldsymbol{y}_i$. The likelihood of the observed data is

$$p(\mathcal{Y}^{\mathcal{O}}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \prod_i \sum_k \pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{C}_k^{\mathcal{O}_i \mathcal{O}_i}), \tag{4.34}$$

where $\boldsymbol{C}_k^{\mathcal{O}_i \mathcal{O}_i} = (\boldsymbol{W}_k^{\mathcal{O}_i})^T \boldsymbol{W}_k^{\mathcal{O}_i} + \sigma_k^2 \boldsymbol{I}$.

Learning

To estimate the likelihood, we use the ECM algorithm, and again summarize the intuitive updates, leaving the full derivation for Appendix B.

The **expectation step** updates the expected class membership:

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{C}_k^{\mathcal{O}_i})}{\sum_k \pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{C}_k^{\mathcal{O}_i})} \tag{4.35}$$

and the low dimensional data statistics which are intuitively computed based on just the known voxels in the original subject space:

$$\boldsymbol{S}_{ki} \leftarrow \sigma_k^2 ((\boldsymbol{W}_k^{\mathcal{O}_i})^T \boldsymbol{W}_k^{\mathcal{O}_i} + \sigma_k^2 \boldsymbol{I})^{-1} \tag{4.36}$$

$$\widehat{\boldsymbol{x}}_{ki} \leftarrow \boldsymbol{S}_{ki} \frac{1}{\sigma_k^2} (\boldsymbol{W}_k^{\mathcal{O}_i})^T (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{\mu}_k^{\mathcal{O}_i}) \tag{4.37}$$

The statistics $\widehat{\boldsymbol{x}}_{ki}$ and $\boldsymbol{S}_{ki} = \sigma_k^2 \boldsymbol{C}_{ki}^{-1}$ can be interpreted as expected low-dimensional patch and precision. The **maximization step** follows standard parameter updates. The mean is the average residual of the estimated SVD projection and the observed values:

$$[\mu_k]_j = \mu_{kj} \leftarrow \frac{1}{\sum_{i \in \mathcal{O}^j} \gamma_{ik}} \sum_{i \in \mathcal{O}^j} \gamma_{ik} (y_{ij} - \boldsymbol{W}_k^j \widehat{\boldsymbol{x}}_{ki}) \tag{4.38}$$

where $\mathcal{O}^j$ is the set of patches that have entry j observed. The principal axes and residual variance are estimated based on the current estimates of data statistics:

$$\boldsymbol{W}_k^j \leftarrow (\sum_{i \in \mathcal{O}^j} \gamma_{ik} (\widehat{\boldsymbol{x}}_{ki} \widehat{\boldsymbol{x}}_{ki}^T + \boldsymbol{S}_{ki}))^{-1} \sum_{i \in \mathcal{O}^j} \gamma_{ik} (y_{i,j} - \mu_{kj}) \boldsymbol{x}_i^T \tag{4.39}$$

$$\sigma_k^2 \leftarrow \sum_{i \in \mathcal{O}^j} \gamma_{ik} \left[(y_{ij} - \boldsymbol{W}_k^j \widehat{\boldsymbol{x}}_{ki} - \mu_{kj})^2 + \boldsymbol{W}_k^j \boldsymbol{S}_{ki} (\boldsymbol{W}_k^j)^T \right]. \tag{4.40}$$

The cluster probabilities are generated with the usual update:

$$\pi_k = \frac{1}{N} \sum_i \gamma_{ik}. \tag{4.41}$$

Finally, we let $\boldsymbol{C}_k = \boldsymbol{W}_k^T \boldsymbol{W}_k + \sigma_k^2 \boldsymbol{I}$ upon convergence.

Restoration

Using the cluster parameters π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{C}_k$, we use the same restoration strategy as presented in 4.2 to reconstruct high-resolution volumes. Again, the image being restored can be from the same low-resolution training collection or be a completely new scan.

4.3.3 Implementation

We stack together the affinely registered images from our entire collection, and split the stack into overlapping subvolumes of size $18 \times 18 \times 18$. Within each subvolume, we learn the mixture model of low-dimensional latent variables presented in Section 4.3.2. Instead of just choosing one patch from each volume at a given location, we collect all patches within the small subvolume centered at that location. This provides more data for each model, which is crucial when given very sparse data. Moreover, it offers robustness in the face of slight misalignment of images. Given the learned parameters at each location, we restore all patches within the subvolume. We use a patch size of 9mm$\times$9mm$\times$9mm, and use $K = 15$ clusters. Our implementation is freely available[3].

4.4 Results

We evaluate our algorithm using 826 T1-weighted brain MR images from the ADNI dataset [63]. We downsample the isotropic 1mm images to 1mm $\times$ 1mm $\times$ 5mm, and use these low-resolution images as our training dataset. We learn the local patch model at dense locations in the downsampled images and reconstruct 25 scans. While we experimented with both methods presented in this chapter, we found the model explicitly estimating missing data (Section 4.3.1) to be prohibitively slow. Instead, we use the

[3]The open-source code can be found at `https://github.com/adalca/papago`. We also implement a general library for working with patches in MATLAB that we heavily use in this project. This library is similarly open source and can be found at `https://github.com/adalca/patchlib`.

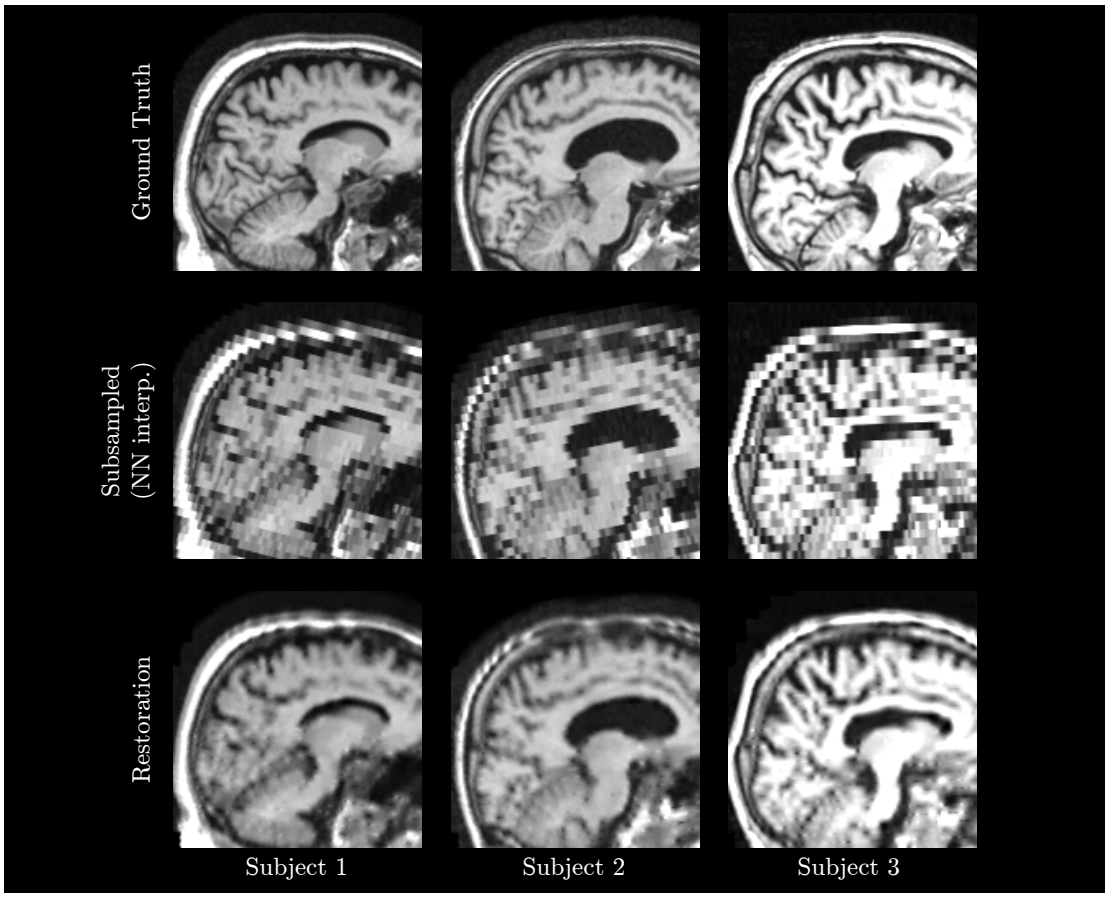

Figure 4.5. Restoration in variable images. Three affinely-aligned volumes illustrate the typical variability in the scans. While most anatomy is grossly aligned, the methods must be robust against significant local variability. The images also have some slightly different intensity profiles, which are hard to normalize due to the quality of the scans. Nonetheless, the proposed method restores the images accurately.

latent low-dimensional patch model (Section 4.3.2) throughout these experiments. We discuss this choice in the next section. We also use three commonly used methods as a baseline for comparison: nearest neighbour interpolation, non-local means (NLM) upsampling [77], and linear interpolation. NLM is often used for image upsampling, although it was not designed for clinical images with such sparse resolution. We validate our algorithm by comparing reconstructions to the original isotropic volumes both visually and quantitatively.

Figure 4.5 shows restorations for three scans, illustrating the variability of the anatomical structure and intensity patterns. Despite significant variability in anatomi-

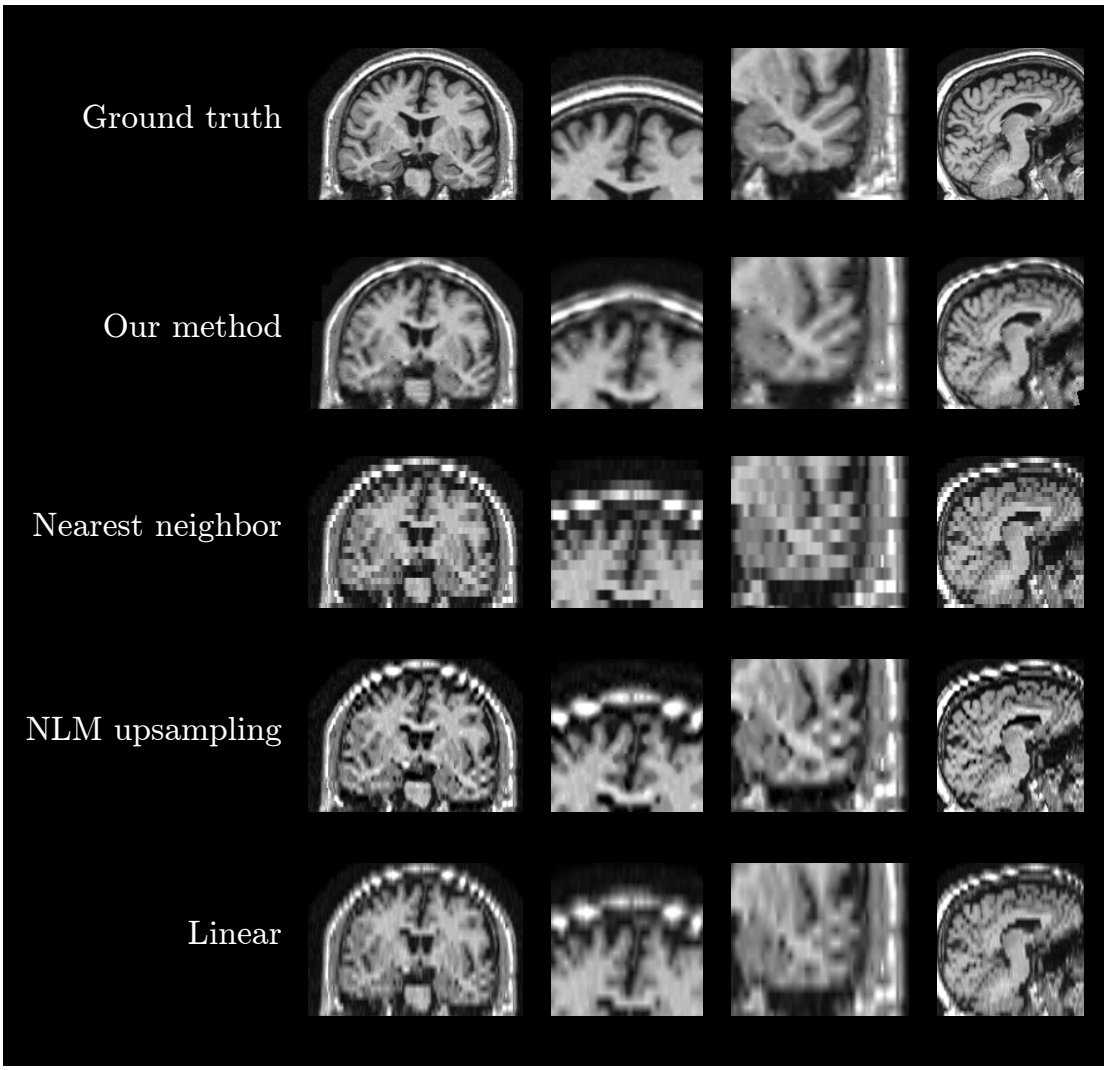

Figure 4.6. Representative restorations. A representative original high resolution image, a reconstruction by our method as well as three standard methods: nearest neighbor, NLM, and linear interpolation. Our method reconstructs more anatomically plausible substructures, as can be seen in both Axial and Saggital planes.

cal shape and intensity, our method restores the fine-scale anatomy. Figure 4.6 shows representative restored images for one subject. Our reconstruction method produces significantly more anatomically plausible regions in both the coronal and saggital slices. The magnified axial crops highlight that our method is able to restore anatomical structures that are almost entirely missing in the other methods, such as the dura mater or

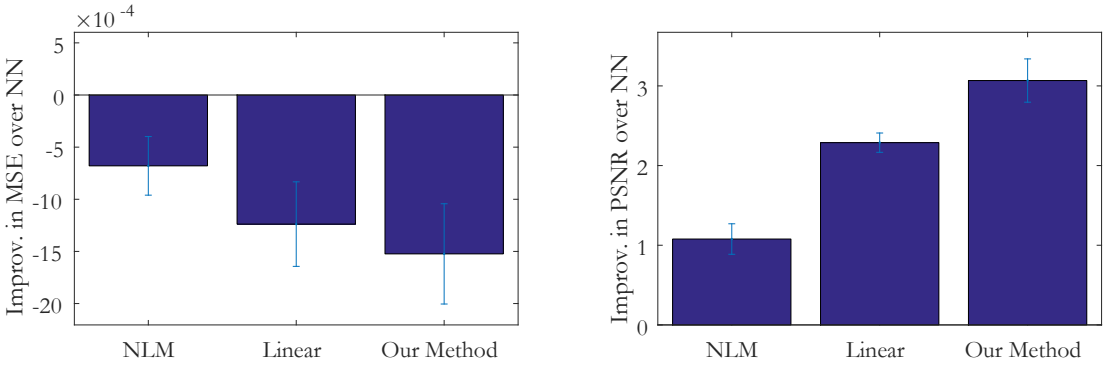

Figure 4.7. Errors for image restorations compared to ground truth isotropic scans. Due to high variability among subject images, we illustrate improvements of NLM upsampling, linear interpolation, and our method over nearest neighbor interpolation. Mean squared error (lower is better), and PSNR (higher is better) illustrate that our method recovers significantly better signal.

sulci of the temporal lobe.

Figure 4.7 reports errors of restored images compared to ground truth. Due to high variability among subject scans, we plot improvements of each method over nearest neighbor interpolation. We first measure the mean squared error (MSE) between a restored scan and the true isotropic image taken over all voxels. Nearest neighbor MSE was 0.003 ± 0.001. Secondly, we measure the peak signal to noise ratio (PSNR), which is the log transform of the ratio of the maximum intensity to the MSE. Nearest neighbor PSNR was 25.3 ± 1.4. Both metrics are commonly used in measuring reconstruction of compressed or noisy signal. Our restoration algorithm offers significant improvement compared to the other methods.

■ 4.5 Discussion

We use a large collection of low-resolution images to infer fine-scale anatomy of a particular subject. We introduce two models that capture structural similarity across subjects in such large clinical image collections and use it to fill in the missing data in a low resolution scan. The results of our algorithm are anatomically plausible volumetric images consistent with low resolution scans. Our method does not require high resolution scans or expert annotations, but can instead build the missing structure by learning from collections of clinical scans of similar quality to that of the low-quality input image.

Model Comparison

The first model (Section 4.3.1) explicitly models missing voxels, and the learning procedure infers missing data statistics at each iteration. Visually, we found this method to perform well in most subvolumes we experimented with. However, the method has a long convergence time, requiring more than 25 iterations to converge, and each iteration is slow due to large matrix multiplication and inverses involved in the update computations for each patch. We also found the method to be sensitive to the initialization used, where initializing the missing voxels with cluster averages yielded dramatically different results compared to initializing with linear interpolation.

The second model (Section 4.3.2) explicitly ignores missing voxels, and models the low-dimensional representation of each patch. This has the advantage of avoiding potential local minima during learning, and is more robust to initialization. Additionally, we observed significantly faster convergence rates and quicker update due to small matrices involved in computations. While this method produces the results shown in the previous section, it had lower performance in particular subvolumes compared to the first model due to overall smoother appearance. In the future, fast approximations to the update equations used in the first method may offer competitive runtimes and results.

The only assumption regarding the missing voxels is that they are missing at random, which generally holds true in most sparse acquisitions. The models can therefore be learned from data of varying sparseness, including restoring data in all acquisition directions simultaneously.

Future Directions

The goal of these models is to facilitate downstream analysis of clinically acquired images. Given the predominance of such images in the medical setting, we plan to apply these models to a diverse set of populations to facilitate previously hindered analyses in new diseases.

The models can be improved in several ways. Probabilistic PCA has been shown to be sensitive to outliers [61], and outliers are likely to be present among medical scans. A future improvement of the model will therefore include robust PCA [58, 144]. Both models currently require significant computational resources, and better approximations in their learning procedures can help improve runtimes. Similarly, we do not impose any structure on the learning of overlapping subvolumes, but such structure could both

improve runtime and regularize learning across entire volumes.

Our method shares similarities with genotype imputation, which uses statistical inference for unobserved genotypes [13, 54, 72]. Many genetic studies include only a sparse sampling of subject genotypes, as described in Chapter 2. Imputation methods use patterns of sequence covariation in high-resolution genotype datasets such as HapMap [46] to infer denser genetic information for the subjects in a study. Many of these methods have biological underpinnings that justify the use of covariation for imputation, similar to our use of anatomical consistency across a population of subjects. While the models are significantly different in the two applications, future analyses should investigate the similar insights learned across the two domains. Also, joint imputation models could help impute genotypes related to anatomy, and hereditary anatomical regions in images.

Chapter 5

Visualization of Large-Scale Results

As imaging studies start routinely processing thousands of subjects, verification of the computational or analysis steps becomes difficult when evaluating subjects one by one. We introduce a novel approach for fast and effective visualization of large image collections in population studies. The key insight is to collapse inherently high-dimensional imaging data onto an interactive two-dimensional canvas native to a computer screen in a way that enables intuitive browsing of the image data. Increasingly, medical image computing research involves exploring large image sets with high intrinsic dimensionality. This includes three dimensions for each medical volume, and many *meta*-dimensions such as subject index, modality type in multimodal studies, time in longitudinal studies, or parameter choice in parameter sweep experiments. Current visualization tools generally display one or few 2D slices or 3D renderings at a time, and do not provide a natural way to explore the meta-dimensions.

We propose a novel visualization approach that enables rapid interactive visualization of high dimensional image data, bridging the gap between single-volume viewers and large dataset statistics. We aim to harness users' innate ability to identify visual patterns and deviations from those patterns. Specifically, our approach and the resulting visualization tool, called tipiX, bridge the gap between single-image visualization offered by most software and statistical population analysis. Our approach enables users to identify important patterns in the data or anomalies that might otherwise be overlooked. We demonstrate that our platform can be used for quick and effective evaluation and analysis, and we believe it will improve research workflow and facilitate novel method development. Our tool is freely available[1].

[1] http://tipix.csail.mit.edu. we also provide a video and live demonstrations.

5.1 Background

Research in medical image analysis involves increasingly large amounts of data, such as clinical cohorts like the ISGC [57, 102, 109, 140], offering thousands of subjects with multi-modal and longitudinal scans. Parameter sweeps for processing steps can also produce large collections of images, even for a small study. These datasets include meta-dimensions – dimensions other than the three spatial ones – such as time, subject index or parameter setting, such as the two studies we tackle in this thesis. This high dimensionality makes it challenging to identify patterns in large cohorts, to evaluate quality of processing steps, and to tune algorithm parameters.

Several powerful visualization methods and tools have been demonstrated for viewing one or few images at a time [40, 42, 50, 94]. Visualization tools are often bundled with state of the art processing and analysis software. Some packages offer a plethora of interactive visual modules that implement powerful image analysis algorithms [94]. Others offer graphics libraries that can interact with image analysis functions [120]. Finally, some tools are built to interact with complex systems for picture archiving and communication systems (PACS) or repository system [79, 108]. While powerful, all of these tools are built for visualization of few subjects at a time, and do not support large image collections.

A few recently demonstrated visualization packages enable visualization of several volumes at once [43, 94], but users are limited to only working with as many 2D images as will reasonably tile on a computer screen. Unfortunately, such grid displays are not feasible in larger datasets with hundreds of scans.

Processing of large image collections is done instead via complex pipelines [40, 105], necessitating fast evaluation of various steps. Automatically computed evaluation metrics are often used to summarize intermediate or final results, to identify outliers, and to evaluate trends [40, 76, 126]. For example, we describe such a pipeline and population statistics in the preprocessing of stroke data, Section 3.6. However, statistics are often limited, task-specific, and do not always capture the complexity inherent in individual tasks. For example, volume measurements of anatomical structures or pathology do not identify spatial patterns. More complex statistics can capture the spatial distribution of locations, but would not identify patterns in shape. In contrast, a researcher can often visually identify complex patterns given the appropriate visualization method.

Our approach improves on single-volume viewers by exploiting the ability of users to visually detect patterns and to identify problems across large collections. Specifically,

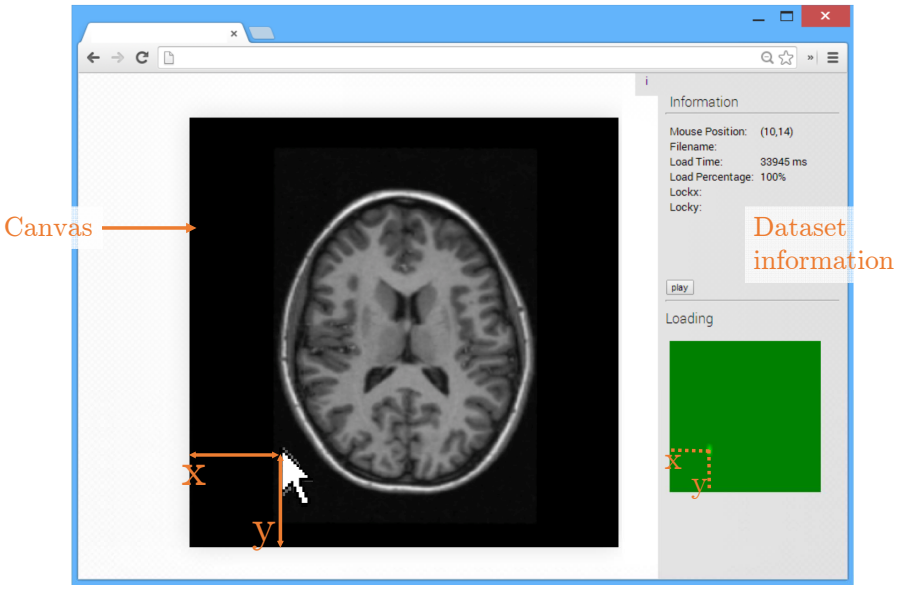

Figure 5.1. User interface that implements our approach. Elements in orange were added to this screenshot for illustration purposes. The position of the cursor, shown as x and y, controls which 2D image is displayed on the canvas. The load matrix on the right information bar (green) offers a visual indication of the position of the current image in the entire dataset.

the method provides rapid intuitive visualization of entire datasets by projecting two user-specified data dimensions onto a screen and providing convenient ways to interact with other dimensions. The tool is freely available, open-source, and does not require downloading or installation. It runs locally in modern web browsers while keeping all data and processing on the user's computer, therefore avoiding the requirement to upload sensitive information to a server. To the best of our knowledge, this is the first method for fast visualization of large image datasets.

■ 5.2 Design and Features

tipiX uses a two-dimensional interactive canvas to capture image data that is inherently high-dimensional. Given an image collection, tipiX displays a two-dimensional cross-section of the data. Through simple movement of the cursor, two more dimensions can be seamlessly explored, determining the two-dimensional image that is currently displayed on the screen. As a dataset may have many meta-dimensions, the user controls which dimensions are chosen for display and navigation. Figure 5.1 provides an overview

of the user interface.

Display Dimensions versus Navigation Dimensions

The main display – or *canvas* – shows two dimensions of the image set that users are familiar with, such as two spatial dimensions of an axial slice from a brain MRI volume. The user can simultaneously explore two more dimensions, which could be either physical dimensions or meta-dimensions. Examples include depth (the third spatial dimension), time (e.g., patient age or time of the scan), subject index in the collection, etc. In the remainder of this section, we use a dataset of 3D images with an extra dimension of subject index to illustrate the key features of the visualization method.

We use the position of the cursor on the drawing canvas to determine the location along the navigation dimensions. For example, moving the cursor to location $(4, 12)$ on the canvas displays the 4^{th} axial slice of the 12^{th} subject. This mode facilitates exploring two other dimensions, such as depth and time, together (if, for example, one of the navigation dimensions is set to subject age or time of the scan). The user can easily select and change display and navigation dimensions.

We avoid using sliders since they are limited to controlling one dimension at a time. This would result in much slower and more cumbersome simultaneous exploration of multiple dimensions in the datasets. In contrast, our method enables data visualization in a way that makes it easy to navigate in its entirety.

Implementation and Features

Our approach enables flexible selection of dimensions to be displayed. With a simple command the user can switch from viewing axial slices through volume depth and different subjects to viewing sagittal slices in the same data, as illustrated in Figure 5.2.

Within each dimension, tipiX provides fine control through keyboard shortcuts. For example, after identifying an outlier image among 100 subjects, a user might want to explore this subject along another dimension, such as a different modality. Our framework allows *locking* the subject (or generally, the current index for any dimension) with keyboard shortcuts to enable the user to explore the other dimension for that particular subject. Once this task is completed, the user can *unlock* the subject, and continue exploring the dataset. Our implementation includes several other useful features such as an information panel and a preview image to summarize the entire dataset.

We provide a freely available and open source implementation of tipiX. The im-

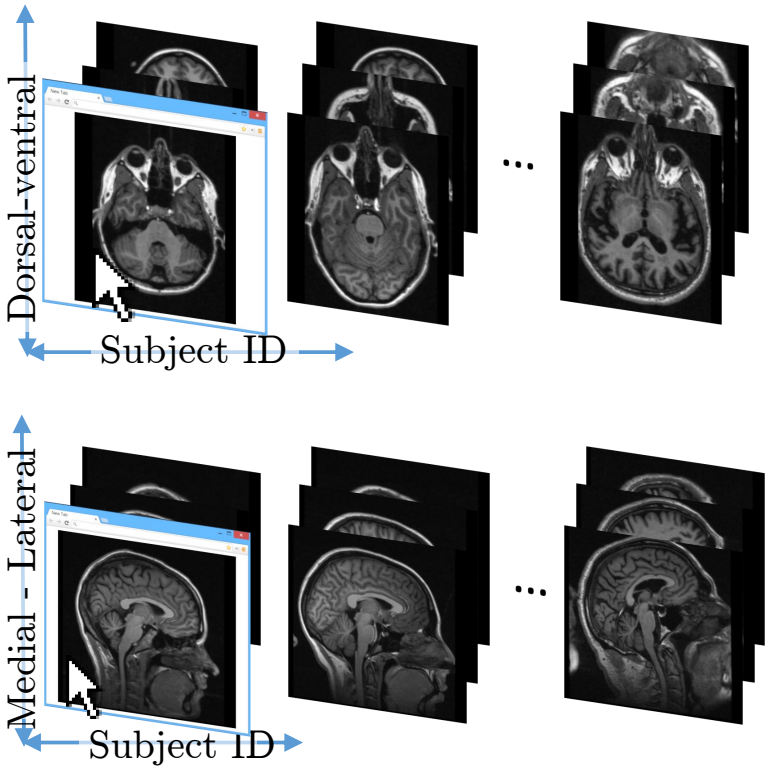

Figure 5.2. Example uses of tipiX on a dataset of 3D volumes with many subjects. Left: one axial slice of a subject is shown to the user at a time. The vertical position of the cursor controls axial slice location in the same subject, and the horizontal position of the cursor controls the subject index in the study. Right: exploring 2D sagittal views in the same dataset.

plementation runs in modern web browsers. This design decision avoids limitations associated with cross-platform functionality and software dependencies. All processing and visualization is performed client-side on the user's computer, thus avoiding security and privacy concerns. Sharing of visualization scenes, for example between technical researchers and clinicians, is simple and reduces to simply providing a unique URL. This functionality assumes both parties have access to the same data, for example if the data is accessible online or through a server on a private network.

The interactive canvas controls and data can also be embedded into another webpage, which is useful when developing tutorials, discussing medical datasets or previewing public data. tipiX employs the XIO library [51] to support input in popular medical

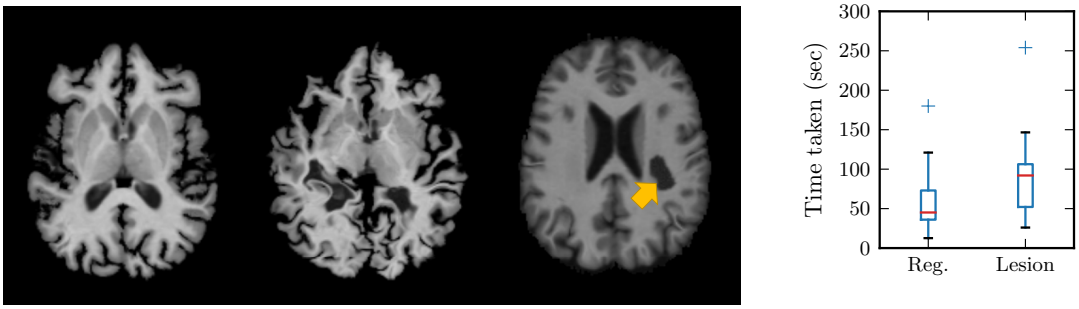

Figure 5.3. Outlier detection. Left panel: a typical brain scan after registration to an atlas and skullstripping, a poorly registered subject, and a simulated hypointense lesion on a subject (yellow arrow). Right plots: boxplots of time taken to identify two outliers in mis-registration and lesion detection tasks.

imaging formats, such as DICOM [3] and NIFTI [18], in addition to regular images, such as PNG and JPG. The data is rescaled to fit on the canvas, whose size adapts to each user's browser window.

■ 5.3 Evaluations

To the best of our knowledge, ours is the first visualization method capable of exploring large medical image collections visually. Our approach directly addresses the problem of visualizing high-dimensional data, while other viewers concentrate on specific volumes. We therefore avoid a direct comparison, as it would be unfair to the baseline methods. Instead, we demonstrate the utility of our platform through several user studies of typical medical image computing tasks. Each study included 9 users who were familiar with images, and who were provided with a brief, one-minute instruction on how to use tipiX.

Outlier Detection

As part of a typical medical image computing workflow, various processing steps must be evaluated for correctness. For example, following registration of a subject cohort, it is important to identify scans that did not register properly to an atlas template. Similarly, in a general cohort of patients, it is useful to quickly identify patients with a particular pathology.

In the first user study, we register 20 T1-weighted brain MRI scans that are part of the Freesurfer brain atlas [40] to a common atlas. We introduce a random perturbation

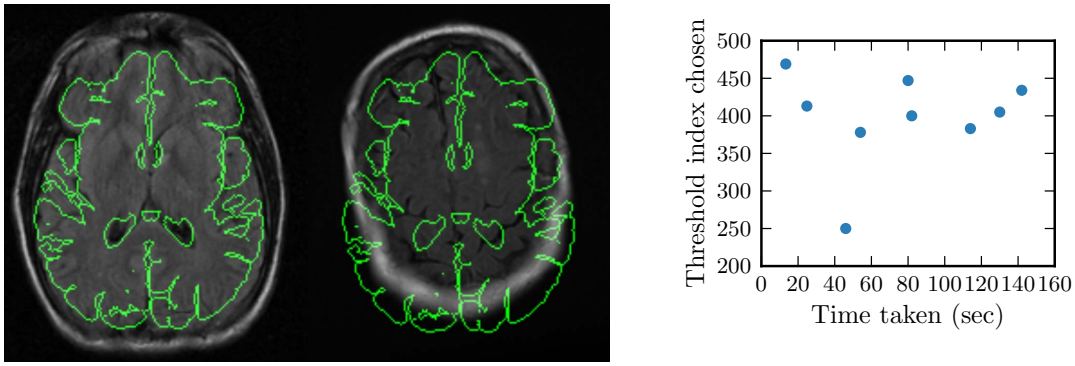

Figure 5.4. Detection of inclusion threshold. Left panel: a typical well-aligned subject image shown next to a typical poorly registered image, with the template brain and ventricle boundaries shown in green. Right plot: the average threshold versus the time taken by users.

in one registration to mimic misalignment. Users are given the entire dataset, and asked to identify a mis-registered subject in the cohort, where the mis-registration is only noticeable in some of the slices. In a similar fashion, we simulate a hypointense lesion in the white matter of one subject, and ask the users to identify the scan that contains the lesion. The lesion is only noticeable in about 11 of 100 brain slices.

These would be difficult tasks for single-volume visualization tools. Since the problems are only noticeable in a subset of the slices, re-factoring the dataset to a single volume of a specific slice across subjects is not feasible. Figure 5.3 illustrates example volume slices and reports the user time required for outlier detection.

We find that users are very adept at identifying outliers when using our visualization tool, taking on average about 45 seconds to identify a mis-registration, and about 90 seconds to identify a subject with a lesion. We observed empirically that users who first explored the dataset across volumes rather than within a single volume tended to complete the tasks faster. This illustrates the power of our approach of visualizing meta-dimensions and image dimensions simultaneously. Quick visual detection promises to substantially improve quality control in complex tasks, where designing robust quality control measures can be time consuming or prohibitive.

Pattern Identification

Similarly to identifying outliers following a pipeline step, we often aim to subdivide a subject cohort. For example, after rigid registration of a large image set to an atlas, we

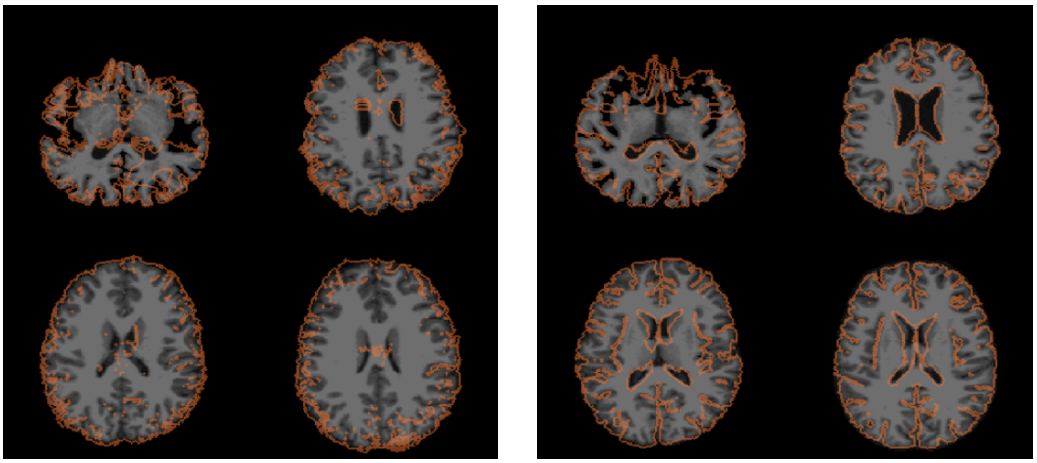

Figure 5.5. A screenshot of two images shown to users in the parameter sweep study. Axial slices from four separate subjects are shown with segmentation contours (orange) implied by the aligned atlas. Moving the cursor left to right sweeps the registration parameters; moving the cursor up and down goes through different axial slices. Left: results from the median parameter choice. Right: a parameter setting chosen by a user who noted that the top-left subject is otherwise mis-registered.

may wish to identify a subset of subjects that have promising registration results for analysis. In this study, we rank 500 non-rigidly registered subjects from the ISGC stroke cohort based on the sum of squared differences between each deformed subject image and the atlas template. Using this ranking, our method enables the users to explore the entire dataset and identify a threshold on the similarity measure that would separate well-registered subjects from those with significant misalignment. In Figure 5.4, we illustrate accurate good and poor registrations and report the agreement level among different users in the study. We found that nearly all users decided quickly on a rough threshold range, but some spent more time deciding on an optimal choice and remarked on the subjective nature of the task.

Parameter exploration is necessary in many applications, such as exploration of new datasets or development of novel methods. Registration quality metrics often do not capture complex patterns or task-specific criteria. Visualization of the behaviour of various parameters on a subset of the population can help to quickly identify the optimal value of the regularization parameter choice. We perform a parameter sweep for non-rigid registration using the Log Domain Diffeomorphic Demons algorithm [35, 136] on four subjects, and ask users to identify the optimal parameter (Figure 5.5). We found that most users chose very similar settings (smoothing kernel size near 1.5mm) in about

two minutes, and spent the majority of that time refining the optimal parameter once a reasonable range was found.

■ 5.4 Discussion

The rapid visualization method presented here, tipiX, is open sourced and in continued development as we receive feedback from the community on desired functionality. More features are possible, and we hope the tool will grow organically as different users contribute by suggesting or implementing functionality to facilitate their research. For example, seamlessly exploring a fifth dimension with the mouse wheel is a topic of future work.

Complex signal, such as change in the shape of a pathology, is hard to discover automatically. In Chapter 6 we describe a classification of stroke subjects based on their age and disease spread, and discuss the use of tipiX in exploring these classes.

TipiX is also used on non-medical datasets, for example with large collection of natural images used in computer vision. Exploration of these datasets facilitates analysis of natural variations in images, such as changes in a scene with the time of day [21]. Several features are implemented for such datasets, such as the video play of timelapses. With continued development, tipiX is likely to expand in usability from radiological scans, to natural images, and media applications.

Chapter 6

Discussion and Conclusions

THE methods and large-scale approach to image analysis introduced in this thesis enable many future clinical and technical directions of study. In this chapter, we suggest several future developments and explore initial clinical analyses using the stroke and ADNI cohorts.

6.1 Technical Directions

Throughout this thesis we sought to improve inference of anatomy and pathology by exploiting information outside the input image. We summarize discussions and expand on future directions.

6.1.1 Genetic and Clinical Factors

In Chapter 2, we demonstrate that subject-specific external factors, including their genotype and clinical information, can have a significant impact on the analysis of a subject's anatomy. We propose a versatile framework that enables heterogeneous external data to be used for inference of anatomical change. The model can be readily expanded, improving phenotypic functional dependencies on external data and *de novo* prediction, removing the need for the baseline observation. The method can be used to predict dramatically varied phenotypes, from disease growth in a medical scan to facial appearance in a photograph.

Predicting anatomical change is just one demonstration of the utility of genetic variants for medical image analysis. Image segmentation, image registration, and image population analysis can all use genetic factors to improve standard methods.

6.1.2 Collection of Subjects with Spatial Disease Burden

In Chapter 3, we demonstrate that a collection of patients can help identify complex spatial pathology in a new patient. The training enables separation of complex pathologies with similar presentation in brain MRI, leading to novel clinical analysis. Later in this chapter, we discuss prospective research directions, some of which are part of the ongoing clinical studies facilitated by the methods in this thesis.

We also discuss several image pre-processing methods in Chapter 5 to more robustly address standard analyses in clinical data. An important insight is that the sparse aspect of the data should be modelled directly or explicitly accounted for as much as possible to improve the image processing.

With this mindset, we are actively developing a method for non-rigid registration of brain images acquired in clinical settings using three-dimensional patches in a discrete registration framework [26]. The method explicitly models the sparse structure characteristic of the problem, leading to a robust patch-based discrete registration algorithm. While most methods use single voxels to asses data similarity, we design an appropriate 3D patch-based similarity function surrounding the desired voxel. We have developed a flexible, fast and open-source implementation of discrete deformable registration, including several patch-based data similarity functions.

6.1.3 Collections of Clinical Scans

In Chapter 4, we demonstrate that a collection of low-resolution scans captures fine-scale anatomy, leading to restoration of new clinical images. The models we propose can be expanded in several ways, with a focus on robust modifications to tackle the heterogeneity of clinical data. The model is general and can be applied to images of different resolutions, even within the same cohort. Prospective future applications can therefore explore automatic standard analysis that was previously infeasible due to the low quality of images.

6.1.4 Rapid Visualization of Image Collections

We present an approach that enables rapid interactive visualization of high dimensional image data. Our method facilitates assessment of computational steps and image analysis results in the large studies tackled in this thesis. As the method is implemented in the browser and is open-source, we hope the tool will grow organically as users implement functionality to facilitate research in medical image analysis, computer vision

and computational photography.

6.2 Clinical Analyses

We briefly illustrate how the methods developed in this thesis can be used to study spatial leukoaraiosis patterns from several large cohorts, and to predict such patterns using patient-specific genotype and clinical information. We deploy the pre-preprocessing steps for clinical images described in Section 3.6, followed by our delineation of different pathologies using Cerebro. We combined our pre-processing and analysis steps using the pipeBuilder system, an image analysis pipeline construction tool we developed to facilitate easy refinement of analysis of image analysis [125]. Altogether, we analyze over 3000 subject scans from 11 different stroke acquisition centers within the ISGS cohort, and nearly 5000 scans of 1000 subjects from the longitudinal ADNI dataset.

6.2.1 Multi Cohort Comparison

We consider the spatial leukoaraiosis distribution across different populations. Figure 6.1 illustrates how the population-wide leukoaraiosis averages are remarkably consistent across centers within the stroke study, despite differences in population sizes, population characteristics and imaging scanners. This suggests that subject cohorts can be combined for large scale analyses of white matter disease associations with clinical and genetic factors. In contrast, we observe more significant differences of the mean leukoaraiosis spread between the stroke and ADNI cohorts. Since the ADNI population has significantly different inclusion criteria, and does not include subjects who have recently had a stroke, it could be used as a control population in future studies. The differences in the leukoariosis maps are an ideal place to start future clinical comparisons. For example, are subjects in the stroke study associated with higher leukoaraiosis volume after normalizing for inclusion criteria across the two studies?

6.2.2 Volume Progression with Age

The evolution of leukoaraiosis burden with respect to clinical variables, such as age, is important for understanding cerebrovascular mechanisms related to stroke [109, 110]. Since leukoaraiosis volume varies dramatically across different patients, we cluster the patients into more homogeneous sub-populations, and then investigate the change with age separately in each sub-population. We use a two-component regression mixture model to capture variability in leukoaraiosis volume growth [64, 99]. Each component

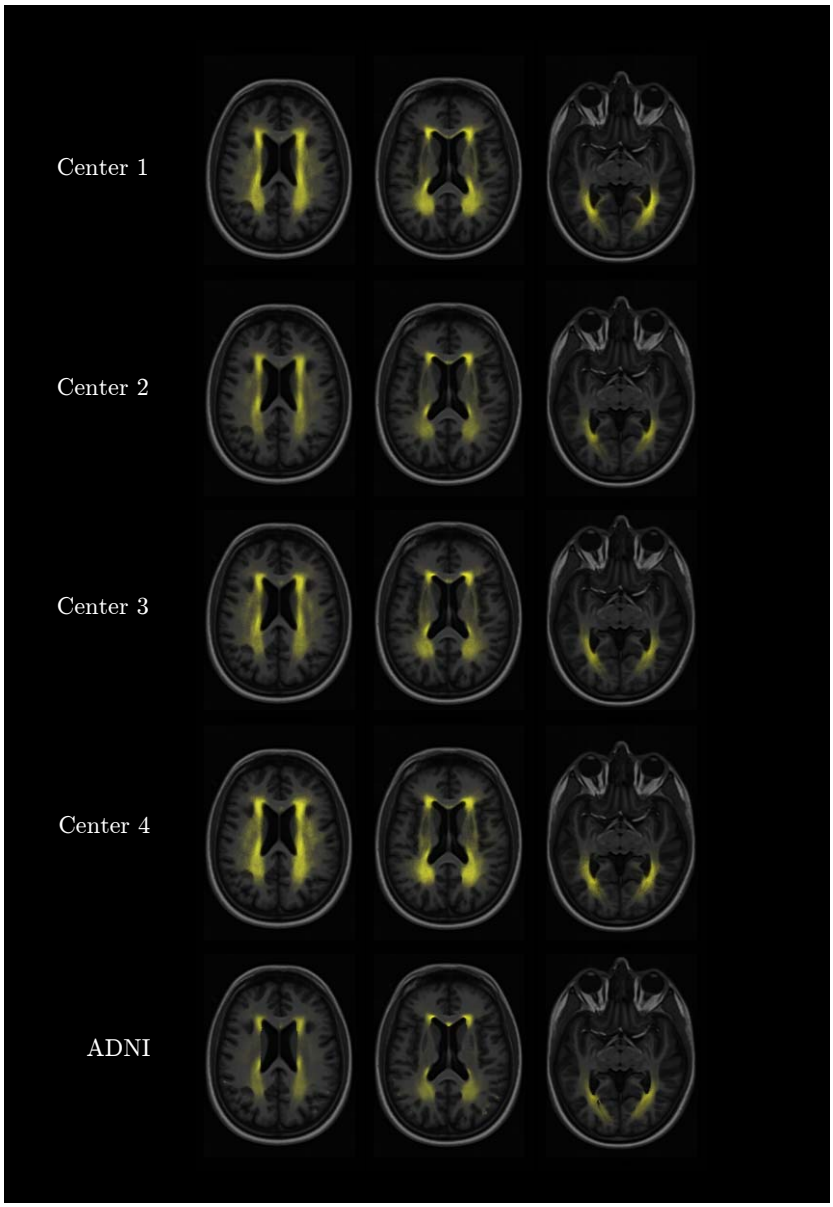

Figure 6.1. Average leukoaraiosis maps for four ISGC centers and ADNI. While the distribution patterns among the centers is consistent, the ADNI cohort, which does not include stroke cases, has a significantly weaker or more constrained spread.

is characterized by a different dependency on age.

Let v_i and z_i be the scalar leukoaraiosis volume and cluster assignment of patient i, respectively. We let X_i be a 2-dimensional feature vector capturing age and intercept,

and assume i.i.d. multinomial priors for cluster membership. Given 2-dimensional regression coefficient vectors $\boldsymbol{\beta}_c$ for each cluster c and fixed variance σ^2, we model the leukoaraiosis volume v_i in patient i as normally distributed with mean $X_i\boldsymbol{\beta}_c$ and fixed variance σ^2:

$$v_i = X_i\beta_c + \epsilon_i, \quad \text{where } \epsilon_i \sim \mathcal{N}(0, \sigma^2). \tag{6.1}$$

To estimate parameters $\boldsymbol{\beta}_c$ we use a hard-assignment iterative algorithm similar to k-means clustering, alternating until convergence between computing the cluster assignments and solving least-squares linear regression for cluster parameters:

$$z_i = \arg\min_c ||v_i - X_i\boldsymbol{\beta}_c||_2^2, \tag{6.2}$$

$$\boldsymbol{\beta}_c = (\boldsymbol{X}^T Z_c \boldsymbol{X})^{-1}\boldsymbol{X}^T Z_c \boldsymbol{v}, \tag{6.3}$$

where $\boldsymbol{v}$ is the vector of all volume values, $\boldsymbol{X}$ is the $N \times 2$ feature matrix, Z_c is a diagonal binary matrix, and $Z_c(i,i) = 1$ if $z_i = c$.

In the ISGC stroke study analysed throughout this thesis, this method identified a cluster of patients for whom age has little to no effect on leukoaraiosis volume ($\beta_1 = 2.27\text{mm}^3/\text{year}$), and another set of patients for whom it grows substantially with age ($\beta_2 = 8.84\text{mm}^3/\text{year}$). This clustering provides an initial subphenotyping of stroke subjects into *progressors* and *non-progressors*, providing a potential source for correlation analysis with external clinical variables.

■ 6.2.3 Distribution Progression with Age

While previous studies examined the overall leukoaraiosis *volume* of each patient using manual segmentations in a limited number of subjects, the method presented in Chapter 3 enables evaluation of leukoaraiosis *spatial distribution* evolution in the brain as a function of age in large-scale studies.

Within each cluster learned in the previous section, we use Nadaraya-Watson kernel regression [28, 88, 139] on the leukoaraiosis label maps to visualize representative images $I(t)$:

$$I(t) = \frac{\sum_{i=1}^{N} K_h(t - t_i) I_i}{\sum_{i=1}^{N} K_h(t - t_i)}, \tag{6.4}$$

where t is the age of interest, N is the number of patients, I_i is the leukoaraiosis label map of patient i warped into atlas space, and $K_h(\cdot)$ is a Gaussian kernel function with

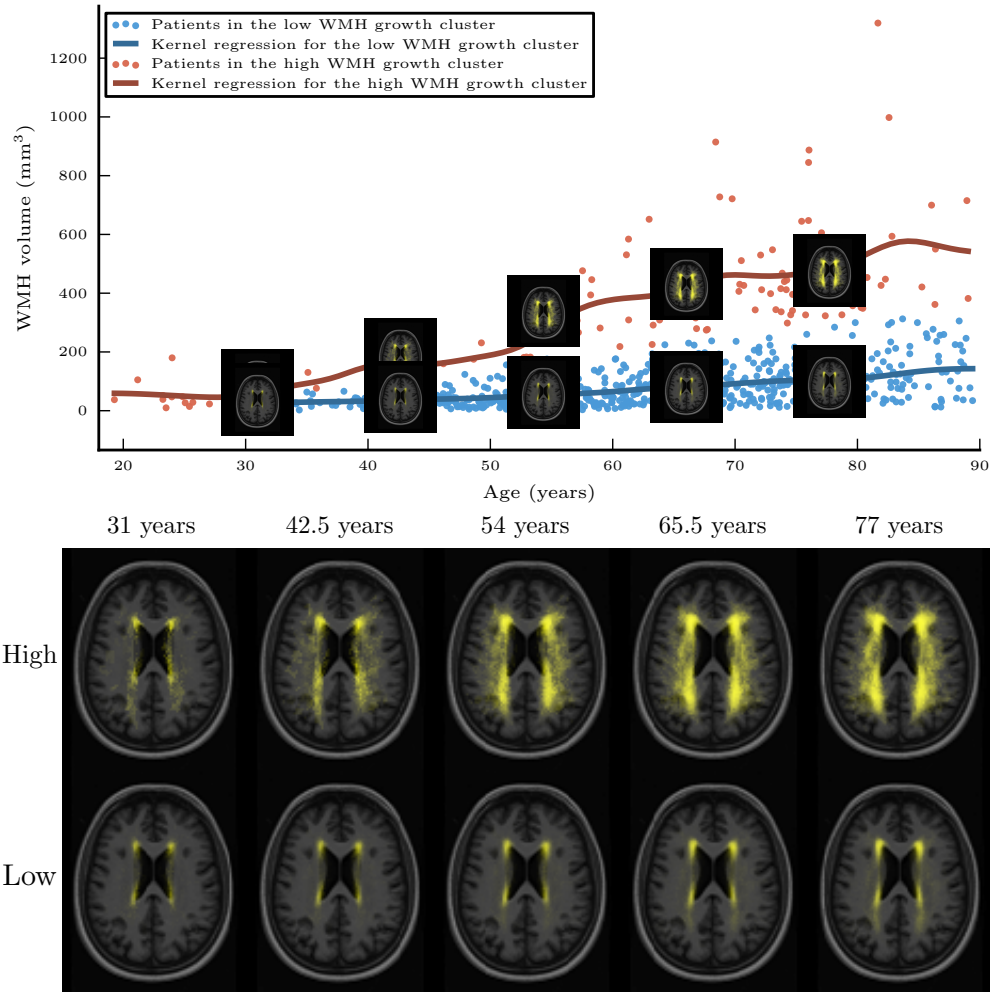

Figure 6.2. Top: the two-component regression mixture model clusters the patients into those with high leukoaraiosis growth as a function of age (red) and those with low leukoaraiosis growth as a function of age (blue). The lines show a kernel regression of leukoaraiosis volume as a function of age in each cluster. The representative images shown are obtained via kernel regression of the leukoaraiosis label maps as a function of age. Bottom: the two sets of representative images in more detail.

standard deviation h and mean 0. Intuitively, a representative leukoaraiosis image is a weighted average of all leukoaraiosis label maps, with patients close to age t contributing more to the average. Figure 6.2 visualizes the progression of the leukoaraiosis distributions with age, as well as representative images at different time points.

For the fast-growing leukoaraiosis burden cluster identified above, leukoaraiosis

tends to spread throughout the white matter, and most strongly in the posterior regions of the white matter. In the other, slow-growing leukoaraiosis burden cluster, the leukoaraiosis remains confined near the ventricles, as expected.

The visualization method presented in Chapter 5 can be used to exploring these two groups simultaneously, giving a concrete idea of the actual growth patterns of the vascular problems[1]. While difficult to quantify, this gives important clinical information in understanding cerebral ischemia.

To the best of our knowledge, this is the first analysis of the progression of spatial leukoaraiosis distribution with age. In the future, we will contrast clustering results across imaging centers and cohorts, and analyse how the spread of leukoaraiosis distribution varies across different clusters.

The disease progression estimated here is based on a cross-sectional cohort (each subject only has one scan). A promising area of future study is the comparison of this population-wide progression with actual disease progression in longitudinal studies, such as ADNI, where individual and population-wide trends can be evaluated.

■ 6.2.4 Prediction of Patient-Specific Growth

In Chapter 2 we introduce a method for predicting phenotypic change given patient-specific external information. In Chapter 3 we propose a method to delineate a cerebrovascular phenotype important in stroke, and describe a low-dimensional embedding of the disease spread. Together, these frameworks facilitate subject-specific prediction of leukoaraiosis growth over time, offering a rich characterization of disease trajectory. The longitudinal ADNI cohort presents an ideal control population, where such subject-specific predictions can be made and validated. Figure 6.3 shows a preliminary analysis, where we automatically delineated leukoaraiosis in 1000 ADNI scans to predict disease burden growth in follow-up scans for specific subjects. This analysis opens the door to spatial predictions in stroke patients, where a single, baseline scan is observed, and subsequent prediction can aid in treatment planning and preventing further cerebrovascular complications.

■ 6.2.5 Genetic Discovery

A central goal of the ISGC is to analyse the white matter disease burden and its genetic predisposition. The leukoaraiosis burden volumes we computed are currently undergoing Genome-Wide Association Study (GWAS) for genetic discovery in the context of

[1]The exploration of this result is explored in the tipiX video demonstration as well.

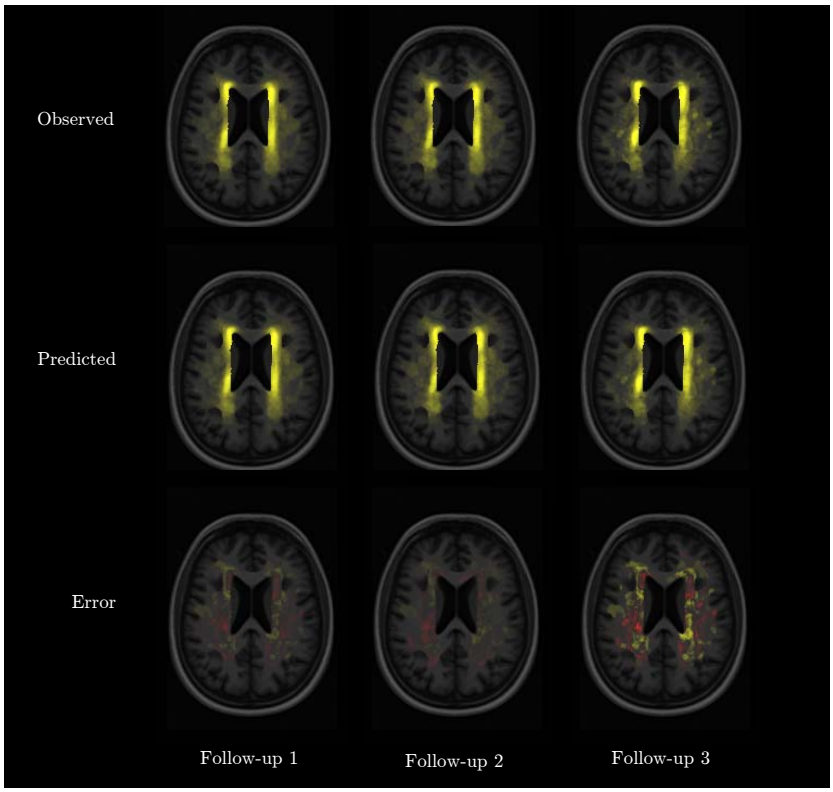

Figure 6.3. Preliminary prediction of leukoaraiosis distribution. For three follow-up scans of one subject, the observed, predicted and error maps of leukoaraisis distribution are projected onto the spatial statistical model of leukoaraiosis. The prediction error is multiplied by 10 for visualization. While the spatial change within one subject is minimal compared to population changes, the method can predict the slow growth over time.

stroke. The total scalar volume is a limited phenotypic characterization of the disease. The principal directions of the low-dimensional leukoaraiosis model in Chapter 3 likely characterize cerebrovascular mechanisms as they capture the spatial modes of variability of white matter disease. Since these measures provide a richer representation of the disease, we propose to analyse the coordinates of each subject in this space and look for correlations with genetic variants and clinical values. The regression-based volume clustering presented above provides an avenue for genetic examination, where we can determine if subjects in the *progressor* group have a genetic risk factor compared to *non-progressors*.

■ 6.3 Conclusions

In this thesis, we explore how variables external to a medical scan, such as genotypes, clinical factors, and large image collections can improve anatomy and images in the new patients. We show how genetic variants and clinical indicators can predict anatomical MRI images. We demonstrate that a cohort of patients with white matter pathology can help predict spatially complex disease burden in a new patient. We use large collections of clinical scans to dramatically improve the resolution of a new image and recover fine-scale anatomy. We present an approach that enables rapid interactive visualization of imaging data in large studies. By bringing these methods together in large scale analyses, we illustrate how our contributions facilitate new avenues of technical and clinical research.

Appendix A

Prediction of Anatomical Change: Parameter Updates

We present parameter updates for learning the generative model (2.5) and (2.3) using a linear mixed effects model (LMM) interpretation, and summarize the phenotype prediction (2.6) updates using the least squares kernel machine (LSKM) interpretation.

We seek the maximum likelihood estimates of parameters $\bar{\beta}$ and $\boldsymbol{\theta} = (\tau_G^2, \tau_C^2, \tau_I^2, \sigma^2)$. Specifically, letting

$$\boldsymbol{y}' = (\boldsymbol{y} - \boldsymbol{y}_0) \odot ([\frac{1}{\boldsymbol{x}_1} ... \frac{1}{\boldsymbol{x}_T}]^T) \tag{A.1}$$

$$\boldsymbol{V} = \sum_D \tau_D^2 \boldsymbol{K}_D + \sigma^2 \mathbb{I} \tag{A.2}$$

$$\boldsymbol{q} = ((\mathbb{1}^T \boldsymbol{V}^{-1} \mathbb{1})^{-1} \mathbb{1}^T \boldsymbol{V}^{-1})^T \tag{A.3}$$

we obtain intuitive iterative updates from standard linear mixed model literature [44, 75] for $\bar{\beta}$ using the best linear unbiased predictor

$$\bar{\beta} = \boldsymbol{q}^T \boldsymbol{y}', \tag{A.4}$$

and for the kernel parameters

$$\theta^{(k+1)} = \theta^{(k)} + \frac{1}{2} \left[\mathcal{I}^{(k)}\right]^{-1} \boldsymbol{R}^{(k)}, \tag{A.5}$$

where

$$\mathcal{I}_{i,j}^{k+1} = \frac{1}{2}\mathrm{tr}(\boldsymbol{P}\boldsymbol{K}_i\boldsymbol{P}\boldsymbol{K}_j) \tag{A.6}$$

$$\boldsymbol{R} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}\boldsymbol{K}_i) + \frac{1}{2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{K}_i\boldsymbol{P}y' \tag{A.7}$$

$$\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\mathbb{1}\boldsymbol{q}^T. \tag{A.8}$$

Essentially, (A.5) projects the residuals $\boldsymbol{R}$ at each step (t) onto the expected rate of change in likelihood.

To perform prediction, the LSKM estimation [44, 68, 75, 138] of the health terms $\boldsymbol{h} = \boldsymbol{\alpha}^T\boldsymbol{K}$ require the estimation of vectors $\boldsymbol{\alpha}_G, \boldsymbol{\alpha}_C$ and $\boldsymbol{\alpha}_I$ via a linear system of equations that involves our estimates of $\hat{\beta}$ and θ:

$$\begin{bmatrix} \mathbb{1}^T\mathbb{1} & \mathbb{1}^T\boldsymbol{K}_G & \mathbb{1}^T\boldsymbol{K}_C & \mathbb{1}^T\boldsymbol{K}_I \\ \boldsymbol{K}_G^T\mathbb{1} & \boldsymbol{K}_G^T\boldsymbol{K}_G + \frac{\tau_G^2}{\sigma^2}\boldsymbol{K}_G^T\mathbb{1} & \boldsymbol{K}_G^T\boldsymbol{K}_C & \boldsymbol{K}_G^T\boldsymbol{K}_I \\ \boldsymbol{K}_C^T\mathbb{1} & \boldsymbol{K}_G^T\mathbb{1} & \boldsymbol{K}_C^T\boldsymbol{K}_C + \frac{\tau_G^2}{\sigma^2}\boldsymbol{K}_C^T\mathbb{1} & \boldsymbol{K}_C^T\boldsymbol{K}_I \\ \boldsymbol{K}_I^T\mathbb{1} & \boldsymbol{K}_I^T\boldsymbol{K}_G & \boldsymbol{K}_I^T\mathbb{1} & \boldsymbol{K}_I^T\boldsymbol{K}_I + \frac{\tau_I^2}{\sigma^2}\boldsymbol{K}_I^T\mathbb{1} \end{bmatrix} \begin{bmatrix} \bar{\beta} \\ \boldsymbol{\alpha}_G \\ \boldsymbol{\alpha}_C \\ \boldsymbol{\alpha}_I \end{bmatrix} = \begin{bmatrix} \mathbb{1}^T\boldsymbol{y}' \\ \boldsymbol{K}_G^T\boldsymbol{y}' \\ \boldsymbol{K}_C^T\boldsymbol{y}' \\ \boldsymbol{K}_I^T\boldsymbol{y}' \end{bmatrix}. \tag{A.9}$$

which essentially translated from LMM parameters τ to LSKM parameters $\boldsymbol{\alpha}$. This facilitates phenotype prediction via:

$$\Delta\boldsymbol{y} = \Delta\boldsymbol{x}\bar{\beta} + \Delta\boldsymbol{x}\left(\boldsymbol{\alpha}_G^T\boldsymbol{K}_G + \boldsymbol{\alpha}_C^T\boldsymbol{K}_C + \boldsymbol{\alpha}_I^T\boldsymbol{K}_I\right). \tag{A.10}$$

Note that we define kernels on baseline data. Standard LMM solutions would be computationally expensive, or even prohibitive, for thousands of observations, as the entire genetic and the image phenotype data is prohibitively large.

Appendix B

Mixture of Principal Components with Missing Data

In this appendix we derive two models for mixture of principal components with missing data introduced in Chapter 4. We first start with the significantly simpler problem of a single Gaussian with missing data [31, 74], and incrementally add complexity to the model until we describe our full models [31, 61, 74, 130, 131].

B.1 Preamble

We assume $\boldsymbol{y}$ is drawn from a multivariate normal of dimension d:

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{B.1}$$

where the multivariate normal $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is parametrized by mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$. In our application, $\boldsymbol{y}$ is a three-dimensional patch whose voxels have been vectorized. Additionally, we assume some entries of vector $\boldsymbol{y}$ are missing, and denote them as $\boldsymbol{y}^{\mathcal{M}}$, and let $\boldsymbol{y}_{\mathcal{O}}$ be the observed entries. Similarly, the notation $\boldsymbol{\Sigma}^{\mathcal{M},\mathcal{O}}$ selects the missing rows and observed columns of the covariance $\boldsymbol{\Sigma}$.

We list the conditional probability of missing data in a multivariable Gaussian sample [34, 55], a result which will be used throughout this derivation:

$$p(\boldsymbol{y}^{\mathcal{M}}|\boldsymbol{y}^{\mathcal{O}}) = \mathcal{N}(\boldsymbol{\mu}^{\mathcal{M}|\mathcal{O}}, \boldsymbol{\Sigma}^{\mathcal{M}|\mathcal{O}}) \tag{B.2}$$

where

$$\boldsymbol{\mu}^{\mathcal{M}|\mathcal{O}} = \boldsymbol{\mu}^{\mathcal{M}} + \boldsymbol{\Sigma}^{\mathcal{M}\mathcal{O}}(\boldsymbol{\Sigma}^{\mathcal{O}\mathcal{O}})^{-1}(\boldsymbol{y}^{\mathcal{O}} - \boldsymbol{\mu}^{\mathcal{O}}) \tag{B.3}$$

$$\boldsymbol{\Sigma}^{\mathcal{M}|\mathcal{O}} = \boldsymbol{\Sigma}^{\mathcal{M}\mathcal{M}} - (\boldsymbol{\Sigma}^{\mathcal{O}\mathcal{M}})^{T}(\boldsymbol{\Sigma}^{\mathcal{O}\mathcal{O}})^{-1}\boldsymbol{\Sigma}^{\mathcal{O}\mathcal{M}} \tag{B.4}$$

The expectations of the unknown data is then:

$$\mathbf{E}[y_j|\boldsymbol{y}^{\mathcal{O}};\boldsymbol{\theta}] = \mu_j + \boldsymbol{\Sigma}_j^{\mathcal{O}}(\boldsymbol{\Sigma}^{\mathcal{O},\mathcal{O}})^{-1}(\boldsymbol{y}^{\mathcal{O}} - \boldsymbol{\mu}^{\mathcal{O}}) \tag{B.5}$$

and

$$\mathbf{E}[y_j y_k|\boldsymbol{y}_i^{\mathcal{O}};\boldsymbol{\theta}] = \mathbf{E}[y_j|\boldsymbol{y}^{\mathcal{O}};\boldsymbol{\theta}]\mathbf{E}[y_k|\boldsymbol{y}^{\mathcal{O}};\boldsymbol{\theta}] + \boldsymbol{\Sigma}_{jk} - (\boldsymbol{\Sigma}^{\mathcal{O}j})^T(\boldsymbol{\Sigma}^{\mathcal{O}\mathcal{O}})^{-1}\boldsymbol{\Sigma}^{\mathcal{O}k} \tag{B.6}$$

where both y_j and y_k are missing.

■ B.2 Single Multivariate Normal

■ B.2.1 Multivariate Normal with Missing Data

We approximate the maximum likelihood estimates of a multivariate normal given samples with missing data [31, 74]. Following extensive literature on this topic, we use this model as a first step in learning a mixture of principal components of rotated 3-D patches with missing data. We assume $\boldsymbol{y}_i$, for $i \in 1..N$, are drawn from a multivariate normal of dimension d:

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{B.7}$$

We assume some entries of each vector $\boldsymbol{y}_i$ are missing, and denote them as $\boldsymbol{y}_i^{\mathcal{M}_i}$, and let $\boldsymbol{y}_i^{\mathcal{O}_i}$ be the observed entries. We model the missing entries as latent random variables, and let $\mathcal{Y} = \{\boldsymbol{y}_i\}_{i=1}^N$, $\mathcal{Y}^{\mathcal{O}} = \{\boldsymbol{y}_{i,\mathcal{O}_i}\}_{i=1}^N$, and $\boldsymbol{\theta} = \{\boldsymbol{u}, \boldsymbol{\Sigma}\}$. The observed data likelihood is:

$$\begin{aligned}\ell(\boldsymbol{\theta}|\mathcal{Y}^{\mathcal{O}}) &= p(\mathcal{Y}^{\mathcal{O}};\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &= \prod_i^N \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i};\boldsymbol{\mu}^{\mathcal{O}_i},\boldsymbol{\Sigma}^{\mathcal{O}_i\mathcal{O}_i}),\end{aligned} \tag{B.8}$$

and the complete data likelihood is:

$$\begin{aligned}\ell(\boldsymbol{\theta}|\mathcal{Y}) &= p(\mathcal{Y};\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &= \prod_i^N \mathcal{N}(\boldsymbol{y}_i;\boldsymbol{\mu},\boldsymbol{\Sigma})\end{aligned} \tag{B.9}$$

from (B.8) and (B.9), the posterior of the missing data is:

$$\begin{aligned} p(\mathcal{Y}^{\mathcal{M}}|\mathcal{Y}^{\mathcal{O}_i}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= p(\mathcal{Y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) / p(\mathcal{Y}^{\mathcal{O}_i}; \boldsymbol{\mu}^{\mathcal{O}_i}, \boldsymbol{\Sigma}^{\mathcal{O}_i\mathcal{O}_i}) \\ &= \prod_i p(\boldsymbol{y}_i^{\mathcal{M}_i}|\boldsymbol{y}_i^{\mathcal{O}_i}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{aligned} \tag{B.10}$$

To estimate the observed likelihood in the presence of missing data, we use the Expectation Conditional Maximization (ECM) variant of the Generalized Expectation Maximization algorithm. The expectation of the log likelihood is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbf{E}_{\mathcal{Y}^{\mathcal{M}}|\mathcal{Y}^{\mathcal{O}_i}, \boldsymbol{\theta}^{(t)}} \left[\log p(\mathcal{Y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] \tag{B.11}$$

$$= -\frac{dN}{2}\log 2\pi - \frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_i^N \mathbf{E}_{\mathcal{Y}^{\mathcal{M}}|\mathcal{Y}^{\mathcal{O}_i}, \boldsymbol{\theta}^{(t)}} \left[(\boldsymbol{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})\right]. \tag{B.12}$$

For clarity, we drop the subscript $\mathcal{Y}^{\mathcal{M}}|\mathcal{Y}^{\mathcal{O}_i}, \boldsymbol{\theta}^{(t)}$ for the rest of this section since all expectations are taken with respect to this conditional distribution. Therefore, the **expectation step** involves the updates of the missing data:

$$\begin{aligned} \mathbf{E}\left[y_{ij}\right] &= y_{ij}^{(t)} \\ y_{ij}^{(t)} &= \begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed} \\ \mu_j^{(t)} + (\boldsymbol{\Sigma}^{j\mathcal{O}_i})^{(t)}(\boldsymbol{\Sigma}^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\boldsymbol{y}_i^{\mathcal{O}_i} - (\boldsymbol{\mu}^{\mathcal{O}_i})^{(t)}) & \text{otherwise} \end{cases} \end{aligned} \tag{B.13}$$

where we explicitly write $y_{ij}^{(t)}$ at iteration t as it uses parameter estimates from time t. Similarly, the missing data covariance terms:

$$\begin{aligned} \mathbf{E}\left[y_{ij}y_{ik}\right] &= \mathbf{E}[y_{ij}]\mathbf{E}[y_{ik}] + s_{i,jk}^{(t)} \\ &= y_{ij}^{(t)} y_{ik}^{(t)} + s_{i,jk}^{(t)} \\ s_{ijk}^{(t)} &= \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ is observed} \\ \boldsymbol{\Sigma}_{j,k}^{(t)} - (\boldsymbol{\Sigma}^{\mathcal{O}_i j})^{(t),T}(\boldsymbol{\Sigma}^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\boldsymbol{\Sigma}^{\mathcal{O}_i k})^{(t)} & \text{otherwise} \end{cases} \end{aligned} \tag{B.14}$$

using (B.5) and (B.6).

For the **maximization step**, we optimize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$:

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \tag{B.15}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\mu}} &= -\frac{1}{2}\sum_{i}^{N}\frac{\partial}{\partial \boldsymbol{\mu}}\mathbf{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu})\right] \\
&= -\sum_{i}\Sigma^{-1}(\boldsymbol{y}_i^{(t)}-\mu) = 0 \\
\sum_i \boldsymbol{\mu}^{(t-1)} &= \sum_i \boldsymbol{y}_i^{(t)} \\
\boldsymbol{\mu}^{(t+1)} &= \frac{1}{N}\sum_i \boldsymbol{y}_i^{(t)}
\end{aligned} \tag{B.16}$$

using [92] equation (85), and

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\Sigma}^{-1}} &= \frac{N}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}}log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\sum_{i}^{N}\frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}}\mathbf{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu})\right] \\
&= \frac{N}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_i \mathbf{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu})(\boldsymbol{y}_i-\boldsymbol{\mu})^T\right] = 0 \\
\boldsymbol{\Sigma}^{(t+1)} &= \frac{1}{N}\sum_i \mathbf{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu})(\boldsymbol{y}_i-\boldsymbol{\mu})^T\right] \\
&= \frac{1}{N}\sum_i \mathbf{E}\left[\boldsymbol{y}_i\boldsymbol{y}_i^T - \boldsymbol{y}_i\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{y}_i^T - \boldsymbol{\mu}\boldsymbol{\mu}^T\right] \\
&= \frac{1}{N}\sum_i \boldsymbol{y}_i^{(t)}\boldsymbol{y}_i^{(t),T} + \boldsymbol{S}_i^{(t)} - \boldsymbol{y}_i^{(t)}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{y}_i^{(t),T} - \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= \frac{1}{N}\sum_i (\boldsymbol{y}_i^{(t)}-\boldsymbol{\mu})(\boldsymbol{y}_i^{(t)}-\boldsymbol{\mu})^T + \boldsymbol{S}_i^{(t)} \\
\boldsymbol{\Sigma}^{(t+1)} &= \frac{1}{N}\sum_i (\boldsymbol{y}_i^{(t)}-\boldsymbol{\mu}^{(t+1)})(\boldsymbol{y}_i^{(t)}-\boldsymbol{\mu}^{(t+1)})^T + \boldsymbol{S}_i^{(t)}
\end{aligned} \tag{B.17}$$

where $\left[\boldsymbol{S}_i^{(t)}\right]_{jk} = s_{ijk}^{(t)}$, via (B.5), (B.6), and [92] equation (57). Both updates have intuitive views explained in section 4.2.

In summary, the expectation step imputes data statistics, estimating $\boldsymbol{y}_i^{(t)}$ and $\boldsymbol{S}_i^{(t)}$ via (B.13) and (B.14), respectively, and the maximization step estimates model parameters $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$ via (B.16) and (B.17), respectively.

■ B.2.2 Multivariate Normal with Transformed Patches with Missing Voxels

As described in section 4.2, we affinely transform patches from the original subject space, where the patches have either missing or observed voxels, to subject space where voxels are interpolated. We let $\boldsymbol{\Gamma}_i$ and $\boldsymbol{R}_i$ be the interpolation matrices that transform

the subject space vectorized patch $\tilde{\boldsymbol{y}}_i$ to atlas space $\boldsymbol{y}_i$, and vice-versa:

$$\boldsymbol{y}_i = \boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i \tag{B.18}$$

$$\tilde{\boldsymbol{y}}_i = \boldsymbol{R}_i \boldsymbol{y}_i \tag{B.19}$$

Since $\boldsymbol{R}_i \Gamma_i = \boldsymbol{I}$, we treat $\boldsymbol{R}_i$ as a generalized inverse of $\boldsymbol{\Gamma}_i$, and for simplicity we further assume R_i is square [1]. We assume variables $\boldsymbol{y}_i$, for $i \in 1..N$, are drawn from a multivariate normal of dimension d in the atlas space:

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{B.20}$$

and therefore:

$$\tilde{\boldsymbol{y}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i), \tag{B.21}$$

where

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{R}_i \boldsymbol{\mu} \tag{B.22}$$

$$\tilde{\boldsymbol{\Sigma}}_i = \boldsymbol{R}_i \boldsymbol{\Sigma} \boldsymbol{R}_i^T. \tag{B.23}$$

We assume some entries of each vector $\boldsymbol{y}_i$ are missing, as before, and model the missing entries as latent random variables. We let $\tilde{\mathcal{Y}} = \{\tilde{\boldsymbol{y}}_i\}_{i=1}^N$, $\tilde{\mathcal{Y}}^{\mathcal{O}} = \{\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i}\}_{i=1}^N$, $\mathcal{R} = \{\boldsymbol{R}_i\}_{i=1}^N$, and $\boldsymbol{\theta} = \{\mathcal{R}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. The observed data likelihood is:

$$\begin{aligned} \ell(\boldsymbol{\theta}|\tilde{\mathcal{Y}}^{\mathcal{O}_i}) &= p(\tilde{\mathcal{Y}}^{\mathcal{O}_i}; \mathcal{R}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \prod_i^N \mathcal{N}(\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i}; \tilde{\boldsymbol{\mu}}_i^{\mathcal{O}_i}, \tilde{\boldsymbol{\Sigma}}_i^{\mathcal{O}_i \mathcal{O}_i}) \end{aligned} \tag{B.24}$$

The complete data likelihood is:

$$\begin{aligned} \ell(\boldsymbol{\theta}|\tilde{\mathcal{Y}}) &= p(\tilde{\mathcal{Y}}; \mathcal{R}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \prod_i^N \mathcal{N}(\tilde{\boldsymbol{y}}_i; \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i) \end{aligned} \tag{B.25}$$

[1] Depending on the rigid transformation model used, $\boldsymbol{R}_i$ might not be square, but will usually be close. Assuming R_i as square simplifies the derivations significantly without significantly affecting the final updates.

from (B.24) and (B.25), the posterior of the missing data is:

$$p(\tilde{\mathcal{Y}}^{\mathcal{M}}|\tilde{\mathcal{Y}}^{\mathcal{O}_i},\mathcal{R},\boldsymbol{\mu},\boldsymbol{\Sigma}) = p(\tilde{\mathcal{Y}};\mathcal{R},\boldsymbol{\mu},\boldsymbol{\Sigma})/p(\tilde{\mathcal{Y}}^{\mathcal{O}_i};\mathcal{R},\boldsymbol{\mu}^{\mathcal{O}_i},\boldsymbol{\Sigma}^{\mathcal{O}_i\mathcal{O}_i})$$
$$= \prod_i^N p(\tilde{\boldsymbol{y}}_i^{\mathcal{M}_i}|\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i};\tilde{\boldsymbol{\mu}}_i,\tilde{\boldsymbol{\Sigma}}_i) \tag{B.26}$$

Similarly to the non-transformed case, to estimate the optimal parameters we use the ECM algorithm:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbf{E}_{\tilde{\mathcal{Y}}^{\mathcal{M}}|\tilde{\mathcal{Y}}^{\mathcal{O}_i},\boldsymbol{\theta}^{(t)}}\left[\log p(\tilde{\mathcal{Y}};\mathcal{R},\boldsymbol{\mu},\boldsymbol{\Sigma})\right] \tag{B.27}$$
$$= -\frac{dN}{2}\log 2\pi - \frac{1}{2}\sum_i \log|\boldsymbol{R}_i\boldsymbol{\Sigma}\boldsymbol{R}_i^T| \tag{B.28}$$
$$-\frac{1}{2}\sum_i^N \mathbf{E}_{\tilde{\mathcal{Y}}^{\mathcal{M}}|\tilde{\mathcal{Y}}^{\mathcal{O}_i},\boldsymbol{\theta}^{(t)}}\left[(\tilde{\boldsymbol{y}}_i-\boldsymbol{R}_i\boldsymbol{\mu})^T(\boldsymbol{R}_i\boldsymbol{\Sigma}\boldsymbol{R}_i^T)^{-1}(\tilde{\boldsymbol{y}}_i-\boldsymbol{R}_i\boldsymbol{\mu})\right] \tag{B.29}$$
$$= -\frac{dN}{2}\log 2\pi - \frac{1}{2}\sum_i \log|\boldsymbol{R}_i\boldsymbol{R}_i^T| - \frac{N}{2}\log(|\boldsymbol{\Sigma}|) \tag{B.30}$$
$$-\frac{1}{2}\sum_i^N \mathbf{E}_{\tilde{\mathcal{Y}}^{\mathcal{M}}|\tilde{\mathcal{Y}}^{\mathcal{O}_i},\boldsymbol{\theta}^{(t)}}\left[(\tilde{\boldsymbol{y}}_i-\boldsymbol{\mu}_i)^T\tilde{\boldsymbol{\Sigma}}_i^{-1}(\tilde{\boldsymbol{y}}_i-\boldsymbol{\mu}_i)\right] \tag{B.31}$$

As before, for clarity, we drop the subscript $\mathcal{Y}^{\mathcal{M}}|\mathcal{Y}^{\mathcal{O}_i},\boldsymbol{\theta}^{(t)}$ for the rest of this section. The **expectation step** involves the updates of the missing data:

$$\mathbf{E}\left[\tilde{y}_{ij}\right] = \tilde{y}_{ij}^{(t)}$$
$$\tilde{y}_{ij}^{(t)} = \begin{cases} \tilde{y}_{ij} & \text{if } \tilde{y}_{ij} \text{ is observed} \\ \tilde{\mu}_{ij}^{(t)} + (\tilde{\boldsymbol{\Sigma}}_i^{j\mathcal{O}_i})^{(t)}(\tilde{\boldsymbol{\Sigma}}_i^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i} - (\tilde{\boldsymbol{\mu}}_i^{\mathcal{O}_i})^{(t)}) & \text{otherwise} \end{cases} \tag{B.32}$$

and the missing data covariance terms:

$$\mathbf{E}\left[\tilde{y}_{ij}\tilde{y}_{ik}\right] = \mathbf{E}[\tilde{y}_{ij}]\mathbf{E}[\tilde{y}_{ik}] + s_{ijk}^{(t)}$$
$$= \tilde{y}_{ij}^{(t)}\tilde{y}_{ik}^{(t)} + s_{ijk}^{(t)}$$
$$s_{ijk}^{(t)} = \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ is observed} \\ \tilde{\boldsymbol{\Sigma}}_{i,jk}^{(t)} - (\tilde{\boldsymbol{\Sigma}}_i^{\mathcal{O}_i j})^{(t),T}(\tilde{\boldsymbol{\Sigma}}_i^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\tilde{\boldsymbol{\Sigma}}_i^{\mathcal{O}_i k})^{(t)} & \text{otherwise} \end{cases} \tag{B.33}$$

using (B.5) and (B.6).

For the **maximization step**, we use the equivalence

$$\mathbf{E}\left[(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{\mu}}_i)^T \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{\mu}}_i)\right] = \mathbf{E}\left[(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})\right] \tag{B.34}$$

and optimize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ for $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$:

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\mu}} &= -\frac{1}{2} \sum_i^N \frac{\partial}{\partial \boldsymbol{\mu}} \mathbf{E}\left[(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})\right] \\
&= -\sum_i \Sigma^{-1} (\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} - \mu) = 0 \\
\boldsymbol{\mu}^{(t+1)} &= \frac{1}{N} \sum_i \boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)}
\end{aligned} \tag{B.35}$$

using [92] equation (85), and

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\Sigma}^{-1}} &= \frac{N}{2} \frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}} log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2} \sum_i^N \frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}} \mathbf{E}\left[(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})\right] \\
&= \frac{N}{2} \boldsymbol{\Sigma} - \frac{1}{2} \sum_i \mathbf{E}\left[(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})^T\right] = 0 \\
\boldsymbol{\Sigma}^{(t+1)} &= \frac{1}{N} \sum_i \mathbf{E}\left[(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i - \boldsymbol{\mu})^T\right] \\
&= \frac{1}{N} \sum_i \boldsymbol{\Gamma}_i (\tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{R}_i \boldsymbol{\mu})(\tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{R}_i \boldsymbol{\mu})^T \boldsymbol{\Gamma}_i^T + \boldsymbol{\Gamma}_i \boldsymbol{S}_i^{(t)} \boldsymbol{\Gamma}_i^T \\
\boldsymbol{\Sigma}^{(t+1)} &= \frac{1}{N} \sum_i (\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})^T + \boldsymbol{\Gamma}_i \boldsymbol{S}_i^{(t)} \boldsymbol{\Gamma}_i^T
\end{aligned} \tag{B.36}$$

where $\left[\boldsymbol{S}_i^{(t)}\right]_{jk} = s_{ijk}^{(t)}$.

In summary, the expectation step imputes data statistics, estimating $\boldsymbol{y}_i^{(t)}$ and $\boldsymbol{S}_i^{(t)}$ via (B.32) and (B.33), respectively, and the maximization step estimates model parameters $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$ via (B.35) and (B.36), respectively. The expectation step computations happen in the original subject space, while the parameter updates are performed in atlas space.

■ B.2.3 PCA with Missing Voxels

As discussed in section 4.2, there are several options for modeling principal components in the context of missing data. We start with the probabilistic PCA formulation [131]

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{\mu} + \boldsymbol{\epsilon} \tag{B.37}$$

where $\boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{I})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$, $\boldsymbol{W}$ is a $d \times p$ matrix, $\boldsymbol{x}$ is a $p \times 1$ vector, where $p << d$ is the dimension of the mbedding space. The likelihood is

$$\begin{aligned} p(\mathcal{Y}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) &= \prod_i \int_{\boldsymbol{x}_i} p(\boldsymbol{y}_i|\boldsymbol{x}_i; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) p(\boldsymbol{x}) d\boldsymbol{x}_I \\ &= \prod_i \int_{\boldsymbol{x}_i} \mathcal{N}(\boldsymbol{y}_i, \boldsymbol{W}\boldsymbol{x}_i + \boldsymbol{\mu}; \sigma^2\boldsymbol{I}) \mathcal{N}(\boldsymbol{x}; 0, \boldsymbol{I}) d\boldsymbol{x}_i \\ &\propto \prod_i \mathcal{N}(\boldsymbol{y}_i, \boldsymbol{\mu}; \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}) \int_{\boldsymbol{x}_i} \mathcal{N}(\boldsymbol{x}; \boldsymbol{m}_c, \boldsymbol{C}_c) d\boldsymbol{x}_i \\ &= \prod_i \mathcal{N}(\boldsymbol{y}_i, \boldsymbol{\mu}; \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}) \end{aligned} \tag{B.38}$$

where

$$\begin{aligned} \boldsymbol{m}_c &= \boldsymbol{C}_c \frac{1}{\sigma^2} \boldsymbol{W}^T (\boldsymbol{y}_i - \boldsymbol{\mu}) \\ \boldsymbol{C}_c &= \sigma^2 (\boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I})^{-1}. \end{aligned} \tag{B.39}$$

Probabilistic PCA models $\boldsymbol{x}$ as a random variable and uses the EM algorithm to learn the principal axes $\boldsymbol{W}$. In the presence of missing data $\boldsymbol{y}^{\mathcal{M}}$, either $\boldsymbol{y}^{\mathcal{M}}$ or $\boldsymbol{x}$ could be modelled as latent variables [61, 131]. In the next two sections we model each of these quantities as a hidden latent variable, and discuss the effects of both models in the main text.

■ B.2.4 ML-PCA for Missing Data

In this section, we model missing voxels $y_i^{\mathcal{M}_i}$ as latent variables, and adopt the likelihood (B.38). Note that the low-dimensional patches $\boldsymbol{x}$ are not modelled. We let $\boldsymbol{C} = \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}$. Maximum likelihood estimates can again be achieved via EM, and the E-

step updates are identical to the ones in the simple multivariate case (B.13) and (B.14):

$$\begin{aligned}
\mathbf{E}\left[y_{i,j}\right] &= y_{ij}^{(t)} \\
y_{ij}^{(t)} &= \begin{cases} y_{i,j} & \text{if } y_{ij} \text{ is observed} \\ \mu_j^{(t)} + (\boldsymbol{C}^{j,\mathcal{O}_i})^{(t)}(\boldsymbol{C}^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\boldsymbol{y}_i^{\mathcal{O}_i} - (\boldsymbol{\mu}^{\mathcal{O}_i})^{(t)}) & \text{otherwise} \end{cases}
\end{aligned} \tag{B.40}$$

$$\begin{aligned}
\mathbf{E}\left[y_{ij}y_{ik}\right] &= \mathbf{E}[y_{ij}]\mathbf{E}[y_{ik}] + s_{ijk}^{(t)} \\
&= y_{ij}^{(t)}y_{ik}^{(t)} + s_{ijk}^{(t)} \\
s_{ijk}^{(t)} &= \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ is observed} \\ \boldsymbol{C}_{jk}^{(t)} - (\boldsymbol{C}^{\mathcal{O}_i,j})^{(t),T}(\boldsymbol{C}^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\boldsymbol{C}^{\mathcal{O}_i k})^{(t)} & \text{otherwise} \end{cases}
\end{aligned} \tag{B.41}$$

Similarly, in the maximization step, we get the same update for $\boldsymbol{\mu}$:

$$\boldsymbol{\mu}^{(t+1)} = \frac{1}{N}\sum_i \boldsymbol{y}_i^{(t)} \tag{B.42}$$

The main difference is apparent when updating $\boldsymbol{W}$:

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{W}} &= -\frac{N}{2}\frac{\partial}{\partial \boldsymbol{W}} log|\boldsymbol{C}| - \frac{1}{2}\sum_i^N \frac{\partial}{\partial \boldsymbol{W}}\mathbf{E}\left[(\boldsymbol{y}_i - \boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})\right] \\
&= -\frac{N}{2}\frac{\partial}{\partial \boldsymbol{W}} log|\boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}| - \frac{1}{2}\sum_i^N \frac{\partial}{\partial \boldsymbol{W}}\mathbf{E}\left[(\boldsymbol{y}_i - \boldsymbol{\mu})^T(\boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})\right]^T \\
&= -\frac{N}{2}(2\boldsymbol{C}^{-1}\boldsymbol{W}) - \frac{1}{2}\sum_i \mathbf{E}\left[-\boldsymbol{W}\boldsymbol{C}^{-1}2(\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})^T\boldsymbol{C}^{-1}\right]^T \\
&= N(-\boldsymbol{C}^{-1}\boldsymbol{W} + \boldsymbol{C}^{-1}\mathbf{E}\left[\sum_i \frac{1}{N}(\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})^T\right]\boldsymbol{C}^{-1}\boldsymbol{W})
\end{aligned} \tag{B.43}$$

following the matrix cookbook [92]. The *empirical* covariance $\boldsymbol{\Sigma}^{(t+1)} = \mathbf{E}\left[\sum_i \frac{1}{N}(\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})^T + \boldsymbol{S}_i^{(t)}\right]$ is defined in (B.17). Therefore:

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{W}} &= N(\boldsymbol{C}^{-1}\boldsymbol{\Sigma}^{(t+1)}\boldsymbol{C}^{-1}\boldsymbol{W} - \boldsymbol{C}^{-1}\boldsymbol{W}) = 0 \\
\boldsymbol{C}^{-1}\boldsymbol{\Sigma}^{(t+1)}\boldsymbol{C}^{-1}\boldsymbol{W} &= \boldsymbol{C}^{-1}\boldsymbol{W} \\
\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{(t+1)}\sigma^2\boldsymbol{I}(\boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{W} &= \boldsymbol{W} \\
\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{(t+1)}((\boldsymbol{W}^T\boldsymbol{W})^{-1} + \frac{1}{\sigma^2}\boldsymbol{I})^{-1} &= \boldsymbol{W}^T\boldsymbol{W} \\
\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{(t+1)} &= \boldsymbol{W}^T\boldsymbol{W}((\boldsymbol{W}^T\boldsymbol{W})^{-1} + \frac{1}{\sigma^2}\boldsymbol{I}) \\
\boldsymbol{\Sigma}^{(t+1)} &= \sigma^2\boldsymbol{I} + \boldsymbol{W}^T\boldsymbol{W}
\end{aligned} \tag{B.44}$$

Consider the singular value decomposition (SVD) of $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{V}^T$, where $\boldsymbol{U}$ is a $d \times p$ matrix (since we only keep the first p components), $\boldsymbol{L}$ is a $p \times p$ diagonal matrix, and $\boldsymbol{V}$ is a $p \times p$ rotation matrix.

$$\begin{aligned}
\boldsymbol{\Sigma}^{(t+1)}\boldsymbol{U}L &= \sigma^2\boldsymbol{U}L + \boldsymbol{U}\boldsymbol{L}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{L}^T\boldsymbol{U}^T\boldsymbol{U}L \\
\boldsymbol{\Sigma}^{(t+1)}\boldsymbol{U}L &= \boldsymbol{U}(\sigma^2\boldsymbol{I} + \mathbf{L}^2)\mathbf{L} \\
\boldsymbol{\Sigma}^{(t+1)}\boldsymbol{u}_j &= (\sigma^2 + l_j^2)\boldsymbol{u}_j
\end{aligned} \tag{B.45}$$

Therefore $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{\Sigma}^{(t+1)}$, with eigenvalues $\lambda_j = \sigma^2 + l_j^2$, and $l_j = \sqrt{\lambda_j - \sigma^2}$. Excluding zero solutions, we have:

$$\begin{aligned}
\boldsymbol{W}^{(t+1)} &= \boldsymbol{U}\boldsymbol{L}\boldsymbol{V}^T \\
&= \boldsymbol{U}(\boldsymbol{\Lambda} - \sigma^2\boldsymbol{I})^{1/2}\boldsymbol{V}^T \\
&= \boldsymbol{U}(\boldsymbol{\Lambda} - \sigma^2\boldsymbol{I})^{1/2}
\end{aligned} \tag{B.46}$$

where $[\boldsymbol{\Lambda}]_j = \lambda_j$, and $\boldsymbol{V}^T$ is an arbitrary rotation vector which we set to $\boldsymbol{I}$. From [131], at optimal $\boldsymbol{W}$,

$$\sigma^{(t+1)2} = \frac{1}{d-q}\sum_{j=q+1}^{d}\lambda_j \tag{B.47}$$

Therefore, the algorithm first computes expectation step via (B.40) and (B.41). In the maximization step, we obtain $\boldsymbol{\mu}^{(t+1)}$ via (B.42), and form $\boldsymbol{\Sigma}^{(t+1)}$ via (B.17), and ob-

tain $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ through eigenvalue decomposition of $\boldsymbol{\Sigma}^{(t+1)}$. Finally, we estimate $\sigma^{2,(t+1)}$ via (B.47), followed by $\boldsymbol{W}^{(t+1)}$ via (B.46).

■ B.2.5 ML-PCA with Transformed Patches with Missing Voxels

As before, adding the notation of transformed patches simply affects which image space each update is computed in. Specifically $\boldsymbol{\Gamma}_i$ and $\boldsymbol{R}_i$ transform the subject space vectorized patch $\tilde{\boldsymbol{y}}_i$ to atlas space $\boldsymbol{y}_i$, and vice-versa via (B.19). The patch $\boldsymbol{y}_i$, for $i \in 1..N$ is drawn from a multivariate normal of dimension d:

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{C}), \tag{B.48}$$

where $\boldsymbol{C} = \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}$, and therefore

$$\tilde{\boldsymbol{y}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{C}}_i), \tag{B.49}$$

where

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{R}_i\boldsymbol{\mu} \tag{B.50}$$

$$\tilde{\boldsymbol{C}}_i = \boldsymbol{R}_i\boldsymbol{C}\boldsymbol{R}_i^T. \tag{B.51}$$

We form the observed data likelihood:

$$p(\tilde{\mathcal{Y}}^{\mathcal{O}_i}; \mathcal{R}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \prod_i \mathcal{N}(\boldsymbol{y}_i, \boldsymbol{\mu}_i^{\mathcal{O}_i}; \boldsymbol{C}_i^{\mathcal{O}_i\mathcal{O}_i}) \tag{B.52}$$

And again use the ECM algorithm. We are led to the subject-space expectation step:

$$\begin{aligned}
\mathbf{E}\left[\tilde{y}_{ij}\right] &= \tilde{y}_{ij}^{(t)} \\
\tilde{y}_{ij}^{(t)} &= \begin{cases} \tilde{y}_{ij} & \text{if } \tilde{y}_{ij} \text{ is observed} \\ \tilde{\mu}_{ij}^{(t)} + (\tilde{\boldsymbol{C}}_i^{j\mathcal{O}_i})^{(t)}(\tilde{\boldsymbol{C}}_i^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\tilde{\boldsymbol{y}}_i^{\mathcal{O}_i} - (\tilde{\boldsymbol{\mu}}_i^{\mathcal{O}_i})^{(t)}) & \text{otherwise} \end{cases}
\end{aligned} \tag{B.53}$$

$$\begin{aligned}
\mathbf{E}\left[\tilde{y}_{ij}\tilde{y}_{ik}\right] &= \mathbf{E}[\tilde{y}_{ij}]\mathbf{E}[\tilde{y}_{ik}] + s_{ijk}^{(t)} \\
&= \tilde{y}_{ij}^{(t)}\tilde{y}_{ik}^{(t)} + s_{ijk}^{(t)} \\
s_{ijk}^{(t)} &= \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ is observed} \\ \tilde{\boldsymbol{C}}_{i,jk}^{(t)} - (\tilde{\boldsymbol{C}}_i^{\mathcal{O}_i j})^{(t),T}(\tilde{\boldsymbol{C}}_i^{\mathcal{O}_i\mathcal{O}_i})^{(t),-1}(\tilde{\boldsymbol{C}}_i^{\mathcal{O}_i k})^{(t)} & \text{otherwise} \end{cases}
\end{aligned} \tag{B.54}$$

while the maximization step updates follow the same form as the previous section, in atlas-space:

$$\boldsymbol{\mu}^{(t+1)} = \frac{1}{N}\sum_i \boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} \tag{B.55}$$

$$\boldsymbol{\Sigma}^{(t+1)} = \frac{1}{N}\sum_i (\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})^T + \boldsymbol{\Gamma}_i \boldsymbol{S}_i^{(t)} \boldsymbol{\Gamma}_i^T, \tag{B.56}$$

and, given $\boldsymbol{U}^{(t+1)}\boldsymbol{\Lambda}^{(t+1)}\boldsymbol{V}^{(t+1)T} = \text{SVD}(\boldsymbol{\Sigma}^{(t+1)})$,

$$\begin{aligned}
\sigma^{(t+1)2} &= \frac{1}{d-q}\sum_{j=q+1}^{d} \lambda_j^{(t+1)} \\
\boldsymbol{W}^{(t+1)} &= \boldsymbol{U}^{(t+1)}(\boldsymbol{\Lambda}^{(t+1)} - \sigma^{(t+1)2}\boldsymbol{I})^{1/2} \\
\boldsymbol{C}^{(t+1)} &= \boldsymbol{W}^{(t+1)T}\boldsymbol{W}^{(t+1)} + \sigma^{(t+1)2}\boldsymbol{I}.
\end{aligned} \tag{B.57}$$

■ B.2.6 PPCA with Missing Data

In this section, we explicitly model the low-dimensional patch representation $\boldsymbol{x}$ as a latent variable. We let $\boldsymbol{C} = \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}$. Following (B.38), the likelihood of the observed data is

$$p(\mathcal{Y}^{\mathcal{O}}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \prod_i \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}^{\mathcal{O}_i}, \boldsymbol{C}^{\mathcal{O}_i\mathcal{O}_i}) \tag{B.58}$$

whereas the joint probability of the observed data and latent variables:

$$p(\mathcal{Y}^{\mathcal{O}}, \boldsymbol{x}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \prod_i p(\boldsymbol{y}^{\mathcal{O}_i}|\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) p(\boldsymbol{x}_i) \tag{B.59}$$

$$= \prod_i \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i + \boldsymbol{\mu}^{\mathcal{O}_i}, \sigma^2\boldsymbol{I}) \mathcal{N}(\boldsymbol{x}_i; 0, \boldsymbol{I}) \tag{B.60}$$

where the notation $\boldsymbol{W}^{\mathcal{O}_i}$ extracts the rows of $\boldsymbol{W}$ corresponding to the observed entries in $\boldsymbol{y}_i$. Finally, the posterior probability is:

$$p(\boldsymbol{x}|\mathcal{Y}^{\mathcal{O}}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = p(\mathcal{Y}^{\mathcal{O}}, \boldsymbol{x}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) / p(\mathcal{Y}^{\mathcal{O}}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \tag{B.61}$$

$$= \prod_i \frac{\mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i + \boldsymbol{\mu}^{\mathcal{O}_i}, \sigma^2\boldsymbol{I}) \mathcal{N}(\boldsymbol{x}_i; 0, \boldsymbol{I})}{\mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}^{\mathcal{O}_i}, \boldsymbol{C}^{\mathcal{O}_i\mathcal{O}_i})} \tag{B.62}$$

$$\propto \prod_i \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{m}_{ci}, \boldsymbol{C}_{ci}) \tag{B.63}$$

where

$$\begin{aligned} \boldsymbol{C}_{ci} &= \sigma^2((\boldsymbol{W}^{\mathcal{O}_i})^T\boldsymbol{W}^{\mathcal{O}_i} + \sigma^2\boldsymbol{I})^{-1} \\ \boldsymbol{m}_{ci} &= \boldsymbol{C}_c \frac{1}{\sigma^2}(\boldsymbol{W}^{\mathcal{O}_i})^T(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{\mu}^{\mathcal{O}_i}) \end{aligned} \tag{B.64}$$

using (B.38) and (B.39).

We use the Expectation Maximization algorithm, and look to optimize:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbf{E}_{\boldsymbol{x}|\mathcal{Y}^{\mathcal{O}}, \boldsymbol{\theta}^{(t)}} \left[\log p(\mathcal{Y}^{\mathcal{O}}, \boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{W}, \sigma^2)\right] \tag{B.65}$$

$$= -\frac{dN}{2}\log 2\pi - \frac{N}{2}\log|\sigma^2\boldsymbol{I}| \tag{B.66}$$

$$- \frac{1}{2}\sum_i^N \frac{1}{\sigma^2}\mathbf{E}\left[(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})^T(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})\right] \tag{B.67}$$

$$- \frac{dN}{2}\log 2\pi - \frac{N}{2}\log|\boldsymbol{I}| - \frac{1}{2}\sum_i^N \mathbf{E}\left[\boldsymbol{x}_i^T\boldsymbol{x}_i\right]. \tag{B.68}$$

The **expectation step** computes the statistics over the latent variables:

$$\begin{aligned}\mathbf{E}\left[\boldsymbol{x}_i\right] &= \boldsymbol{x}_i^{(t)} \\ &= \int_{\boldsymbol{x}_i} \boldsymbol{x}_i \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{m}_{ci}, \boldsymbol{C}_{ci}) d\boldsymbol{x}_i \\ &= \boldsymbol{m}_{ci}^{(t)}\end{aligned} \tag{B.69}$$

and

$$\begin{aligned}\mathbf{E}\left[\boldsymbol{x}_i \boldsymbol{x}_i^T\right] &= \mathbf{E}\left[\boldsymbol{x}_i\right] \mathbf{E}\left[\boldsymbol{x}_i\right]^T + cov(\boldsymbol{x}_i, \boldsymbol{x}_i) \\ &= \boldsymbol{x}_i^{(t)} \boldsymbol{x}_i^{(t),T} + \boldsymbol{C}_{ci}^{(t)}\end{aligned} \tag{B.70}$$

The **maximization step** updates the model parameters:

$$\begin{aligned}\frac{\partial Q}{\partial \boldsymbol{\mu}} &= \sum_i \frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\mu}} \mathbf{E}\left[(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})^T (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})\right] \\ &= \sum_i \frac{1}{2\sigma^2} \mathbf{E}\left[-2(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})\right] \\ &= \sum_i \frac{1}{\sigma^2} (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i^{(t)} - \boldsymbol{\mu}^{\mathcal{O}_i}) = 0 \\ \mu_j &= \frac{1}{|\mathcal{O}^j|} \sum_{i \in \mathcal{O}^j} (y_{i,j} - \boldsymbol{W}_j \boldsymbol{x}_i^{(t)})\end{aligned} \tag{B.71}$$

where $\mathcal{O}^j$ is the set of patches that have entry j observed.

$$\begin{aligned}\frac{\partial Q}{\partial \boldsymbol{W}} &= \sum_i \frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{W}} \mathbf{E}\left[(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})^T (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})\right] \\ &= \sum_i \frac{1}{2\sigma^2} \mathbf{E}\left[2(\boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{\mu}^{\mathcal{O}_i}))\boldsymbol{x}_i^T\right] \\ &= \frac{1}{\sigma^2} \sum_i \boldsymbol{W}^{\mathcal{O}_i}(\boldsymbol{x}_i^{(t)} \boldsymbol{x}_i^{(t),T} + \boldsymbol{C}_{ci}^{(t)}) - (\boldsymbol{y}_i^{\mathcal{O}_i} - (\boldsymbol{\mu}^{\mathcal{O}_i})^{(t)}) \boldsymbol{x}_i^{(t),T} = 0 \\ \boldsymbol{W}_j^{(t+1)} &= (\sum_{i \in \mathcal{O}^j} (\boldsymbol{x}_i^{(t)} \boldsymbol{x}_i^{(t),T} + \boldsymbol{C}_{ci}^{(t)}))^{-1} \sum_{i \in \mathcal{O}^j} (y_{i,j} - \mu_j^{(t+1)}) \boldsymbol{x}_i^{(t),T}.\end{aligned} \tag{B.72}$$

and, finally

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma^2} &= \sum_i -\frac{\partial}{\partial \sigma^2}\frac{1}{2\sigma^2}\mathbf{E}\left[(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})^T(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})\right] - \frac{\partial}{\partial \sigma^2}\frac{N}{2}log\sigma^2 \\
&= \sum_i \frac{1}{2\sigma^4}\mathbf{E}\left[(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})^T(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})\right] - \frac{N}{2\sigma^2} = 0 \\
\sigma^{2,(t+1)} &= \sum_i \mathbf{E}\left[(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})^T(\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}_i}\boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})\right] \\
&= \sum_{j,i\in\mathcal{O}^j} (y_{i,j} - \boldsymbol{W}_j^{(t+1)}\boldsymbol{x}_i - \mu_j^{(t+1)})^2 + \boldsymbol{W}_j\boldsymbol{C}_{ci}^{(t+1)}\boldsymbol{W}_j^{(t+1),T}
\end{aligned} \tag{B.73}$$

Therefore, the EM algorithm updates the expectations in (B.69) and (B.70), and the parameters $\boldsymbol{\mu}^{(t+1)}$ via (B.71), $\boldsymbol{W}^{(t+1)}$ via (B.72), finally $\sigma^{2,(t+1)}$ via (B.73).

■ B.3 Mixture of Multivariate Normals

We model the data as coming from a mixture of multivariate normals instead of a single normal as in the previous section. Having built up the theory step by step, this final addition is now manageable.

■ B.3.1 Gaussian Mixture with Missing Data

We model the observations $\boldsymbol{y}$ as coming from a Gaussian Mixture Model:

$$p(\boldsymbol{y}|\pi, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_k \pi_k \mathcal{N}(\boldsymbol{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{B.74}$$

The clusters are captured by binary vector variable $\boldsymbol{z}, z_k = \{0, 1\}$, and $p(z_k = 1) = \pi_k$. We let $\mathcal{Y} = \{\boldsymbol{y}_i\}_{i=1}^N$, $\mathcal{Y}_\mathcal{M} = \{\boldsymbol{y}_i^{\mathcal{M}_i}\}_{i=1}^N$ and $\mathcal{Y}^\mathcal{O} = \{\boldsymbol{y}_i^{\mathcal{O}_i}\}_{i=1}^N$, where we model $\boldsymbol{y}_i^{\mathcal{M}_i}$ as latent

variables. We can write the joint probability as:

$$\begin{aligned} p(\mathcal{Y}, \boldsymbol{z}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_i p(\boldsymbol{y}_i, \boldsymbol{z}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \prod_i \prod_k \left(p(z_k = 1; \boldsymbol{\pi}) p(\boldsymbol{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right)^{z_k} \\ &= \prod_i \prod_k \left(\pi_k \mathcal{N}(\boldsymbol{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right)^{z_k} \\ &= \prod_i \prod_k \left(\pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{\Sigma}_k^{\mathcal{O}_i}) \mathcal{N}(\boldsymbol{y}_i^{\mathcal{M}_i}; \boldsymbol{\mu}_{k,\mathcal{M}_i|\mathcal{O}_i}, \boldsymbol{\Sigma}_{k,\mathcal{M}_i|\mathcal{O}_i})\right)^{z_k} \end{aligned} \tag{B.75}$$

from which we can derive the likelihood of our observed data:

$$\begin{aligned} & p(\mathcal{Y}^{\mathcal{O}}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \int_{\mathcal{Y}_{\mathcal{M}_i}} \sum_{\boldsymbol{z}} p(\boldsymbol{Y}, \boldsymbol{z}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \dagger \mathcal{Y}_{\mathcal{M}_i} \\ &= \prod_i \sum_{\boldsymbol{z}} \prod_k \int_{\boldsymbol{y}_{\mathcal{M}_i}} \left(\pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{\Sigma}_k^{\mathcal{O}_i}) \mathcal{N}(\boldsymbol{y}_i^{\mathcal{M}_i}; \boldsymbol{\mu}_{k,\mathcal{M}_i|\mathcal{O}_i}, \boldsymbol{\Sigma}_{k,\mathcal{M}_i|\mathcal{O}_i})\right)^{z_k} d\boldsymbol{y}_{\mathcal{M}_i} \\ &= \prod_i \sum_k \pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{\Sigma}_k^{\mathcal{O}_i}) \int_{\boldsymbol{y}_{\mathcal{M}_i}} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{M}_i}; \boldsymbol{\mu}_{k,\mathcal{M}_i|\mathcal{O}_i}, \boldsymbol{\Sigma}_{k,\mathcal{M}_i|\mathcal{O}_i}) d\boldsymbol{y}_{\mathcal{M}_i} \\ &= \prod_i \sum_k \pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{\Sigma}_k^{\mathcal{O}_i}) \end{aligned} \tag{B.76}$$

where the third step is due to the fact that exactly one entry of $\boldsymbol{z}$ is 1 at the same time. Finally, the posterior can be written as:

$$p(\mathcal{Y}^{\mathcal{M}}, \boldsymbol{z} | \mathcal{Y}^{\mathcal{O}}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_i \frac{p(\boldsymbol{y}_i, \boldsymbol{z}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{p(\boldsymbol{y}^{\mathcal{O}_i}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})} \tag{B.77}$$

As before, we maximize the log likelihood via EM:

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) =& \mathbf{E}_{\mathcal{Y}^{\mathcal{M}}, \boldsymbol{z} | \mathcal{Y}^{\mathcal{O}}; \boldsymbol{\pi}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}} \left[\log p(\mathcal{Y}^{\mathcal{O}}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] \\ =& \mathbf{E}[\sum_i \sum_k z_k (-\frac{d}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}))] \\ =& \sum_i \sum_k \mathbf{E}[z_k (-\frac{d}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}))] \end{aligned} \tag{B.78}$$

The **expectation step** updates the expectations:

$$\begin{aligned}
\mathbf{E}[z_k] &= \int_{\boldsymbol{y}_{\mathcal{M}_i}} \sum_{z_k} z_k \frac{p(\boldsymbol{y}_i, \boldsymbol{z}; \boldsymbol{\pi}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)})}{p(\boldsymbol{y}^{\mathcal{O}_i}; \boldsymbol{\pi}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)})} d\boldsymbol{y}_{\mathcal{M}_i} \\
&= \frac{\pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{\Sigma}_k^{\mathcal{O}_i})^{(t)})}{\sum_k \pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{\Sigma}_k^{\mathcal{O}_i})^{(t)})} \int_{\boldsymbol{y}_{\mathcal{M}_i}} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{M}_i}; \boldsymbol{\mu}_k^{\mathcal{M}_i|\mathcal{O}_i}, \boldsymbol{\Sigma}_k^{\mathcal{M}_i|\mathcal{O}_i}) d\boldsymbol{y}_{\mathcal{M}_i} \\
&= \frac{\pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{\Sigma}_k^{\mathcal{O}_i})^{(t)})}{\sum_k \pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{\Sigma}_k^{\mathcal{O}_i})^{(t)})} \equiv \gamma_{ik}^{(t)}
\end{aligned} \tag{B.79}$$

$$\begin{aligned}
\mathbf{E}[z_k y_{ij}] &= \int_{\boldsymbol{y}_{\mathcal{M}_i}} \sum_{z_k} z_k y_{ij} \frac{p(\boldsymbol{y}_i, \boldsymbol{z}; \boldsymbol{\pi}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)})}{p(\boldsymbol{y}^{\mathcal{O}_i}; \boldsymbol{\pi}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)})} d\boldsymbol{y}_{\mathcal{M}_i} \\
&= \frac{\pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{\Sigma}_k^{\mathcal{O}_i})^{(t)})}{\sum_k \pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{\Sigma}_k^{\mathcal{O}_i})^{(t)})} \int_{\boldsymbol{y}_{\mathcal{M}_i}} y_{ij} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{M}_i}; \boldsymbol{\mu}_k^{\mathcal{M}_i|\mathcal{O}_i}, \boldsymbol{\Sigma}_k^{\mathcal{M}_i|\mathcal{O}_i}) d\boldsymbol{y}_{\mathcal{M}_i} \\
&= \gamma_{ik}^{(t)} E[y_{ij}] \\
&= \gamma_{ik}^{(t)} y_{ij}^{(t)}
\end{aligned} \tag{B.80}$$

where $y_{ij}^{(t)}$ is the expectation in (B.13). Similarly

$$\begin{aligned}
\mathbf{E}[z_k y_{ij} y_{il}] &= = \gamma_{ik}^{(t)} E[y_{ij} y_{il}] \\
&= \gamma_{ik}^{(t)} (y_{ij}^{(t)} y_{ik}^{(t)} + s_{ijl}^{(t)}).
\end{aligned} \tag{B.81}$$

The **maximization step** follows:

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\mu}_k} &= -\frac{1}{2} \sum_i^N \frac{\partial}{\partial \boldsymbol{\mu}_k} \mathbf{E}\left[z_k (\boldsymbol{y}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_k) \right] \\
&= -\sum_i \Sigma^{-1} \gamma_{ik}^{(t)} (\boldsymbol{y}_i^{(t)} - \boldsymbol{\mu}_k) = 0 \\
\boldsymbol{\mu}_k^{(t+1)} &= \frac{1}{\sum_i \gamma_{ik}^{(t)}} \sum_i \gamma_{ik}^{(t)} \boldsymbol{y}_i^{(t)}
\end{aligned} \tag{B.82}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\Sigma}^{-1}} &= \sum_i \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}}\mathbf{E}[z_k] log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}}\mathbf{E}\left[z_k(\boldsymbol{y}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})\right] \\
&= \sum_i \frac{1}{2}\gamma_{ik}^{(t)}\boldsymbol{\Sigma} - \frac{1}{2}\sum_i \mathbf{E}\left[z_k(\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})^T\right] = 0 \\
\boldsymbol{\Sigma}^{(t+1)} &= \frac{1}{\sum_i \gamma_{ik}^{(t)}}\sum_i \gamma_{ik}^{(t)}\left[(\boldsymbol{y}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})(\boldsymbol{y}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})^T + \boldsymbol{C}_i^{(t)}\right] \qquad \text{(B.83)}
\end{aligned}$$

$$\frac{\partial Q}{\partial \pi_k} = \sum_i \mathbf{E}[z_k]\frac{1}{\pi_k} + \lambda = 0 \qquad \text{(B.84)}$$

$$\pi_k = \frac{-1}{\lambda}\sum_i \gamma_{ik} \qquad \text{(B.85)}$$

$$\sum_k \pi_k = \frac{-1}{\lambda}\sum_{ik} \gamma_{ik} \qquad \text{(B.86)}$$

$$\lambda = -N \qquad \text{(B.87)}$$

$$\pi_k^{(t+1)} = \frac{1}{N}\sum_i \gamma_{ik}^{(t)} \qquad \text{(B.88)}$$

where we use a Lagrange multiplier λ to enforce $\sum_k \pi_k = 1$.

Similar changes apply to both the expectation and maximization steps of learning of the multivariate normal with transformed patches model, and multivariate normal with ML-PCA model. We omit these updates since they are not insightful, and give the final updates for a mixture model of low dimensional Gaussians for transformed patches with missing voxels in the next section.

■ B.3.2 Mixture Model of Principal Analysers for Transformed Patches with Missing Voxels

As before, adding the notation of transformed patches simply affects which image space we do each update of the EM algorithm in. Let $\boldsymbol{z}, z_k = \{0, 1\}$, and $p(z_k = 1) = \pi_k$., and $\boldsymbol{y}_i$, for $i \in 1..N$ be drawn from a Gaussian Mixture Model:

$$\boldsymbol{y}_i \sim \sum_k \pi_k \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{C}_k), \qquad \text{(B.89)}$$

where $\boldsymbol{C}_k = \boldsymbol{W}_k^T \boldsymbol{W}_k + \sigma_k^2 \boldsymbol{I}$, and

$$\tilde{\boldsymbol{y}}_i \sim \sum_k \pi_k \mathcal{N}(\boldsymbol{\mu}_{k,i}, \boldsymbol{C}_{k,i}), \tag{B.90}$$

where $\tilde{\boldsymbol{\mu}}_{ki} = \boldsymbol{R}_i \boldsymbol{\mu}_k$, and $\tilde{\boldsymbol{C}}_{k,i} = \boldsymbol{R}_i \boldsymbol{C}_k \boldsymbol{R}_i^T$.

We form the observed data likelihood:

$$p(\tilde{\mathcal{Y}}^{\mathcal{O}_i}; \mathcal{R}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \prod_i \sum_k \pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}, \boldsymbol{\mu}_{ki}^{\mathcal{O}_i}; \boldsymbol{C}_{ki}^{\mathcal{O}_i \mathcal{O}_i}) \tag{B.91}$$

and as before, we maximize the log likelihood via EM:

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) =& \mathbf{E}_{\tilde{\mathcal{Y}}^{\mathcal{M}}, \boldsymbol{z}|\mathcal{Y}^{\mathcal{O}}; \boldsymbol{\pi}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{C}^{(t)}} [\log p(\mathcal{Y}^{\mathcal{O}}; \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{C})] \\ =& \sum_i \sum_k \mathbf{E}[z_k(-\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{C}_{k,i}| - \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{k,i})^T \boldsymbol{C}_{k,i}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{k,i}))]. \end{aligned} \tag{B.92}$$

The **expectation step** follows:

$$\mathbf{E}[z_k] = \frac{\pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_{ki}^{\mathcal{O}_i})^{(t)}, (\boldsymbol{C}_{ki}^{\mathcal{O}_i})^{(t)})}{\sum_k \pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_{ki}^{\mathcal{O}_i})^{(t)}, (\boldsymbol{C}_{ki}^{\mathcal{O}_i})^{(t)})} \equiv \gamma_{ik}^{(t)} \tag{B.93}$$

$$\begin{aligned} \mathbf{E}[z_k y_{i,j}] &= \gamma_{ik}^{(t)} y_{ij}^{(t)} \\ y_{ij}^{(t)} &= \begin{cases} y_{i,j} & \text{if } y_{ij} \text{ is observed} \\ \mu_j^{(t)} + (\boldsymbol{C}_{ij\mathcal{O}_i})^{(t)} (\boldsymbol{C}_i^{\mathcal{O}_i \mathcal{O}_i})^{(t),-1} (\boldsymbol{y}_i^{\mathcal{O}_i} - (\boldsymbol{\mu}^{\mathcal{O}_i})^{(t)}) & \text{otherwise} \end{cases} \end{aligned} \tag{B.94}$$

$$\begin{aligned} \mathbf{E}[z_k y_{ij} y_{ik}] &= \gamma_{ik}^{(t)} (\mathbf{E}[y_{ij}]\mathbf{E}[y_{ik}] + s_{ijk}^{(t)}) \\ &= \gamma_{ik}^{(t)} (y_{ij}^{(t)} y_{ik}^{(t)} + s_{ijk}^{(t)}) \\ s_{ijk}^{(t)} &= \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ is observed} \\ \boldsymbol{C}_{i,jk}^{(t)} - (\boldsymbol{C}_i^{\mathcal{O}_i j})^{(t),T} (\boldsymbol{C}_i^{\mathcal{O}_i \mathcal{O}_i})^{(t),-1} (\boldsymbol{C}_i^{\mathcal{O}_i k})^{(t)} & \text{otherwise} \end{cases} \end{aligned} \tag{B.95}$$

while the maximization step updates are computed in atlas-space:

$$\begin{aligned}
\boldsymbol{\mu}^{(t+1)} &= \frac{1}{\sum_i \gamma_{ik}^{(t)}} \sum_i \gamma_{ik}^{(t)} \boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} \\
\boldsymbol{\Sigma}^{(t+1)} &= \frac{1}{\sum_i \gamma_{ik}^{(t)}} \sum_i \gamma_{ik}^{(t)} \left[(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})(\boldsymbol{\Gamma}_i \tilde{\boldsymbol{y}}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})^T + \boldsymbol{\Gamma}_i \tilde{\boldsymbol{S}}_i^{(t)} \boldsymbol{\Gamma}_i^T \right] \\
\pi_k^{(t+1)} &= \frac{1}{N} \sum_i \gamma_{ik}^{(t)},
\end{aligned} \tag{B.96}$$

and, given $\boldsymbol{U}^{(t+1)} \boldsymbol{\Lambda}^{(t+1)} \boldsymbol{V}^{(t+1)T} = \text{SVD}(\boldsymbol{\Sigma}^{(t+1)})$,

$$\begin{aligned}
\sigma^{(t+1)2} &= \frac{1}{d-q} \sum_{j=q+1}^{d} \lambda_j^{(t+1)} \\
\boldsymbol{W}^{(t+1)} &= \boldsymbol{U}^{(t+1)} (\boldsymbol{\Lambda}^{(t+1)} - \sigma^{(t+1)2} \boldsymbol{I})^{1/2} \\
\boldsymbol{C}^{(t+1)} &= \boldsymbol{W}^{(t+1)T} \boldsymbol{W}^{(t+1)} + \sigma^{(t+1)2} \boldsymbol{I}.
\end{aligned} \tag{B.97}$$

■ B.3.3 Mixture Model of Probabilistic Principal Analysers for Transformed Patches with Missing Voxels

In this section, we extend the probabilistic PCA with missing data model B.2.6. We model binary variable $\boldsymbol{z}, z_k = \{0, 1\}$, and $p(z_k = 1) = \pi_k$. We let $\boldsymbol{C}_k = \boldsymbol{W}_k^T \boldsymbol{W}_k + \sigma_k^2 \boldsymbol{I}$. The likelihood of the observed data is

$$p(\mathcal{Y}^{\mathcal{O}}; \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \prod_i \sum_k \pi_k \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; \boldsymbol{\mu}_k^{\mathcal{O}_i}, \boldsymbol{C}_k^{\mathcal{O}_i \mathcal{O}_i}) \tag{B.98}$$

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbf{E}_{\boldsymbol{x}|\mathcal{Y}^{\mathcal{O}}, \boldsymbol{\theta}^{(t)}} \left[\log p(\mathcal{Y}^{\mathcal{O}}, \boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{W}, \sigma^2) \right] \tag{B.99}$$

$$= \sum_{i,k} \mathbf{E}[z_k (-\frac{d}{2} \log 2\pi - \frac{1}{2} \log |\sigma^2 \boldsymbol{I}| \tag{B.100}$$

$$- \frac{1}{2\sigma^2} (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}} \boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i})^T (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{W}^{\mathcal{O}} \boldsymbol{x}_i - \boldsymbol{\mu}^{\mathcal{O}_i}) \tag{B.101}$$

$$- \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{I}| - \frac{1}{2} \boldsymbol{x}_i^T \boldsymbol{x}_i)]. \tag{B.102}$$

The **expectation step** computes the statistics over the latent variables:

$$\mathbf{E}[z_k] = \frac{\pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{C}_k^{\mathcal{O}_i})^{(t)})}{\sum_k \pi_k^{(t)} \mathcal{N}(\boldsymbol{y}_i^{\mathcal{O}_i}; (\boldsymbol{\mu}_k^{\mathcal{O}_i})^{(t)}, (\boldsymbol{C}_k^{\mathcal{O}_i})^{(t)})} \equiv \gamma_{ik}^{(t)} \tag{B.103}$$

$$\mathbf{E}\left[z_k \boldsymbol{x}_i\right] = \gamma_{ik}^{(t)} \boldsymbol{m}_{cki}^{(t)} \tag{B.104}$$

$$\mathbf{E}\left[z_k \boldsymbol{x}_i \boldsymbol{x}_i^T\right] = \gamma_{ik}^{(t)} \left(\boldsymbol{x}_i^{(t)} \boldsymbol{x}_i^{(t),T} + \boldsymbol{C}_{cki}^{(t)}\right) \tag{B.105}$$

Where

$$\begin{aligned} \boldsymbol{m}_{ci} &= \boldsymbol{C}_c \frac{1}{\sigma^2} (\boldsymbol{W}^{\mathcal{O}})^T (\boldsymbol{y}_i^{\mathcal{O}_i} - \boldsymbol{\mu}^{\mathcal{O}_i}) \\ \boldsymbol{C}_{ci} &= \sigma^2 ((\boldsymbol{W}^{\mathcal{O}})^T \boldsymbol{W}^{\mathcal{O}} + \sigma^2 \boldsymbol{I})^{-1}. \end{aligned} \tag{B.106}$$

The **maximization step** follows standard derivatives with respect to the parameters:

$$\mu_j^{(t+1)} = \frac{1}{\sum_{i \in \mathcal{O}^j} \gamma_{ik}^{(t)}} \sum_{i \in \mathcal{O}^j} \gamma_{ik}^{(t)} (y_{ij} - \boldsymbol{W} \boldsymbol{x}_i^{(t)}) \tag{B.107}$$

$$\boldsymbol{W}^{(t+1)} = (\sum_{i \in \mathcal{O}^j} \gamma_{ik}^{(t)} (\boldsymbol{x}_i^{(t)} \boldsymbol{x}_i^{(t),T} + \boldsymbol{C}_{ci}^{(t)}))^{-1} \sum_{i \in \mathcal{O}^j} \gamma_{ik}^{(t)} (y_{ij} - \mu_j^{(t+1)}) \boldsymbol{x}_i^{(t),T} \tag{B.108}$$

$$\sigma^{2,(t+1)} = \sum_{i \in \mathcal{O}^j} \gamma_{ik}^{(t)} \left[(y_{ij} - \boldsymbol{W}^{(t+1)} \boldsymbol{x}_i - \mu_j^{(t+1)})^2 + \boldsymbol{W} \boldsymbol{C}_{ci}^{(t+1)} \boldsymbol{W}^{(t+1),T} \right] \tag{B.109}$$

$$\pi_k^{(t+1)} = \frac{1}{N} \sum_i \gamma_{ik}^{(t)}. \tag{B.110}$$

Therefore, the EM algorithm consists of computing the expectation step via (B.105). In the maximization step, we obtain $\boldsymbol{\mu}^{(t+1)}$, $\boldsymbol{W}^{(t+1)}$, and $\sigma^{2,(t+1)}$ via (B.110).

Bibliography

[1] Petronella Anbeek, Koen L Vincken, Matthias JP Van Osch, Robertus HC Bisschops, and Jeroen Van Der Grond. Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage*, 21(3):1037–1044, 2004.

[2] Alzheimer's Association et al. 2011 Alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 7(2):208, 2011.

[3] National Electrical Manufacturers Association, American College of Radiology, et al. *Digital imaging and communications in medicine (DICOM)*. National Electrical Manufacturers Association, 1998.

[4] Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.

[5] Nematollah K Batmanghelich, Adrian V Dalca, Mert R Sabuncu, and Polina Golland. Joint modeling of imaging and genetics. *Information Processing in Medical Imaging*, LNCS 7917:766–777, 2013.

[6] Nematollah K Batmanghelich, Adrian V Dalca, Gerald Quon, Mert R Sabuncu, and Polina Golland. Probabilistic modeling of imaging, genetics and diagnosis. *IEEE Transactions on Medical Imaging*, 35(7):1765–79, 2016.

[7] Komal Kumar Bhatia, Akhila Rao, Anthony N Price, Robin Wolz, Joseph V Hajnal, and Daniel Rueckert. Hierarchical manifold learning for regional image analysis. *IEEE Transactions on Medical Imaging*, 33(2):444–461, 2014.

[8] Alessandro Biffi, Christopher D Anderson, Rahul S Desikan, Mert Sabuncu, Lynelle Cortellini, Nick Schmansky, David Salat, and Jonathan Rosand. Genetic

variation and neuroimaging measures in alzheimer disease. *Archives of neurology*, 67(6):677–685, 2010.

[9] Rémi Blanc, Christof Seiler, Gabor Székely, Lutz-Peter Nolte, and Mauricio Reyes. Statistical model based shape prediction from a combination of direct observations and various surrogates: application to orthopaedic research. *Med Image Anal*, 16(6):1156–1166, 2012.

[10] Kaj Blennow, Harald Hampel, Michael Weiner, and Henrik Zetterberg. Cerebrospinal fluid and plasma biomarkers in alzheimer disease. *Nature Reviews Neurology*, 6(3):131–144, 2010.

[11] Carolina Bonilla, Lesley-Anne Boxill, Stacey Ann Mc Donald, Tyisha Williams, Nadeje Sylvester, Esteban J Parra, Sonia Dios, Heather L Norton, Mark D Shriver, and Rick A Kittles. The 8818g allele of the agouti signaling protein (asip) gene is ancestral and is associated with darker skin color in african americans. *Human genetics*, 116(5):402–406, 2005.

[12] Ron Brookmeyer, Elizabeth Johnson, Kathryn Ziegler-Graham, and H Michael Arrighi. Forecasting the global burden of Alzheimers disease. *Alzheimer's & dementia*, 3(3):186–191, 2007.

[13] Brian L Browning and Sharon R Browning. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics*, 84(2):210–223, 2009.

[14] Eyal Carmi, Siuyan Liu, Noga Alon, Amos Fiat, and Daniel Fiat. Resolution enhancement in MRI. *Magnetic resonance imaging*, 24(2):133–154, 2006.

[15] Matthew C Clark, Lawrence O Hall, Dmitry B Goldgof, Robert Velthuizen, F Reed Murtagh, and Martin S Silbiger. Automatic tumor segmentation using knowledge-based techniques. *IEEE Transactions on Medical Imaging*, 17(2): 187–201, 1998.

[16] Timothy F Cootes, Andrew Hill, Christopher J Taylor, and Jane Haslam. Use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–365, 1994.

[17] Pierrick Coupé, José V Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles, and Louis D Collins. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954, 2011.

[18] Robert W Cox, John Ashburner, Hester Breman, Kate Fissell, Christian Haselgrove, Colin J Holmes, Jack L Lancaster, David E Rex, Stephen M Smith, Jeffrey B Woodward, et al. A (sort of) new image data format standard: Nifti-1. *Human Brain Mapping*, 25:33, 2004.

[19] Marc Cruts, Ilse Gijselinck, Julie van der Zee, Sebastiaan Engelborghs, Hans Wils, Daniel Pirici, Rosa Rademakers, Rik Vandenberghe, Bart Dermaut, Jean-Jacques Martin, et al. Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature*, 442(7105):920–924, 2006.

[20] Antonio Culebras, Carlos S Kase, Joseph C Masdeu, Allan J Fox, R Nick Bryan, C Barrie Grossman, Donald H Lee, Harold P Adams, William Thies, et al. Practice guidelines for the use of imaging in transient ischemic attacks and acute stroke A report of the Stroke Council, American Heart Association. *Stroke*, 28 (7):1480–1497, 1997.

[21] Adrian V Dalca. The MIT Boston Timescape Project. http://timescape.mit.edu/, 2016.

[22] Adrian V Dalca, Giovanna Danagoulian, Ron Kikinis, Ehud Schmidt, and Polina Golland. Segmentation of nerve bundles and ganglia in spine mri using particle filters. *MICCAI: International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 6893:537–545, 2011.

[23] Adrian V Dalca, Ramesh Sridharan, Lisa Cloonan, Kaitlin M Fitzpatrick, Allison Kanakis, Karen L Furie, Jonathan Rosand, Ona Wu, Mert Sabuncu, Natalia S Rost, and Polina Golland. Segmentation of cerebrovascular pathologies in stroke patients with spatial and shape priors. *MICCAI: International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 8674:773–780, 2014.

[24] Adrian V Dalca, Ramesh Sridharan, Natalia Rost, and Polina Golland. tipiX:

Rapid visualization of large image collections. In *MICCAI-IMIC Interactive Medical Image Computing Workshop*, 2014.

[25] Adrian V Dalca, Ramesh Sridharan, Mert R Sabuncu, and Polina Golland. Predictive modeling of anatomy with genetic and clinical data. *MICCAI: International Conference on Medical Image Computing and Computer Assisted Intervention*, LNCS 9351:519–526, 2015.

[26] Adrian V Dalca, Andreea Bobu, Natalia S Rost, and Polina Golland. Patch-based discrete registration of clinical brain images. In *MICCAI-PATCHMI Patch-based Techniques in Medical Imaging*. Springer, 2016.

[27] Manasi Datar, Prasanna Muralidharan, Abhishek Kumar, Sylvain Gouttard, Joseph Piven, Guido Gerig, et al. Mixed-effects shape models for estimating longitudinal changes in anatomy. *Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data*, LNCS 7570:76–87, 2012.

[28] Bradley C Davis, P Thomas Fletcher, Elizabeth Bullitt, and Sarang Joshi. Population shape regression from random design data. *IEEE International Conference on Computer Vision*, pages 1–7, 2007.

[29] Bradley C Davis, P Thomas Fletcher, Elizabeth Bullitt, and Sarang Joshi. Population shape regression from random design data. *International Journal of Computer Vision*, 90(2):255–266, 2010.

[30] Charles-Alban Deledalle, Joseph Salmon, and Arnak S Dalalyan. Image denoising with patch based PCA: local versus global. *British Machine Vision Conference*, 81:425–455, 2011.

[31] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 39(1):1–38, 1977.

[32] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

[33] James D Doecke, Simon M Laws, Noel G Faux, William Wilson, Samantha C Burnham, Chiou-Peng Lam, Alinda Mondal, Justin Bedo, Ashley I Bush, Belinda Brown, et al. Blood-based protein biomarkers for diagnosis of alzheimer disease. *Archives of neurology*, 69(10):1318–1325, 2012.

[34] Nikos Drakos and Ross Moore. Gaussiaon process. http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html.

[35] Florence Dru and Tom Vercauteren. An ITK implementation of the symmetric log-domain diffeomorphic demons algorithm. *The Insight Journal*, 2009.

[36] Bruno Dubois, Howard H Feldman, Claudia Jacova, Steven T DeKosky, Pascale Barberger-Gateau, Jeffrey Cummings, André Delacourte, Douglas Galasko, Serge Gauthier, Gregory Jicha, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, 6(8): 734–746, 2007.

[37] Stanley Durrleman, Xavier Pennec, Alain Trouvé, José Braga, Guido Gerig, and Nicholas Ayache. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *International journal of computer vision*, 103(1):22–59, 2013.

[38] Alexei A Efros and William T Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of Annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 341–346. ACM, 2001.

[39] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006.

[40] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[41] Greg M Fleishman, Boris A Gutman, P Thomas Fletcher, and Paul M Thompson. Simultaneous longitudinal registration with group-wise similarity prior. *Information Processing in Medical Imaging*, LNCS 9123:746–757, 2015.

[42] Karl J Friston, John T Ashburner, Stefan J Kiebel, Thomas E Nichols, and William D Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images*. Academic Press, 2011.

[43] Yang Gao. Mriwatcher, 2013. URL http://www.nitrc.org/projects/mriwatcher/.

[44] Tian Ge, Thomas E Nichols, Debashis Ghosh, Elizabeth C Mormino, Jordan W Smoller, Mert R Sabuncu, and Alzheimer's Disease Neuroimaging Initiative. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. *NeuroImage*, 109:505–514, 2015.

[45] Ezequiel Geremia, Olivier Clatz, Bjoern H Menze, Ender Konukoglu, Antonio Criminisi, and Nicholas Ayache. Spatial decision forests for ms lesion segmentation in multi-channel magnetic resonance images. *NeuroImage*, 57(2):378–390, 2011.

[46] Richard A Gibbs, John W Belmont, Paul Hardenbol, Thomas D Willis, Fuli Yu, Huanming Yang, Lan-Yang Ch'ang, Wei Huang, Bin Liu, Yan Shen, et al. The international hapmap project. *Nature*, 426(6968):789–796, 2003.

[47] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *IEEE International Conference on Computer Vision*, pages 349–356. IEEE, 2009.

[48] Polina Golland, W Eric L Grimson, Martha E Shenton, and Ron Kikinis. Detection and analysis of statistical differences in anatomical shape. *Medical image analysis*, 9(1):69–86, 2005.

[49] Tom Haeck, Frederik Maes, and Paul Suetens. ISLES challenge 2015: Automated model-based segmentation of ischemic stroke in MR images. *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, LNCS 9556:246–253, 2015.

[50] Daniel Haehn. Slice:drop: collaborative medical imaging in the browser. In *ACM SIGGRAPH Computer Animation Festival*, pages 1–1. ACM, 2013.

[51] Daniel Haehn, Nicolas Rannou, Banu Ahtam, Ellen Grant, and Rudolph Pienaar. Neuroimaging in the browser using the X Toolkit. *Frontiers in Neuroinformatics*, 101, 2014.

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

[53] Harry Hochheiser, Bruce J Aronow, Kristin Artinger, Terri H Beaty, James F Brinkley, Yang Chai, David Clouthier, Michael L Cunningham, Michael Dixon, Leah Rae Donahue, et al. The FaceBase consortium: a comprehensive program to facilitate craniofacial research. *Developmental biology*, 355(2):175–182, 2011.

[54] Bryan N Howie, Peter Donnelly, and Jonathan Marchini. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5(6):e1000529, 2009.

[55] Gary Huang. *Conditional and Marginal Distributions of a multivariate Gaussian*. Retrieved July 2015, 2010. URL https://gbhqed.wordpress.com/2010/02/21/conditional-and-marginal-distributions-of-a-multivariate-gaussian/.

[56] Hai-Dong Huang, Chun-Min Yang, Hai-Feng Shu, Yong-Qin Kuang, Tao Yang, Wei-Qi He, Kai Zhao, Xun Xia, Jing-Min Cheng, Yuan Ma, et al. Genetic predisposition of stroke: understanding the evolving landscape through meta-analysis. *International journal of clinical and experimental medicine*, 8(1):1315, 2015.

[57] Helen B Hubert, Manning Feinleib, Patricia M McNamara, and William P Castelli. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the framingham heart study. *Circulation*, 67(5): 968–977, 1983.

[58] Mia Hubert and Sanne Engelen. Robust pca and classification in biosciences. *Bioinformatics*, 20(11):1728–1736, 2004.

[59] Michael Hughes, Gregory YH Lip, Guideline Development Group for the NICE National Clinical Guideline for Management of Atrial Fibrillation in Primary, Secondary Care, et al. Stroke and thromboembolism in atrial fibrillation: a systematic review of stroke risk factors, risk stratification schema and cost effectiveness data. *Thrombosis and haemostasis*, 99(2):295–304, 2008.

[60] Juan E Iglesias, Ender Konukoglu, Darko Zikic, Ben Glocker, Koen Van Leemput, and Bruce Fischl. Is synthesizing mri contrast useful for inter-modality analysis? *MICCAI: International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 8149:631–638, 2013.

[61] Alexander Ilin and Tapani Raiko. Practical approaches to principal component

analysis in the presence of missing values. *The Journal of Machine Learning Research*, 11:1957–2000, 2010.

[62] Tommi S Jaakkola. Tutorial on variational approximation methods. *Advanced mean field methods: theory and practice*, pages 129–159, 2001.

[63] Clifford R Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

[64] Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87, 1991.

[65] Amod Jog, Aaron Carass, and Jerry L Prince. Improving magnetic resonance resolution with supervised learning. In *ISBI - IEEE International Symposium on Biomedical Imaging*, pages 987–990. IEEE, 2014.

[66] Brian Johnston, M Stella Atkins, Blair Mackiewich, and M Anderson. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Transactions on Medical Imaging*, 15(2):154–169, 1996.

[67] András Kelemen, Gábor Székely, and Guido Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18(10):828–839, 1999.

[68] George Kimeldorf and Grace Wahba. Some results on tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1):82–95, 1971.

[69] Ender Konukoglu, Andre van der Kouwe, Mert R Sabuncu, and Bruce Fischl. Example-based restoration of high-resolution magnetic resonance image acquisitions. *MICCAI: International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 8149:131–138, 2013.

[70] Richard E Latchaw, Mark J Alberts, Michael H Lev, John J Connors, Robert E Harbaugh, Randall T Higashida, Robert Hobson, Chelsea S Kidwell, Walter J Koroshetz, Vincent Mathews, et al. Recommendations for imaging of acute ischemic stroke a scientific statement from the American Heart Association. *Stroke*, 40(11):3646–3678, 2009.

[71] Michael E Leventon, W Eric L Grimson, and Olivier Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 316–323. IEEE, 2000.

[72] Yun Li, Cristen Willer, Serena Sanna, and Gonçalo Abecasis. Genotype imputation. *Annual review of genomics and human genetics*, 10:387, 2009.

[73] Chia-Wei Liao and Jun S Huang. Stroke segmentation by Bernstein-Bezier curve fitting. *Pattern Recognition*, 23(5):475–484, 1990.

[74] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.

[75] Dawei Liu, Xihong Lin, and Debashis Ghosh. Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics*, 63(4):1079–1088, 2007.

[76] Zhexing Liu, Yi Wang, Guido Gerig, Sylvain Gouttard, Ran Tao, Thomas Fletcher, and Martin Styner. Quality control of diffusion weighted images. In *SPIE Medical Imaging*, pages 76280J–76280J. International Society for Optics and Photonics, 2010.

[77] José V Manjón, Pierrick Coupé, Antonio Buades, Vladimir Fonov, D Louis Collins, and Montserrat Robles. Non-local MRI upsampling. *Medical image analysis*, 14(6):784–792, 2010.

[78] José V Manjón, Pierrick Coupé, Antonio Buades, D Louis Collins, and Montserrat Robles. New methods for MRI denoising based on sparseness and self-similarity. *Medical image analysis*, 16(1):18–27, 2012.

[79] Daniel S Marcus, Timothy R Olsen, Mohana Ramaratnam, and Randy L Buckner. The extensible neuroimaging archive toolkit. *Neuroinformatics*, 5(1):11–33, 2007.

[80] Peter McCullagh. Generalized linear models. *European Journal of Operational Research*, 16(3):285–292, 1984.

[81] Charles E McCulloch and John M Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001.

[82] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.

[83] James F Meschia, Donna K Arnett, Hakan Ay, Robert D Brown, Oscar R Benavente, John W Cole, Paul IW De Bakker, Martin Dichgans, Kimberly F Doheny, Myriam Fornage, et al. Stroke genetics network (sign) study design and rationale for a genome-wide association study of ischemic stroke subtypes. *Stroke*, 44(10): 2694–2702, 2013.

[84] Chandan Misra, Yong Fan, and Christos Davatzikos. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *Neuroimage*, 44(4):1415–1422, 2009.

[85] Marc Modat, Ivor JA Simpson, Manual Jorge Cardoso, David M Cash, Nicolas Toussaint, Nick C Fox, and Sébastien Ourselin. Simulating neurodegeneration through longitudinal population analysis of structural and diffusion weighted mri data. *MICCAI: International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 8675:57–64, 2014.

[86] Dariush Mozaffarian, Emelia J Benjamin, Alan S Go, Donna K Arnett, Michael J Blaha, Mary Cushman, Sandeep R Das, Sarah de Ferranti, Jean-Pierre Després, Heather J Fullerton, et al. Executive summary: Heart disease and stroke statistics-2016 update: A report from the american heart association. *Circulation*, 133(4):447, 2016.

[87] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4):869–877, 2005.

[88] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[89] JS Paulsen, DR Langbehn, JC Stout, Elizabeth Aylward, CA Ross, Martha Nance, Mark Guttman, Shannon Johnson, M MacDonald, LJ Beglinger, et al. Detection of Huntingtons disease decades before diagnosis: the Predict-HD study. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(8):874–880, 2008.

[90] Rita Peila, Beatriz L Rodriguez, and Lenore J Launer. Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies The Honolulu-Asia Aging Study. *Diabetes*, 51(4):1256–1262, 2002.

[91] Zhigang Peng, Jia Zhong, William Wee, and Jing-huei Lee. Automated vertebra detection and segmentation from the whole spine mr images. In *IEEE-EMBS: Annual International Conference of the Engineering in Medicine and Biology Society*, pages 2527–2530. IEEE, 2006.

[92] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7:15, 2008.

[93] Adolf Pfefferbaum, Torsten Rohlfing, Margaret J Rosenbloom, Weiwei Chu, Ian M Colrain, and Edith V Sullivan. Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85years) measured with atlas-based parcellation of MRI. *Neuroimage*, 65:176–193, 2013.

[94] Steve Pieper, Michael Halle, and Ron Kikinis. 3D Slicer. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pages 632–635. IEEE, 2004.

[95] Esben Plenge, Dirk HJ Poot, Wiro J Niessen, and Erik Meijering. Super-resolution reconstruction using cross-scale self-similarity in multi-slice MRI. *MICCAI: Medical Image Computing and Computer-Assisted Intervention*, LNCS 8151:123–130, 2013.

[96] Kilian M Pohl, John Fisher, Sylvain Bouix, Martha Shenton, Robert W McCarley, W Eric L Grimson, Ron Kikinis, and William M Wells. Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis*, 11 (5):465–477, 2007.

[97] Steven G Potkin, Guia Guffanti, Anita Lakatos, Jessica A Turner, Frithjof Kruggel, James H Fallon, Andrew J Saykin, Alessandro Orro, Sara Lupoli, Erika Salvi, et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for alzheimer's disease. *PloS one*, 4(8):e6501, 2009.

[98] Bruce M Psaty, Christopher J O'Donnell, Vilmundur Gudnason, Kathryn L Lunetta, Aaron R Folsom, Jerome I Rotter, André G Uitterlinden, Tamara B Harris, Jacqueline CM Witteman, Eric Boerwinkle, et al. Cohorts for heart and

aging research in genomic epidemiology (charge) consortium design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circulation: Cardiovascular Genetics*, 2(1):73–80, 2009.

[99] Richard E Quandt and James B Ramsey. Estimating mixtures of normal distributions and switching regressions. *Journal of the American statistical Association*, 73(364):730–738, 1978.

[100] David C Queller and Keith F Goodnight. Estimating relatedness using genetic markers. *Evolution*, pages 258–275, 1989.

[101] Miriam R Raffeld, Stephanie Debette, and Daniel Woo. International stroke genetics consortium update. *Stroke*, 47(4):1144–1145, 2016.

[102] Elizabeth A Regan, John E Hokanson, James R Murphy, Barry Make, David A Lynch, Terri H Beaty, Douglas Curran-Everett, Edwin K Silverman, and James D Crapo. Genetic epidemiology of copd (COPDgene) study design. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 7(1):32–43, 2011.

[103] Eric M Reiman, Jennifer A Webster, Amanda J Myers, John Hardy, Travis Dunckley, Victoria L Zismann, Keta D Joshipura, John V Pearson, Diane Hu-Lince, Matthew J Huentelman, et al. GAB2 alleles modify Alzheimer's risk in APOE 4 carriers. *Neuron*, 54(5):713–720, 2007.

[104] John H Relethford. Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology*, 118(4): 393–398, 2002.

[105] David E Rex, Jeffrey Q Ma, and Arthur W Toga. The LONI pipeline processing environment. *Neuroimage*, 19(3):1033–1048, 2003.

[106] Shannon L Risacher, Li Shen, John D West, Sungeun Kim, Brenna C McDonald, Laurel A Beckett, et al. Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. *Neurobiology of aging*, 31(8):1401–1418, 2010.

[107] Torsten Rohlfing, Edith V Sullivan, and Adolf Pfefferbaum. Regression models of atlas appearance. *International Conference on Information Processing in Medical Imaging*, LNCS 5636:151–162, 2009.

[108] Antoine Rosset, Luca Spadola, and Osman Ratib. Osirix: an open-source software for navigating in multidimensional dicom images. *Journal of Digital Imaging*, 17 (3):205–216, 2004.

[109] Natalia S Rost, Kaitlin Fitzpatrick, Alessandro Biffi, Allison Kanakis, William Devan, Christopher D Anderson, Lynelle Cortellini, Karen L Furie, and Jonathan Rosand. White matter hyperintensity burden and susceptibility to cerebral ischemia. *Stroke*, 41(12):2807–2811, 2010.

[110] Natalia S Rost, Rosanna M Rahman, Alessandro Biffi, Eric E Smith, Allison Kanakis, Kaitlin Fitzpatrick, Fabricio Lima, Bradford B Worrall, James F Meschia, Robert D Brown, et al. White matter hyperintensity volume is increased in small vessel stroke subtypes. *Neurology*, 75(19):1670–1677, 2010.

[111] François Rousseau, Kio Kim, and Colin Studholme. A groupwise super-resolution approach: application to brain MRI. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 860–863. IEEE, 2010.

[112] Frederic Rousseau, Piotr A Habas, and Colin Studholme. A supervised patch-based approach for human brain labeling. *IEEE Transactions on Medical Imaging*, 30(10):1852–1862, 2011.

[113] Mert R Sabuncu and Koen Van Leemput. The relevance voxel machine (RVoxM): a bayesian method for image-based prediction. *MICCAI: International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 6893:99–106, 2011.

[114] Ralph L Sacco, Robert Adams, Greg Albers, Mark J Alberts, Oscar Benavente, Karen Furie, Larry B Goldstein, Philip Gorelick, Jonathan Halperin, Robert Harbaugh, et al. Guidelines for prevention of stroke in patients with ischemic stroke or transient ischemic attack a statement for healthcare professionals from the american heart association/american stroke association council on stroke: Co-sponsored by the council on cardiovascular radiology and intervention: The american academy of neurology affirms the value of this guideline. *Circulation*, 113 (10):e409–e449, 2006.

[115] Neda Sadeghi, Marcel Prastawa, P Thomas Fletcher, Clement Vachet, Bo Wang, et al. Multivariate modeling of longitudinal mri in early brain development with

confidence measures. In *IEEE International Symposium on Biomedical Imaging*, pages 1400–1403. IEEE, 2013.

[116] Balasrinivasa Rao Sajja, Sushmita Datta, Renjie He, Meghana Mehta, Rakesh K Gupta, Jerry S Wolinsky, and Ponnada A Narayana. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Annals of Biomedical Engineering*, 34(1):142–151, 2006.

[117] Serena Sanna, Anne U Jackson, Ramaiah Nagaraja, Cristen J Willer, Wei-Min Chen, Lori L Bonnycastle, Haiqing Shen, Nicholas Timpson, Guillaume Lettre, Gianluca Usala, et al. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nature genetics*, 40(2):198–203, 2008.

[118] Isaac J Schoenberg and Isaac Jacob Schoenberg. *Cardinal spline interpolation*, volume 12. SIAM, 1973.

[119] Jonathan M Schott, Jonathan W Bartlett, Nick C Fox, and Josephine Barnes. Increased brain atrophy rates in cognitively normal older adults with low cerebrospinal fluid Aβ1-42. *Annals of neurology*, 68(6):825–834, 2010.

[120] William J Schroeder, Lisa Sobierajski Avila, and William Hoffman. Visualizing with VTK: a tutorial. *IEEE Computer Graphics and Applications*, 20(5):20–27, 2000.

[121] Florent Ségonne, Anders M Dale, Evelina Busa, Maureen Glessner, David Salat, Horst K Hahn, and Bruce Fischl. A hybrid approach to the skull stripping problem in mri. *Neuroimage*, 22(3):1060–1075, 2004.

[122] Li Shen, Sungeun Kim, Shannon L Risacher, Kwangsik Nho, Shanker Swaminathan, John D West, Tatiana Foroud, Nathan Pankratz, Jason H Moore, Chantel D Sloan, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage*, 53(3):1051–1063, 2010.

[123] Wenzhe Shi, Jose Caballero, Christian Ledig, Xiahai Zhuang, Wenjia Bai, et al. Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. *MICCAI: Medical Image Computing and Computer-Assisted Intervention*, LNCS 8151:9–16, 2013.

[124] Ramesh Sridharan, Adrian V Dalca, Kaitlin M Fitzpatrick, Lisa Cloonan, Allison Kanakis, Ona Wu, Karen L Furie, Jonathan Rosand, Natalia S Rost, and Polina Golland. Quantification and analysis of large multimodal clinical image studies: Application to stroke. *MICCAI-MBIA Multimodal Brain Image Analysis Workshop, LNCS*, LNCS 8159:18–30, 2013.

[125] Ramesh Sridharan, Adrian V Dalca, and Polina Golland. An interactive visualization tool for nipype medical image computing pipelines. In *MICCAI-IMIC Interactive Medical Image Computing Workshop*, 2014.

[126] Tony Stöcker, Frank Schneider, Martina Klein, Ute Habel, Thilo Kellermann, Karl Zilles, and N Jon Shah. Automated quality assurance routines for fmri data applied to a multicenter study. *Human brain mapping*, 25(2):237–246, 2005.

[127] Martin Styner, Jeffrey A Lieberman, Dimitrios Pantazis, and Guido Gerig. Boundary and medial shape analysis of the hippocampus in schizophrenia. *Medical Image Analysis*, 8(3):197–203, 2004.

[128] Ming-Xin Tang, Yaakov Stern, Karen Marder, Karen Bell, Barry Gurland, Rafael Lantigua, Howard Andrews, Lin Feng, Benjamin Tycko, and Richard Mayeux. The apoe- 4 allele and the risk of alzheimer disease among african americans, whites, and hispanics. *Jama*, 279(10):751–755, 1998.

[129] Paul M Thompson, Tian Ge, David C Glahn, Neda Jahanshad, and Thomas E Nichols. Genetics of the connectome. *Neuroimage*, 80:475–488, 2013.

[130] Michael E Tipping and Christopher M Bishop. Mixtures of probabilistic principal component analyzers. *Neural computation*, 11(2):443–482, 1999.

[131] Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[132] Matthew Traylor, Martin Farrall, Elizabeth G Holliday, Cathie Sudlow, Jemma C Hopewell, Yu-Ching Cheng, Myriam Fornage, M Arfan Ikram, Rainer Malik, Steve Bevan, et al. Genetic risk factors for ischaemic stroke and its subtypes (the metastroke collaboration): a meta-analysis of genome-wide association studies. *The Lancet Neurology*, 11(11):951–962, 2012.

[133] Andy Tsai, Anthony Yezzi, William Wells, Clare Tempany, Dewey Tucker, Ayres Fan, W Eric Grimson, and Alan Willsky. Model-based curve evolution technique for image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–463. IEEE, 2001.

[134] John W Tukey. *Exploratory data analysis*. Reading, Mass., 1977.

[135] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan Colchester, and Paul Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8):677–688, 2001.

[136] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. *MICCAI: International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 5241:754–761, 2008.

[137] Maria Vounou, Thomas E Nichols, Giovanni Montana, and Alzheimer's Disease Neuroimaging Initiative. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage*, 53(3):1147–1159, Nov 2010.

[138] Grace Wahba. *Spline models for observational data*, volume 59. Siam, 1990.

[139] Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.

[140] Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack, William Jagust, Enchi Liu, et al. The alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *Alzheimer's & Dementia*, 9(5):e111–e194, 2013.

[141] William M Wells, W Eric L Grimson, Ron Kikinis, and Ferenc A Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[142] Anderson M Winkler, Peter Kochunov, John Blangero, Laura Almasy, Karl Zilles, Peter T Fox, Ravindranath Duggirala, and David C Glahn. Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *Neuroimage*, 53(3):1135–1146, 2010.

[143] Max Wintermark, Gregory W Albers, Joseph P Broderick, Andrew M Demchuk, Jochen B Fiebach, Jens Fiehler, James C Grotta, Gary Houser, Tudor G Jovin, Kennedy R Lees, et al. Acute stroke imaging research roadmap II. *Stroke*, 44(9): 2628–2639, 2013.

[144] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust PCA via outlier pursuit. In *Advances in Neural Information Processing Systems*, pages 2496–2504, 2010.

[145] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010.

[146] Jianhua Yao, Stacy D O'Connor, and Ronald M Summers. Automated spinal column extraction and partitioning. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pages 390–393. IEEE, 2006.

[147] Ruoqiao Zhang, Charles A Bouman, Jean-Baptiste Thibault, and Ken D Sauer. Gaussian mixture markov random field for image denoising and reconstruction. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1089–1092. IEEE, 2013.